SMART MOBILE DATA COLLECTION IN THE CONTEXT OF NEUROSCIENCE

EDITED BY: Rüdiger Christoph Pryss, Thomas Probst, Winfried Schlee, Manfred Reichert, Myra Spiliopoulou and Berthold Langguth

PUBLISHED IN: Frontiers in Neuroscience, Frontiers in Public Health and Frontiers in Medicine

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-069-0
DOI 10.3389/978-2-88971-069-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SMART MOBILE DATA COLLECTION IN THE CONTEXT OF NEUROSCIENCE

Topic Editors:
Rüdiger Christoph Pryss, Julius Maximilian University of Würzburg, Germany
Thomas Probst, Danube University Krems, Austria
Winfried Schlee, University of Regensburg, Germany
Manfred Reichert, University of Ulm, Germany
Myra Spiliopoulou, Otto von Guericke University Magdeburg, Germany
Berthold Langguth, University of Regensburg, Germany

Citation: Pryss, R. C., Probst, T., Schlee, W., Reichert, M., Spiliopoulou, M., Langguth, B., eds. (2021). Smart Mobile Data Collection in the Context of Neuroscience. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-069-0

Table of Contents

frontiers in Neuroscience

EDITORIAL
published: 25 May 2021
doi: 10.3389/fnins.2021.698597

Editorial: Smart Mobile Data Collection in the Context of Neuroscience

Rüdiger Pryss[1]*, *Berthold Langguth*[2], *Thomas Probst*[3], *Winfried Schlee*[2], *Myra Spiliopoulou*[4] and *Manfred Reichert*[5]

[1] Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany, [2] Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany, [3] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems an der Donau, Austria, [4] Institute of Technical and Business Information Systems, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany, [5] Institute of Databases and Information Systems, Ulm University, Ulm, Germany

Keywords: mobile data collection, ecological momentary assessments, reviews, meta-analyses, multi-modal data fusion, experience sampling, digital phenotyping, mobile crowdsensing

Editorial on the Research Topic

Smart Mobile Data Collection in the Context of Neuroscience

The Covid-19 pandemic demonstrates both the potential of mobile technology in medicine and the need to exploit this potential (e.g., Zhang et al., 2021). On the flip side, many challenges have to be addressed, for which still no suitable answer exists. This ranges from technical frameworks to the reliability of research data, which have been collected with the help of mobile technology. It is striking that the number of mobile data collection strategies in healthcare grows on a frequent basis. Therefore, overarching topics and considerations are increasingly needed to keep pace with these trends, particularly through their categorization and evaluation. In this Research Topic, such an attempt was pursued for mobile data collection in the context of neuroscience.

Aim of this Research Topic: Digital phenotyping, experience sampling, digital health, and ecological momentary assessments (EMA) are only a few methods and strategies that have been presented at the intersection of mobile technology and healthcare. The need to involve multiple disciplines constitutes another important recognition in the given field. Therefore, efforts are constantly needed to review and categorize presented research works. We started this Research Topic in 2018, as we saw and still see many open questions when using mobile technology in healthcare scenarios, especially in the context of clinical neuroscience. The submitted works addressed interesting and novel aspects, also driven through the insight of the identified submission categories. For example, a category multi-modal data fusion could be identified, which will certainly become increasingly important to foster evidence in the context of mobile technology and the neuroscience domain. Altogether, many initially planned aims have been actually pursued by the submissions, but also other directions were presented.

Overview: The works submitted to this topic (including rejected papers) cover many aspects of the pursued topic goals. However, we were able to identify three major categories for the contributions: *Experience Sampling*, *Multi-modal Data Fusion*, and *Reviews and Meta-Analyses*, particularly showing two objectives of the research field of mobile data collection when being used for neuroscience questions. The collection of ecologically valid data through experience sampling or multi-modal data fusion constitutes the first objective. Note that the latter method enables cross-validation of mobile data with other data sources collected through already established methods to gain better insights on the achieved data validity. The second objective, in turn, is concerned with general insights based on Reviews and Meta-Analyses.

OPEN ACCESS

Edited and reviewed by:
Michele Giugliano, International School for Advanced Studies (SISSA), Italy

***Correspondence:**
Rüdiger Pryss
ruediger.pryss@uni-wuerzburg.de

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *21 April 2021*
Accepted: *30 April 2021*
Published: *25 May 2021*

Citation:
Pryss R, Langguth B, Probst T, Schlee W, Spiliopoulou M and Reichert M (2021) Editorial: Smart Mobile Data Collection in the Context of Neuroscience. Front. Neurosci. 15:698597. doi: 10.3389/fnins.2021.698597

EXPERIENCE SAMPLING

The terms Experience Sampling, Ambulatory Assessment and EMA describe a new research method allowing to systematically collect self-reports of emotions, cognition, and behavior in the real-world settings of the participants. A smartphone, wearable or other mobile signaling device is used to prompt the participant with a short questionnaire asking questions on the current situation. This method can then be used to assess fluctuations of clinical symptoms within and between days with high ecological validity and low recall bias (e.g., Probst, 2017; Pryss, 2019).

In this Research Topic, two studies took advantage of this research method: Stieger and Kuhlmann used an experience sampling app to collect a dataset on dreams and nightmares in 92 participants over a period of 22 days. In their paper, they report a detailed item-analysis on the Nightmare Distress Questionnaire and identified those items that reliably discriminate between bad sleep and nightmare. In a study by Weierstall-Pust et al., tablets were used to assess the Post-Traumatic Stress Syndrome in 463 soldiers in Burundi. Based on this field data, the authors were able to discriminate subgroups of soldiers with distinct symptom profiles.

The authors Wurzer and Hauptmann describe an approach, in which a mobile device was used for the treatment of chronic tinnitus. Using a single-arm study design, an auditory stimulation device was tested on a group of 25 tinnitus patients with a treatment period of 16 weeks.

MULTI-MODAL DATA FUSION

The early contribution of Sariyska et al. on feasibility of linking molecular genetic markers to real-world social network size tracked on smartphones linked genetic data to data about the social behavior of people, as recorded with a smartphone app, and reported on an association between genetic expression and social network size of the study participants. This is an example of combining genetic and mobile data for digital phenotyping, and points also to practical limitations: while mobile data are easily scalable, the acquisition of genetic data is an elaborate task, which requires time and money, and for which volunteer recruitment is less easy. This results in smaller samples and less robust models.

Huckins et al. contributed a study on fusing mobile phone sensing and brain imaging to assess depression in college students, in which they identified associations between the time a user had their smartphone unlocked and functional brain activity in brain regions associated with depression. The authors combined passive smartphone sensing data, EMA and functional brain scans and investigated how smartphone usage can be linked to brain connectivity metrics. Their results, derived in one cohort and verified in a second cohort, demonstrate that multi-modal data fusion can lead to new ways of assessing mental health, but also revealed several challenges for the immediate future. Among them, the curse of dimensionality implies that the extraction of knowledge from the high-dimensional fused data requires very large participant samples and dimensionality reduction approaches.

In toward personalized tinnitus treatment: an exploratory study based on internet crowdsensing, Simoes et al. analyzed data of a self-help platform for tinnitus patients to identify predictors of treatment response. Among other findings, the authors showed that treatment duration is the variable explaining most of the variance concerning treatment outcome. Such findings indicate the potential of internet crowdsensing for generating hypotheses for personalized treatments.

In mental condition monitoring based on multimodality biometry, Kiguchi et al. captured multi-modal data to assess mental distress in the workplace. The authors used devices for activity tracking, sleep monitoring, and logging of interactions with office PCs, and they showed that the tracked data agreed with the perceived mental condition of the volunteers, as recorded in questionnaires. Their results suggest that data tracking at the work place has the potential to inform about mental stress.

In motorized shoes induce robust sensorimotor adaptation in walking, Aucie et al. investigated whether results on locomotor adaptation outside the lab agree with insights won in a controlled laboratory environment. They juxtaposed the locomotory behavior of a control group to that of a group wearing motorized shoes with elaborate gear over their normal shoes, and they found that the two groups exhibited mostly the same adaptation patterns. Their results show the potential of wearable devices in modeling movement and gait under real conditions.

REVIEWS AND META-ANALYSES

Research on smartphones in neuroscience is strongly increasing. In PubMed.gov, the search terms "smartphone AND neuroscience" revealed 51 publications for 2016, 71 publications for 2017, 97 publications for 2018, 122 publications for 2019, and 170 publications for 2020. Reviews and meta-analyses are necessary to synthesize findings for both researchers and practitioners. The review and meta-analysis presented by Goreis et al. examined the potential of smartphone apps to reduce post-traumatic stress symptoms. While a reduction of post-traumatic stress symptoms was observed in participants using apps ($g = 0.55$), the effect was not significantly stronger than in participants not using the apps ($g = 0.09$).

Seifert et al. summarized six challenges (collecting data in real-life environments, real-time measurements, within-person data, passive data, smartphone device, data security, and ethical issues) for a smarter smartphone research in neuroscience and provided valuable ideas to overcome these challenges in the future.

The use of smartphone apps to combine mobile crowdsensing and EMA is the focus of the manuscript of Kraft et al. The authors review their experience gathered in various smartphone app projects (TrackYourTinnitus, TrackYourStress, TrackYourHearing, TrackYourDiabetes, Intersession, KINDEX, TinnitusTipps) and provide recommendations for the developments of platforms that combine mobile crowdsensing and ecological momentary assessments.

SUMMARY

The articles in this special topic provide an overview about the rapidly growing area of mobile data collection in clinical

neuroscience. The presented research indicates their huge potential as well as the manifold and multifaceted areas of application.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to all contributors of this Research Topic. Ninety-eight different authors contributed with research and review articles. Furthermore, we thank the reviewers who helped us and the authors to create an interesting and high-quality Research Topic. We hope that readers will enjoy reading this Research Topic as much as we have enjoyed editing it.

REFERENCES

Probst, T. (2017). Does tinnitus depend on time-of-day? An ecological momentary assessment study with the "TrackYourTinnitus" application. *Front. Aging Neurosci.* 9:253. doi: 10.3389/fnagi.2017.00253

Pryss, R. (2019). Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the TrackYourTinnitus mobile platform. *Int. J. Data Sci. Anal.* 8, 327–338. doi: 10.1007/s41060-018-0111-4

Zhang, M. W. B., Chow, A., Ho, R. C. M., and Smith, H. E. (2021). An overview of commercially available apps in the initial months of the covid-19 pandemic. *Front. Psychiatry* 12:557299. doi: 10.3389/fpsyt.2021.557299

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 10 October 2018
doi: 10.3389/fmed.2018.00288

Adapted Acoustic CR Neuromodulation in Patients With Chronic Tonal Tinnitus and Hearing Loss

Hannes Wurzer[1] *and Christian Hauptmann*[2]*

[1] Tinnitus Center, Munich, Germany, [2] Desyncra Operating GmbH, Bad Neuenahr-Ahrweiler, Germany

Chronic tonal tinnitus is often accompanied by sensorineural hearing loss which is associated with altered tuning curves and bandwidth of alternating masking. In this feasibility study the so-called hearing threshold adapted coordinated reset (HTA-CR) neuromodulation was investigated. This method is based on CR neuromodulation, which has been demonstrated to be an effective treatment for chronic tonal tinnitus. It applies four stimulation tones that are determined by the patient's individual tinnitus frequency and hearing impairment. The HTA-CR neuromodulation was programmed to the Desyncra™ for Tinnitus Therapy System and treatment was applied to 25 patients for 4 months on average and 4 h daily. Regular check-ups were done every 4–6 weeks. Therapy outcome was assessed by the tinnitus questionnaire (*Tinnitusfragebogen*, TF) as per Goebel and Hiller. After 4 months the mean TF score was reduced by 27.4%. A reduction of $\geq$ 15 points was found in 40% of the patients while for further 32% of the patients a reduction of 6–14 points was found. Thus, a positive response rate of 72% was observed after 4 months of HTA-CR neuromodulation. Our results suggest that HTA-CR neuromodulation might be at least comparable to standard CR neuromodulation providing another effective therapeutic option for the treatment of chronic tonal tinnitus.

Keywords: tinnitus, neuromodulation, acoustic stimulation, therapy, chronic disease

OPEN ACCESS

Edited by:
Winfried Schlee,
University of Regensburg, Germany

Reviewed by:
Jae-Jin Song,
Seoul National University Bundang Hospital, South Korea
Ricardo Rodrigues Figueiredo,
Faculdade de Medicina de Valença, Brazil

***Correspondence:**
Christian Hauptmann
christian.hauptmann@desyncra.com

Specialty section:
This article was submitted to Translational Medicine, a section of the journal Frontiers in Medicine

Received: *08 August 2018*
Accepted: *21 September 2018*
Published: *10 October 2018*

Citation:
Wurzer H and Hauptmann C (2018) Adapted Acoustic CR Neuromodulation in Patients With Chronic Tonal Tinnitus and Hearing Loss. Front. Med. 5:288. doi: 10.3389/fmed.2018.00288

INTRODUCTION

Chronic tinnitus is an otorhinolaryngological disease affecting ~10–15% of the general population in industrialized countries (1). The permanent perception of sound in the absence of a corresponding sound source is often associated with hearing loss secondary, among others, to noise exposure or aging. Recent studies using advanced imaging techniques and functional approaches indicated that chronic tinnitus is a neural consequence of acoustic acquired sensory deprivation (2–5) leading to an imbalance of excitatory and inhibitory neural networks in the central auditory pathway. These alterations result in elevated spontaneous activity (6) and synchronization of neurons (7, 8) in the auditory cortex, virtually a "tinnitus generator," which is perceived as tinnitus. An overview of involved neuropathophysiological circuits has been published recently by Rauschecker et al. (9).

Because only 20% of patients with chronic tinnitus are affected by a severe impairment of quality of life (10) it is assumed that the limbic system, the "emotional brain," and the vegetative neuronal network including the *formatio recticularis* impact the level of suffering experienced by

patients (11–14). This aspect is recognized and implemented in various therapeutical approaches for tinnitus treatment such as use of psychotropic drugs, relaxation techniques (e.g., progressive relaxation, Tai Chi, etc.), physiotherapy, cognitive behavioral therapy (CBT), and tinnitus retraining therapy (TRT) (15–17). In particular in Germany, TRT has been modified substantially and is often combined with a sound therapy or tinnitus noisers, i.e., noise generators used to mask or cover up the tinnitus sensation (18). So far, only for CBT efficiency in the treatment of tinnitus has been demonstrated in a randomized clinical trial (19).

Various therapies for chronic tinnitus have been developed aiming to manipulate the hyperactive "tinnitus generator" in the auditory cortex by neuromodulation via acoustic stimulation. Beside some special German methods like the Heidelberger music therapy (20) and the tinnitus-centered music therapy (TIM) developed by Cramer (21), tinnitus noisers and maskers are widely used (18). However, the recommendations for therapy of chronic tinnitus with these instruments differ substantially, and sufficiently large randomized clinical trials to demonstrate the efficacy of commonly used interventions are lacking (22). Moreover, none of the in Germany currently available tinnitus noisers is able to efficiently mask tinnitus tones with frequencies above 11 kHz. Thus, more advanced noisers have been developed which offer additional acoustic stimulation such as sounds of rainfall, ocean waves or wind chime.

Another approach of acoustic neuromodulation is the so-called tailor-made notch music therapy (TMNMT), where patients listen to music from which a frequency band of one octave around their individual tinnitus frequency had been removed (23). In this way neurons in the auditory cortex coding the tinnitus frequency are subject to lateral inhibition by neighboring neurons, whereas the afferent input to these neurons is negligible. Results from magnetoencephalography (MEG) have shown that the auditory evoked cortex activity was reduced after 12 months and patients reported reduced subjective tinnitus loudness (23). However, a recent double-blind randomized controlled trial applying TMNMT for 3 months on 100 patients with chronic tonal tinnitus failed to show an improvement on tinnitus distress assessed by the Tinnitus Questionnaire (TQ) (24).

A different approach is the acoustic coordinated reset (CR) neuromodulation, which was developed by the Research Center Juelich, Germany, and is based on theoretical and clinical studies (25–28). The method aims to desynchronize the synchronous hyperactive neuron population coding the tinnitus frequency by sequential stimulation of different subpopulations of the target population (25, 26, 29). For this purpose, an extensive pitch-matching procedure is used to determine the individual tinnitus frequency (30). Four acoustic stimulation tones are then generated based on a computationally developed CR algorithm. These tones are of different frequencies centered around the patient's tinnitus frequency, and their loudness is individually adapted to the loudness of the tinnitus. Via dedicated sound generators and ear phones the patients are exposed to the acoustic stimulation for several hours per day. The first blinded study reported by Tass et al. in 2012 with 63 patients proved the safety and showed clinical efficacy of CR neuromodulation (28). Since then several clinical studies with more than 500 patients suffering from chronic tonal tinnitus have been conducted. The results revealed that 60–75% of patients treated with CR neuromodulation respond well to this therapy as evident from decrease in tinnitus questionnaire scores and improvement of visual analog scale (VAS) scores for tinnitus loudness and annoyance by ∼40% (14, 31, 32, Wurzer et al. submitted). Importantly, these improvements are persistent and stable. MEG and electroencephalography (EEG) data revealed specific alterations in patients with chronic tonal tinnitus such as increased oscillatory power in the delta frequency range and decreased alpha power in the auditory cortex region (12, 33, 34). Upon CR neuromodulation a normalization of the EEG pattern was observed in therapy responding patients (14, 35). Despite these promising results for the majority of chronic tinnitus patients ∼30% of these patients experience no improvement of their tinnitus after CR neuromodulation, the so-called non-responders. Accordingly, efforts are made to further optimize the therapy.

The currently used CR neuromodulation stimulation tones depend exclusively on the tinnitus frequency. Theoretically, a 25–30% overlap of the stimulation tone's tuning curves would be optimal. However, several years of clinical experience support the notion that the patient's hearing ability should be considered in the individual adaption of the CR neuromodulation therapy. Sensorineural hearing loss is known to be associated with altered tuning curves, altered bandwidth of alternating masking and a changed discrimination that might be relevant for the acoustic overlap of the stimulation tones (36). Therefore, in this feasibility study for the first time the patient's tinnitus frequency, the intensity of the adjusted stimulation tones as well as his/her individual audiogram were taken into account when calculating the stimulation tones for the so-called hearing threshold adapted CR neuromodulation therapy (HTA-CR). Basically, the intensity of the stimulation tones was identified to be responsible for the neuronal recruitment. Stronger stimulation tones recruit more neurons and have therefore a flatter tuning curve (36). Since the individual subject is asked to adjust the stimulation such that the tones are audible the adjusted tone intensity strongly depends on the hearing threshold. The experiments done by Hopkins and Moore (36) provide the database linking the tone intensities and hearing loss with the shape of the tuning curve, which is used to calculate the adapted stimulation tones in an iterative approach. The simple study design lacks any control group, since the goal of this feasibility study was to test, if a hearing threshold adapted CR neuromodulation could be applied and to obtain first results which can be used as input for a larger, randomized and controlled study.

The Desyncra™ for Tinnitus medical devices (30) used in this study were programmed accordingly. The primary objective of this study was to verify a clinically significant improvement of chronic tinnitus by this adapted CR neuromodulation within 3–4 months, which was assessed by the tinnitus questionnaire (*Tinnitus-Fragebogen*, TF) as per Goebel and Hiller [German version of Hallam's TQ (37)].

MATERIALS AND METHODS

Study Participants

Twenty-five patients attending a specialized consultation session for tinnitus were enrolled in this clinical study on acoustic CR neuromodulation, which was conducted in a specialized center run by an ear, nose and throat (ENT) specialist located in Munich, Germany. All patients were comprehensively informed about the scope, aim, benefits, and risks of study participation, and a written informed consent was obtained from all participants according to the Declaration of Helsinki and Good Clinical Practice. The study was approved by the relevant ethics committee (Ethics Committee of the Bavarian State Medical Association, BLAEK 2016-136). The inclusion and exclusion criteria are listed in **Table 1**.

Screening data obtained during the initial examination (visit 0) are summarized in **Table 2**. The patients' age was between 21 years and 82 years (mean 51.3 years); the male to female proportion was 16–9.The hearing ability ranged from normal to severe hearing loss. The average hearing loss was calculated from the sum of hearing loss at 0.25, 0.5, 1, 2, 4, 6, and 8 kHz. The frequencies 6 and 8 kHz were included in the calculation taking the hearing loss in the high frequency range of the patients into account. Patients were classified into three hearing groups: group 1 with normal hearing ability (mean hearing impairment < 20 dB); group 2 with mild hearing loss (mean hearing impairment between 20 and 40 dB), and group 3 with moderate to severe hearing loss (mean hearing impairment > 40 dB, **Table 2**) (31, 38). Five patients using hearing aids were included in the study.

TABLE 1 | Overview of inclusion and exclusion criteria.

Inclusion criteria
1. Age ≥ 18 years
2. Primary chronic tinnitus ≥3 months [defined by AAO HNSF guidelines (40)]
3. Tonal tinnitus
4. Tinnitus frequency of 0.4–10 kHz (in exceptional cases up to 12 kHz)
5. TF score > 30 (i.e., at least severity grade II)
6. Able to hear all stimulation tones
7. Commitment to wear the device for 4–6 h/day
8. No other tinnitus treatment in the period of the clinical investigation
Exclusion criteria
1. Secondary / somatic tinnitus [defined by AAO HNSF guidelines (40)]
2. Atonal, pulsatile, or intermittent tinnitus
3. Hearing loss > 70 dB HL between 0.25 to 10 kHz
4. Tinnitus main frequency differs >10% between right and left ear
5. Health related or other reasons that might prevent the patient to complete the study
6. Use of medication that might cause tinnitus, i.e., daily high-dosed NSAIDs (≥1,000 mg/d) and salicylates at doses higher than for cardio-protection, loop diuretics, chemotherapy agents (e.g., cisplatin)
7. Permanent conductive hearing loss ≥15 dB for more than two frequencies in one ear
8. Persistent eardrum defect
9. Atresia or malformation of the outer ear
10. Acute otorrhoea
11. Acute or fast progressing hearing loss within the last 90 days
12. Severe psychiatric disorders

Chronic tinnitus duration was between 1 and 35 years (median 8.0 years). The mean TF score at inclusion was 50.4 ± 12.9 points, and most of the patients suffered from tinnitus with severity grade ≥ 3. Moreover, all patients have been treated before study start with at least two different therapeutic approaches (cortisone infusion, noisers, acupuncture, etc.) without success.

Description of the Medical Device

For HTA-CR neuromodulation the Desyncra™ for Tinnitus Therapy System was used, which is a Class IIa medical device (CE-0123, certified 2016 by TÜV Süd, Germany; FDA certified 2016 by FDA, no. K151558). To adapt the system for HTA-CR neuromodulation, the relation of the stimulation tones to each other and to the individual tinnitus frequency was modified taking into account the individually adjusted tone intensity and measured hearing loss (**Figure 1**). The algorithm of time and tone sequence as used for the standard acoustic CR neuromodulation was left unaltered. The adjustments were done according to the guidelines and controlled after 4–6 weeks. The patients were asked to use the device daily for 4 h on average. The total duration of the study was 4 months.

Study Conduct

During the initial examination (visit 0) medical and tinnitus history was assessed, a medical examination of ear, nose, throat, and the stomatognathic system was performed as well as manual examination of the cervical spine. A hearing threshold and high frequency (11–18 kHz) audiogram was recorded, and a speech audiogram and evaluation of hearing aid performance was performed if required. Moreover, measurement of impedance and otoacoustic emissions, repeated tinnitus pitch matching and tuning curve measurement at 1 kHz were conducted. Patients completed the TF and, if found eligible and interested in study

TABLE 2 | Screening data of study population ($n = 25$).

Variable	Mean ± SD (range)
AGE (YEARS)	
Overall	51.3 ± 13.6 (21–82)
By Hearing Impairment	
Group 1 (< 20 dB, $n = 11$)	46.3 ± 12.2 (21–64)
Group 2 (20–40 dB, $n = 7$)	47.9.4 ± 6.8 (39–59)
Group 3 (> 40 dB, $n = 7$)	62.6 ± 15.5 (37–82)
Tinnitus duration (years)	8.0* (1–35)
TF score	50.4 ± 12.9 (32–73)
	Number of patients (%)
SEX	
# Male	16 (64.0%)
# Female	9 (36.0%)
TINNITUS SEVERITY	
Grade 1 (0–30 points)	0 (0.0%)
Grade 2 (31–46 points)	12 (48.0%)
Grade 3 (47–59 points)	6 (24.0%)
Grade 4 (60–84 points)	7 (28.0%)

Value given as median.

participation, they received the EC approved patient information and informed consent form.

At the start of the study (= visit 1) patients were fully informed about the study and signed the informed consent form. ENT examinations, audiogram recording, and tinnitus pitch matching procedures were repeated, and patients completed the following questionnaires to assess various aspects of the tinnitus: TF, *Tinnitus-Beeintraechtigungsbogen* (TBF-12), Tinnitus Handicap Inventory (THI, German version), and Tinnitus Functional Index (TFI, only available in English). The TFI contains numeric rating scales (NRS) which were used to assess tinnitus loudness (NRS-L) and annoyance (NRS-A). The baseline scores of the questionnaires are summarized in **Table 3**. After completing all study assessments of visit 1 patients received their individually programmed mobile device and were also provided with detailed instructions and information relating to device usage at home.

Check-ups were performed after 5 ± 1 weeks (= visit 2) and 9 ± 1 weeks (= visit 3). During the visits questionnaires were repeated, tinnitus pitch matching performed and the stimulation tones adjusted to the latest tinnitus frequency. Adverse events, adverse device effects and general problems with the device were collected.

The final visit took place after 16 ± 2 weeks (visit 4) with determination of the latest tinnitus frequency, tuning curve assessment, audiogram control, and completion of TF, TBF12, THI, and TFI by the patient.

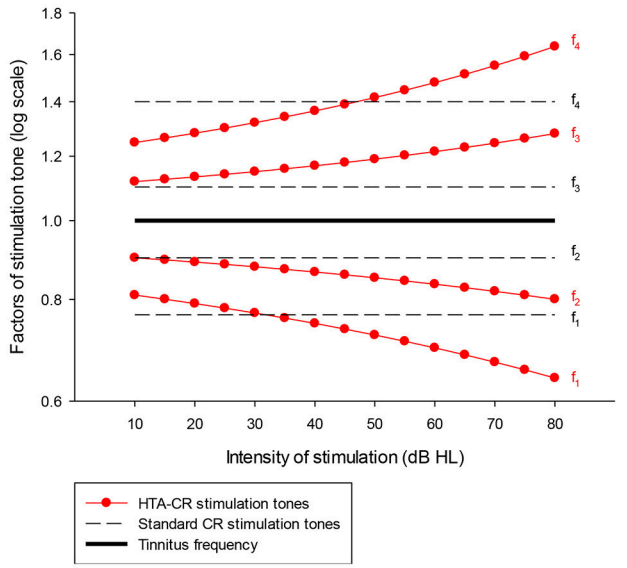


FIGURE 1 | Graphic presentation of the factors used to determine the frequency of the four stimulation tones (f_1-f_4) for HTA-CR neuromodulation (line with dots) and standard CR neuromodulation (dash line). It is assumed that the stimulation tones are perceived ∼10 dB above the individual hearing threshold. A patient with normal hearing ability hears the tones with 10 dB HL whereas a patient with a hearing loss of 40 dB at the tinnitus frequency hears the tones with 50 dB HL. Taking the last case as example, the factors to determine the stimulation tones are 0.72 (tone 1 = f_1), 0.85 (f_2), 1.19 (f_3), and 1.42 (f_4) using the HTA-CR neuromodulation corresponding to a deviation of −5.58% (f_1), −5.51% (f_2), 8.20% (f_3), and 1.18% (f_4) compared to the standard CR neuromodulation factors. These factors are then multiplied with the individual tinnitus frequency to calculate the frequency of the stimulation tones.

Outcome Variables

For all questionnaires the scores were calculated according to the instructions coming with the questionnaire.

To evaluate the therapy outcome TF score change from visit 1 to visit 4 was calculated (primary endpoint). An at least 15-point reduction in TF score during the study period (16 ± 2 weeks) was regarded as success ("Winner," i.e., great responder), a reduction of 6–14 points was defined as an improvement ("Responder"). If a change between −5 and +5 points was observed, the result was classified as unchanged ("Non-responder"), and increases by at least 6 points were judged as deterioration ("Loser," i.e., worsened case).

Statistical Analysis

Scores obtained from questionnaires are summarized as mean ± standard deviation (SD). Changes from baseline (scores before treatment minus scores after treatment) concerning scores for all questionnaires were analyzed by paired Student's *t*-test. To assess potential relationships between continuous variables (age, tinnitus duration, tinnitus severity, and therapy outcome) Pearson's coefficient of correlation (r) was calculated. Response rates were compared between sex groups and between groups of hearing impairment by use of Fisher's exact test for contingency tables. All statistical analyses were performed by M.A.R.C.O. GmbH & Co.KG, Institute for Clinical Research and Statistics (Düsseldorf, Germany) using SAS® version 9.3. A $p < 0.05$ was considered statistically significant.

RESULTS

All patients completed the study. Due to technical reasons (i.e., vacation dates) the duration of the study was slightly longer than originally planned. On average patients used the Desyncra™ for Tinnitus Therapy System for 16 weeks (from visit 1 to visit 4).

At baseline a positive correlation was found between age and duration of the tinnitus ($r = 0.43$, $p = 0.033$), whereas age and tinnitus severity were not correlated ($r = 0.31$, $p = 0.13$).

HTA-CR neuromodulation resulted in a significant reduction of tinnitus and associated discomfort. At the beginning of the study, the mean TF score was 44.9 ± 12.1 (**Table 3**). At the end of the study it was reduced to 32.6 ± 15.5 points,

TABLE 3 | Baseline scores of questionnaires ($n = 25$). NRS-L and NRS-A scores were taken from the TFI questionnaire.

Variable	Mean ± SD (range)
TF score	44.9 ± 12.1 (27–71)
TBF-12 score	13.6 ± 4.7 (5–23)
THI score	52.5 ± 20.9 (10–92)
TFI score	55.1 ± 18.7 (11.2–86.6)
NRS-L score	6.5 ± 2.2 (1–10)
NRS-A score	5.4 ± 2.7 (1–9.5)

corresponding to reduction by 27.4% ($p < 0.001$) compared to baseline (**Figure 2A**). As shown in **Figure 2B**, for 10 out of 25 study participants (40%) the TF score was strongly reduced by more than 15 points, which classified them as "Winners." Moreover, for additional 8 out of 25 patients (32%) a reduction in the range of 6 to 14 points was observed ("Responders"). Thus, the overall response rate ("Winners" plus "Responders") results to 72% in this study. No change of tinnitus was observed in 5 out of 25 patients (20% "Non-responder") treated with HTA-CR neuromodulation while 2 patients (8%) experienced a deterioration of the pre-existing conditions ("Losers").

Interestingly, mean TF scores were higher at the initial examination compared to study start (visit 0: 50.4 ± 12.9 points; visit 1: 44.9 ± 12.1 points; see **Tables 2**, **3**), corresponding to a reduction of 10.9% ($p = 0.002$).

Table 4 shows descriptive statistics relating therapy outcome with baseline characteristics. Response to the therapy (i.e., improvement of TF score) did not seem to depend on age ($r = 0.07$, $p = 0.733$) and tinnitus duration ($r = 0.31$, $p = 0.135$), whereas it appears to increase with tinnitus severity ($r = 0.36$, $p = 0.079$). Response rates were not significantly different between male and female patients (62.5 vs. 88.9%, $p = 0.355$). With regard to the degree of hearing impairment, therapy response was not significantly different between the hearing groups ($p = 0.65$). Both in group 1 and group 3 more than 40% of the patients experienced such an improvement of TF score that they were classified as winners.

The mean score obtained from TBF-12 decreased significantly by 19.7% ($p = 0.006$) from 13.6 ± 4.7 points at baseline to 10.9 ± 5.4 points after 16 ± 2 weeks of HTA-CR neuromodulation therapy (**Figure 3A**). For the THI questionnaire, with a maximum score of 100 points, a significant reduction of 24.5% (visit 1: 52.5 ± 20.9 points; visit 4: 39.6 ± 22.4 points; $p = 0.013$) was observed (**Figure 3B**). Similarly, a significant decrease by 22.1% ($p = 0.003$) from 55.1 ± 18.7 points to 42.9 ± 23.3 points was found for the TFI (**Figure 3C**). To assess tinnitus loudness and annoyance, NRS scores were evaluated. For NRS-L (**Figure 4A**), the baseline value at visit 1 was 6.5 ± 2.2 points and

TABLE 4 | Therapy outcome related to baseline characteristics.

	Winners ($n = 10$)	Responders ($n = 8$)	Non-responders ($n = 5$)	Losers ($n = 2$)
Mean age (range)	53.5 years (40–82)	47.8 years (21–71)	53.6 years (39–77)	48.5 years (37–60)
Median tinnitus duration (range)	8.5 years (1–35)	8.0 years (3–20)	7.0 years (3–18)	6.0 years (1–11)
SEX				
# male ($n = 16$)	7 (43.7%)	3 (18.8%)	4 (25.0%)	2 (12.5%)
# female ($n = 9$)	3 (33.3%)	5 (55.6%)	1 (11.1%)	0 (0.0%)
TINNITUS SEVERITY (# PATIENTS)				
Grade 1 ($n = 2$)	1 (50.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)
Grade 2 ($n = 15$)	5 (33.3%)	5 (33.3%)	3 (20.0%)	2 (13.3%)
Grade 3 ($n = 6$)	3 (50.0%)	1 (16.7%)	2 (33.3%)	0 (0.0%)
Grade 4 ($n = 2$)	1 (50.0%)	1 (50.0%)	0 (0.0%)	0 (0.0%)
HEARING IMPAIRMENT (# PATIENTS)*				
Group 1 ($n = 11$)	5 (45.5%)	4 (36.4%)	2 (18.2%)	0 (0.0%)
Group 2 ($n = 7$)	2 (28.6%)	3 (42.9%)	2 (28.6%)	0 (0.0%)
Group 3 ($n = 7$)	3 (42.9%)	1 (14.3%)	1 (14.3%)	2 (28.6%)

*Refer to ***Table 2***.

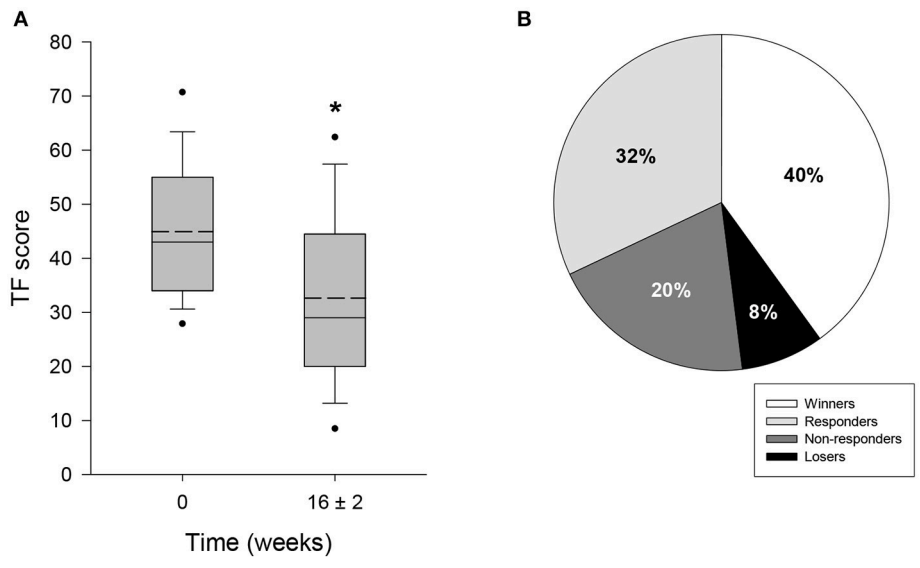


FIGURE 2 | Evaluation of tinnitus by TF questionnaire. 25 patients participated in the study and completed TF at visit 1 (baseline, 0 weeks) and visit 4, i.e., after 16 ± 2 weeks of therapy with HTA-CR neuromodulation. **(A)** Box plots of TF scores showing median (solid line), mean (dash line), 5th and 95th percentile; $n = 25$, *$p < 0.001$ (paired t-test). **(B)** Percentage of patients classified as "Winners" (TF score reduced by ≥15 points), "Responders" (TF score reduced by 6–14 points), "Non-responders" (change between −5 and +5 points), and "Losers" (TF score increased by ≥ 6 points).

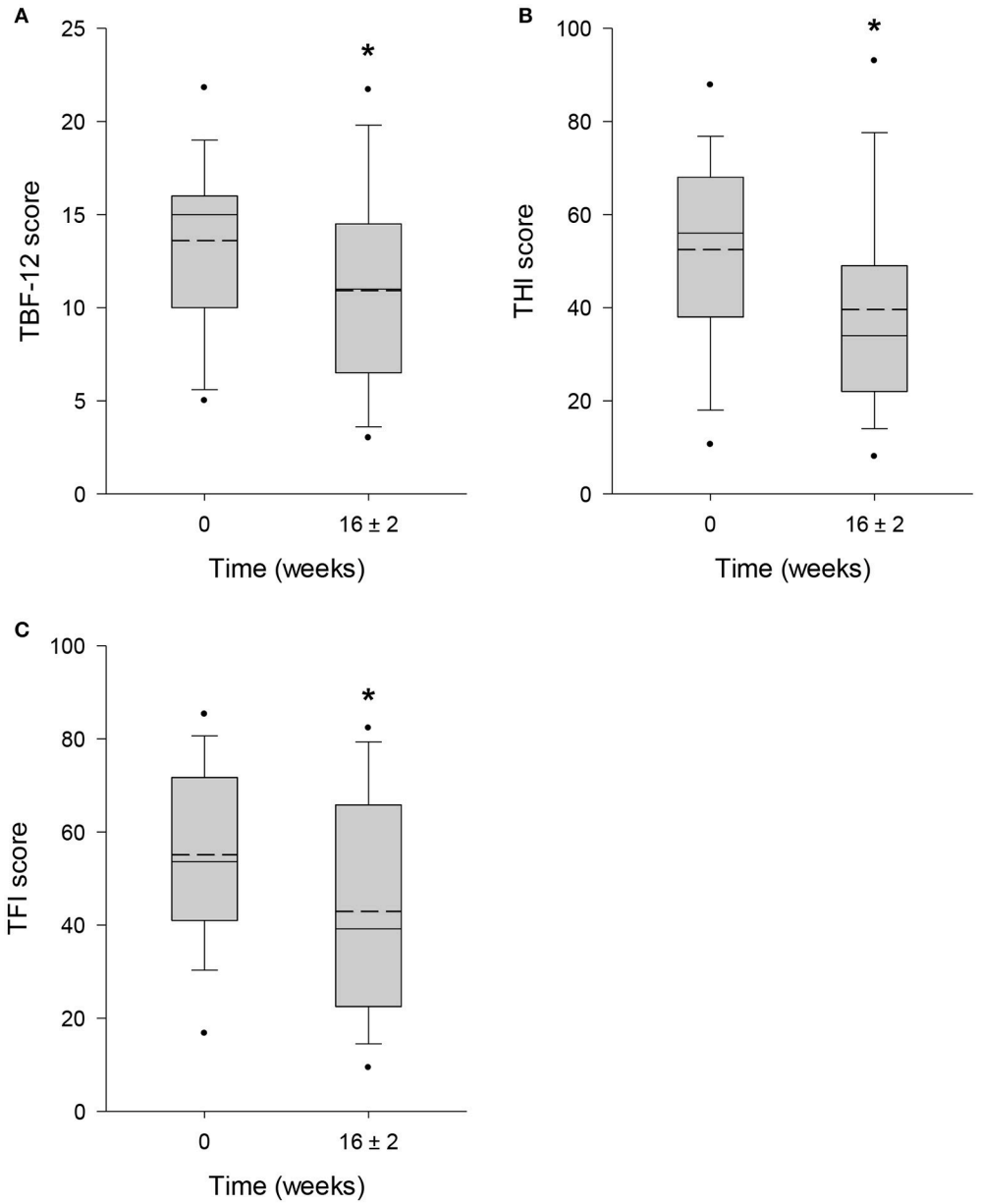


FIGURE 3 | Evaluation of tinnitus by TBF-12 **(A)**, THI **(B)**, and TFI **(C)** questionnaires. All patients completed the questionnaires at visit 1 (baseline, 0 weeks) and visit 4 (16 ± 2 weeks). Results are shown as box plots with median (solid line), mean (dash line), 5th and 95th percentile; $n = 25$, $^*p < 0.05$ (paired t-test).

5.5 ± 2.9 points at visit 4 corresponding to a decrease of 14.9% ($p = 0.022$). NRS-A scores were consistently reduced by 16.7% (visit 1: 5.4 ± 2.7 points; visit 4: 4.5 ± 2.8 points; $p = 0.106$) after 16 ± 2 weeks of HTA-CR neuromodulation therapy (**Figure 4B**).

During the study six adverse events defined as any untoward medical occurrence, and two serious adverse events (SAEs) were reported. Six out of 25 patients reported a temporary worsening of tinnitus symptoms, sensation of head pressure and headache. These side effects were reversible and rapidly disappeared after adjustment of the stimulation or daily stimulation duration. One patient experienced an acute psychosis (anxiety disorder) at the end of the study and was hospitalized for 14 days. Another patient had an unknown cyclothymia and experienced a serious depressive phase during the study resulting in hospitalization for several weeks. Both SAEs were evaluated as not related to the therapy with HTA-CR neuromodulation.

DISCUSSION

This study investigating HTA-CR neuromodulation with use of the Desyncra™ for Tinnitus therapy system is based on an adapted form of the standard CR neuromodulation,

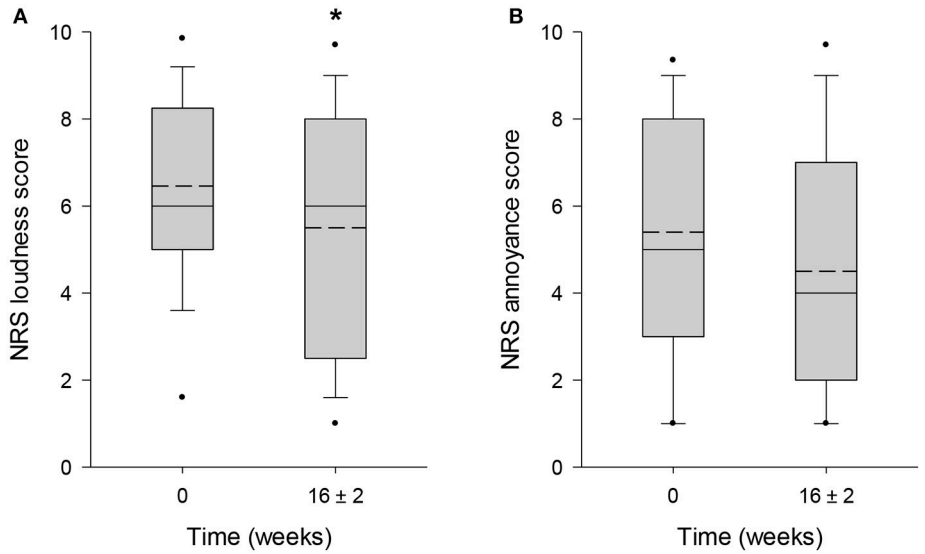


FIGURE 4 | Evaluation of tinnitus by NRS for loudness **(A)** and annoyance **(B)**. Patients completed the scales as part of the TFI questionnaires at visit 1 (baseline, 0 weeks) and visit 4 (16 ± 2 weeks). Results are shown as box plots with median (solid line), mean (dash line), 5th and 95th percentile; $n = 25$; $^*p < 0.05$ (paired t-test).

which was developed through many years of research at the Research Center Juelich, Germany. First results with standard CR neuromodulation from 2012 revealed response rates of 75% in patients with chronic tonal tinnitus (28), which has been confirmed in recent studies comprising meanwhile more than 500 patients (14, 31, 32, Wurzer et al. submitted). A larger double blind study to prove the effectiveness of this method has been carried out in England (RESET2, NCT01541969), but due to serious technical errors the results were not useable (39). Thus, new randomized double blind studies are currently being conducted. Until the results from these studies are available, the effectiveness of CR neuromodulation is critically evaluated in daily clinical practice using non-interventional study approaches (31, 32, Wurzer et al. submitted).

The stimulation tones used for standard CR neuromodulation are determined based on theoretical considerations and experimental results. In an outpatient setting, 67% of patients suffering from tonal tinnitus experienced significant tinnitus improvements after applying this method for 6–12 months (31, Wurzer et al. submitted). However, the altered processing of acoustic stimuli in the whole auditory system associated with sensorineural hearing loss and further experimental results suggest that the stimulation tones might and should be further optimized by adjusting them to the patient's individual tone thresholds. Thus, we applied this idea in our study using the individual stimulation tone intensity and the experimental results obtained by Hopkins and Moore (36). The technical implementation of HTA-CR neuromodulation caused no problem.

The primary objective of the study was to evaluate whether the adapted stimulation with HTA-CR neuromodulation leads to considerable tinnitus improvements. Our results show that despite the small number of patients ($n = 25$) and shorter duration of the study (i.e., ∼4 months) HTA-CR neuromodulation has a remarkable response rate of 72%. A similar response rate was seen with standard CR neuromodulation only in the first trial (28) or after 6 months (31, Wurzer et al. submitted). Moreover, most of the patients reported first subjective changes, e.g., subjective change of tinnitus frequency, already after 2–3 weeks whereas such observations were reported not until 6–8 weeks with standard CR neuromodulation (31, Wurzer et al. submitted).

Unfortunately, the modified stimulation caused a temporary deterioration of tinnitus symptoms in two patients; another control examination after study end revealed that their TF scores were similar to before study start. Based on these first observations it seems reasonable to assume that HTA-CR neuromodulation might induce accelerated changes of tinnitus symptoms as compared to standard CR neuromodulation whereas the overall improvements are similar.

We also reported here a 10.9% improvement of TF score from screening visit 0 to visit 1. This change occurred before therapy start and reflects positive expectations on the therapy outcome. However, most studies do not report detailed data on this "pre-study effect" although it is well-known from clinical experience. Importantly, the effect of HTA-CR neuromodulation reported in our study was calculated as TF score difference between baseline visit 1 (mean TF score: 44.9 ± 12.1 points) and end of study visit 4 (mean TF score: 32.6 ± 15.5 points) resulting in a mean TF score change of 12.3 points. If we consider the TF scores from screening (mean TF score: 50.4 ± 12.9 points), the overall result would even be better resulting in a mean TF score change of 17.8 points and a response rate of 88% (i.e., 60% "Winners" plus 28% "Responders"). Only for 3 of 25

patients (12%) HTA-CR neuromodulation would result in no improvement of their tinnitus symptoms.

In summary, the results of this small feasibility study suggest that the new HTA-CR neuromodulation is at least comparable to the standard CR neuromodulation and might provide another therapeutic option in the treatment of chronic tonal tinnitus.

DATA AVAILABILITY

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

HW planned the study, executed the study and wrote the manuscript together with the second author. CH designed the adapted therapy, planned the study, supported the execution of the study and wrote the manuscript together with the first author.

FUNDING

The study was funded by Desyncra Operating GmbH, Germany.

ACKNOWLEDGMENTS

We thank Prof. Hasso von Wedel and Prof. Brian Moore for useful discussions. Furthermore, we thank the team of Desyncra for the implementation of this novel stimulation algorithm into the Desyncra™ for Tinnitus System. We also thank Petra Reichardt and Nicole Hauptmann for excellent assistance with the individual adaptation of the Desyncra™ for Tinnitus Therapy System for our patients. We thank M.A.R.C.O. GmbH & Co.KG, Institute for Clinical Research and Statistics (Düsseldorf, Germany) for support with statistical analyses and medical writing.

REFERENCES

1. Eggermont JJ, Roberts LE. The neuroscience of tinnitus. *Trends Neurosci.* (2004) 27:676–82. doi: 10.1016/j.tins.2004.08.010
2. Eggermont JJ, Tass PA. Maladaptive neural synchrony in tinnitus: origin and restoration. *Front Neurol.* (2015) 6:29. doi: 10.3389/fneur.2015.00029
3. Eggermont JJ, Kral A. Somatic memory and gain increase as preconditions for tinnitus: insights from congenital deafness. *Hear Res.* (2016) 333:37–48. doi: 10.1016/j.heares.2015.12.018
4. Lee SY, Nam DW, Koo JW, De Ridder D, Vanneste S, Song JJ. No auditory experience, no tinnitus: lessons from subjects with congenital- and acquired single-sided deafness. *Hear Res.* (2017) 354:9–15. doi: 10.1016/j.heares.2017.08.002
5. Roberts LE, Eggermont JJ, Caspary DM, Shore SE, Melcher JR, Kaltenbach JA. Ringing ears: the neuroscience of tinnitus. *J Neurosci.* (2010) 30:14972–9. doi: 10.1523/JNEUROSCI.4028-10.2010
6. Schaette R, Turtle C, Munro KJ. Reversible induction of phantom auditory sensations through simulated unilateral hearing loss. *PLoS ONE* (2012) 7:e35238. doi: 10.1371/journal.pone.0035238
7. De Ridder D, Vanneste S, Langguth B, Llinas R. Thalamocortical dysrhythmia: a theoretical update in tinnitus. *Front Neurol.* (2015) 6:124. doi: 10.3389/fneur.2015.00124
8. Llinás R, Urbano FJ, Leznik E, Ramirez RR, van Marle HJ. Rhythmic and dysrhythmic thalamocortical dynamics: GABA systems and the edge effect. *Trends Neurosci.* (2005) 28:325–33. doi: 10.1016/j.tins.2005.04.006
9. Rauschecker JP, May ES, Maudoux A, Ploner M. Frontostriatal gating of tinnitus and chronic pain. *Trends Cogn Sci.* (2015) 19:567–78. doi: 10.1016/j.tics.2015.08.002
10. Langguth B. Tinnitus: the end of therapeutic nihilism. *Lancet* (2012) 379:1926–8. doi: 10.1016/S0140-6736(12)60561-3
11. Dohrmann K, Elbert T, Schlee W, Weisz N. Tuning the tinnitus percept by modification of synchronous brain activity. *Restor Neurol Neurosci.* (2007) 25:371–8.
12. Weisz N, Moratti S, Meinzer M, Dohrmann K, Elbert T. Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography. *PLoS Med.* (2005) 2:e153. doi: 10.1371/journal.pmed.0020153
13. Song JJ, De Ridder D, Van de Heyning P, Vanneste S. Mapping tinnitus-related brain activation: an activation-likelihood estimation metaanalysis of PET studies. *J Nucl Med.* (2012) 53:1550–7. doi: 10.2967/jnumed.112.102939
14. Adamchic I, Hauptmann C, Tass PA. Changes of oscillatory activity in pitch processing network and related tinnitus relief induced by acoustic CR neuromodulation. *Front Syst Neurosci.* (2012) 6:18. doi: 10.3389/fnsys.2012.00018
15. Martinez-Devesa P, Perera R, Theodoulou M, Waddell A. Cognitive behavioural therapy for tinnitus. *Cochrane Database Syst Rev.* (2010):CD005233. doi: 10.1002/14651858.CD005233.pub3
16. Jastreboff PJ. Tinnitus retraining therapy. *Prog Brain Res.* (2007) 166:415–23. doi: 10.1016/S0079-6123(07)66040-3
17. Minen MT, Camprodon J, Nehme R, Chemali Z. The neuropsychiatry of tinnitus: a circuit-based approach to the causes and treatments available. *J Neurol Neurosurg Psychiatr.* (2014) 85:1138–44. doi: 10.1136/jnnp-2013-307339
18. Zenner HP, Delb W, Kroner-Herwig B, Jager B, Peroz I, Hesse G, et al. A multidisciplinary systematic review of the treatment for chronic idiopathic tinnitus. *Eur Arch Otorhinolaryngol.* (2017) 274:2079–91. doi: 10.1007/s00405-016-4401-y
19. Cima RF, Maes IH, Joore MA, Scheyen DJ, El Refaie A, Baguley DM, et al. Specialised treatment based on cognitive behaviour therapy versus usual care for tinnitus: a randomised controlled trial. *Lancet* (2012) 379:1951–9. doi: 10.1016/S0140-6736(12)60469-3
20. Argstatter H, Krick C, Plinkert P, Bolay HV. [Music therapy for noisiform tinnitus. Concept development and evaluation]. *HNO* (2010) 58:1085–93. doi: 10.1007/s00106-010-2113-y
21. Cramer A. *Wirksame Selbsthilfe mit Musiktherapie*. Stuttgart: Georg Thieme Verlag (2012).
22. Hoare DJ, Kowalkowski VL, Kang S, Hall DA. Systematic review and meta-analyses of randomized controlled trials examining tinnitus management. *Laryngoscope* (2011) 121:1555–64. doi: 10.1002/lary.21825
23. Okamoto H, Stracke H, Stoll W, Pantev C. Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proc Natl Acad Sci USA.* (2010) 107:1207–10. doi: 10.1073/pnas.0911268107
24. Stein A, Wunderlich R, Lau P, Engell A, Wollbrink A, Shaykevich A, et al. Clinical trial on tonal tinnitus with tailor-made notched music training. *BMC Neurol.* (2016) 16:38. doi: 10.1186/s12883-016-0558-7
25. Tass PA. A model of desynchronizing deep brain stimulation with a demand-controlled coordinated reset of neural subpopulations. *Biol Cybern.* (2003) 89:81–8. doi: 10.1007/s00422-003-0425-7
26. Tass PA, Majtanik M. Long-term anti-kindling effects of desynchronizing brain stimulation: a theoretical study. *Biol Cybern.* (2006) 94:58–66. doi: 10.1007/s00422-005-0028-6

27. Tass PA, Popovych OV. Unlearning tinnitus-related cerebral synchrony with acoustic coordinated reset stimulation: theoretical concept and modelling. *Biol Cybern.* (2012) 106:27–36. doi: 10.1007/s00422-012-0479-5
28. Tass PA, Adamchic I, Freund HJ, von Stackelberg T, Hauptmann C. Counteracting tinnitus by acoustic coordinated reset neuromodulation. *Restor Neurol Neurosci.* (2012) 30:137–59. doi: 10.3233/RNN-2012-110218
29. Hauptmann C, Tass PA. Therapeutic rewiring by means of desynchronizing brain stimulation. *Biosystems* (2007) 89:173–81. doi: 10.1016/j.biosystems.2006.04.015
30. Hauptmann C, Wegener A, Poppe H, Williams M, Popelka G, Tass PA. Validation of a mobile device for acoustic coordinated reset neuromodulation tinnitus therapy. *J Am Acad Audiol.* (2016) 27:720–31. doi: 10.3766/jaaa.15082
31. Hauptmann C, Strobel A, Williams M, Patel N, Wurzer H, von Stackelberg T, et al. Acoustic coordinated reset neuromodulation in a real life patient population with chronic tonal tinnitus. *Biomed Res. Int.* (2015) 2015:569052. doi: 10.1155/2015/569052
32. Williams M, Hauptmann C, Patel N. Acoustic CR neuromodulation therapy for subjective tonal tinnitus: a review of clinical outcomes in an independent audiology practice setting. *Front Neurol.* (2015) 6:54. doi: 10.3389/fneur.2015.00054
33. Llinás RR, Ribary U, Jeanmonod D, Kronberg E, Mitra PP. Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proc Nat Acad Sci USA*. (1999) 96:15222–7. doi: 10.1073/pnas.96.26.15222
34. De Ridder D, Elgoyhen AB, Romo R, Langguth B. Phantom percepts: tinnitus and pain as persisting aversive memory networks. *Proc Natl Acad Sci USA*. (2011) 108:8075–80. doi: 10.1073/pnas.1018466108
35. Adamchic I, Toth T, Hauptmann C, Walger M, Langguth B, Klingmann I, et al. Acute effects and after-effects of acoustic coordinated reset neuromodulation in patients with chronic subjective tinnitus. *Neuroimage Clin.* (2017) 15:541–58. doi: 10.1016/j.nicl.2017.05.017
36. Hopkins K, Moore BC. The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J Acoust Soc Am.* (2011) 130:334–49. doi: 10.1121/1.3585848
37. Hallam RS. *TQ. Manual of the Tinnitus Questionnaire.* London: Psychological Corporation (1996).
38. Feldmann H, Brusis T. *Das Gutachten des Hals-Nasen-Ohren-Arztes*. Stuttgart: Thieme (2012).
39. Haller M, Hall DA. Evaluation of the acoustic coordinated reset (CR(R)) neuromodulation therapy for tinnitus: update on findings and conclusions. *Front Psychol.* (2017) 8:1893. doi: 10.3389/fpsyg.2017.01893
40. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER Jr, et al. Clinical practice guideline: tinnitus. *Otolaryngol Head Neck Surg.* (2014) 151 (Suppl. 2):S1–40. doi: 10.1177/0194599814547475

Conflict of Interest Statement: HW is an ENT physician with a tinnitus clinic in Munich, Germany and also a consultant for Brook Henderson Group (Desyncra Technologies Ltd.) since 2014 (Medical Advisor Germany). CH is employed by Desyncra Operating GmbH, Bad Neuenahr-Ahrweiler (part of Desyncra Technologies Ltd, London, UK) and worked with Juelich Research Center between 2002 and 2016, and has received research funding from the European Community, the Federal Ministry of Education and Research (Germany), the Deutsche Forschungsgemeinschaft, and the Helmholtz Association.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 05 December 2018
doi: 10.3389/fnins.2018.00901

Validating Psychometric Questionnaires Using Experience-Sampling Data: The Case of Nightmare Distress

Stefan Stieger[1,2]* *and Tim Kuhlmann*[2]

[1] Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria, [2] Department of Psychology, Research Methods, Assessment and iScience, University of Konstanz, Konstanz, Germany

Nightmares are a comparatively frequent phenomenon. They are often accompanied by emotional distress and gain clinical relevance when recurrent. To assess how much distress nightmares cause the individual, the Nightmare Distress Questionnaire (NDQ, Belicki, 1992) is probably the most often used measure. However, its validity is still disputed. To analyze the validity of the proposed three NDQ subscales in more detail, we conducted an experience sampling study, gathering data either in real-time or short retrospective timeframes over the course of 22 days twice per day (N = 92 participants). The measurements were implemented via a mobile app using participants' own smartphones. Besides the dream quality, we assessed concepts on a daily basis that past research found to be related to dreams. These included critical life events, alcohol consumption, eating behavior, and well-being. We found that only the subscales "general nightmare distress" and "impact on sleep" showed convergent as well as divergent validity. The validity of the subscale "impact on daily reality perception" is unclear. If at all, this subscale is rather indirectly associated with nightmare distress. Furthermore, all of the NDQ items did not differentiate between a bad dream and a nightmare, which suggests that the NDQ might rather be a measure of negative dreams in general and not nightmares in particular. Based on the present experience sampling design, we propose to advance the validation process by further possibilities, such as an item-level, person-level, and multi-level approach. This approach seems to be especially fruitful for concepts which are not very salient (e.g., laughter), can hardly be remembered retrospectively (e.g., dream content), or are potentially threatened by recall biases (e.g., alcohol consumption).

Keywords: nightmare distress, validation, experience sampling, smartphone, psychometrics, questionnaire

OPEN ACCESS

Edited by:
Thomas Probst, Danube University Krems, Austria

Reviewed by:
Keisuke Takano, Ludwig Maximilian University of Munich, Germany
Philip Samuel Santangelo, Karlsruher Institut für Technologie (KIT), Germany

***Correspondence:**
Stefan Stieger
stefan.stieger@kl.ac.at

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *12 July 2018*
Accepted: *19 November 2018*
Published: *05 December 2018*

Citation:
Stieger S and Kuhlmann T (2018) Validating Psychometric Questionnaires Using Experience-Sampling Data: The Case of Nightmare Distress. Front. Neurosci. 12:901. doi: 10.3389/fnins.2018.00901

INTRODUCTION

Nightmares are frightening dysphoric dream sequences accompanied by feelings of fear, usually leading to the awakening of the dreamer. Interestingly, nightmares are a comparatively frequent phenomenon (Wood and Bootzin, 1990) being more prevalent in children (compared to the elderly; Salvio et al., 1992) and women (compared to men; Cuddy and Belicki, 1992; Levin, 1994).

Epidemiological studies found prevalence rates of 2–6% of the population having even *recurrent* nightmares (more than once a week; Belicki and Belicki, 1982; Janson et al., 1995; Ohayon et al., 1997; Schredl, 2010).

Although nightmares are a phenomenon well known to the general public, they are often studied in connection to psycho-pathological symptoms such as sleep disturbances, PTSD, anxiety, or neuroticism (e.g., Levin, 1994; for a review, see Nielsen and Levin, 2007) to name just a few. Neurophysiological studies have found that in people with a nightmare disorder, the activation in parts of the anterior cingulate cortex and parietal lobule were increased and in the frontal and occipital gyri decreased (Shen et al., 2016). It was also shown that people with frequent idiopathic nightmares had differences in the density of slow and fast spindles compared to a control group (Picard-Deland et al., 2018). These studies using fMRI and EEG suggest that there are neurophysiological differences between people having frequent nightmares and those that do not.

Nightmares are often characterized by considerable emotional distress possibly having a dream function by regulating emotions (Blagrove et al., 2004). In general, there are several models of nightmare production (e.g., psychoanalytic models; personality and evolutionary models; neurobiological models) of which the neurocognitive model – as one of the latest suggestions – seems to integrate most of the principles of past models (Nielsen and Levin, 2007). This model assumes that nightmares are the result of a dysfunction in a network of affective processes. These processes serve as an adaptive function to extinct fear memory. If this fear extinction process is disturbed, nightmares emerge (Nielsen and Levin, 2007).

Besides these models of nightmare genesis, a large corpus of studies focuses on the *consequences* of having nightmares mostly manifested in (psychological) distress. This so-called nightmare distress has been conceptualized differently in the past (e.g., nightmare intensity, nightmare effects, nightmare related symptoms; for a review, see Böckermann et al., 2014). One conceptualization that has been frequently studied is the subjective appraisal of how much distress nightmares cause in the sufferers. To assess these subjective feelings, the Nightmare Distress Questionnaire (NDQ, Belicki, 1992) is probably the most often used questionnaire. It consists of thirteen questions with Likert-type answering options purportedly assessing trait-like distress caused by nightmares.

Although the NDQ has good reliability (Cronbach α) and is frequently used in research, its factor structure and validity is still debated. For example, Martínez et al. (2005) found three subscales in a sample of 162 university students. However, their study only included 12 participants, who reported having nightmares on a weekly basis. A more recent, well-powered study by Böckermann et al. (2014) also analyzed the factor structure of the NDQ by recruiting 213 individuals having one or more nightmares in a typical week. Again, a 3-factor solution was found in a principal component analysis (PCA) with the factors *general nightmare distress*, *impact on sleep*, and *impact on daily reality perception*. Although the authors found convergent as well as divergent validity between the three subscales with other nightmare related concepts (e.g., nightmare frequency, sleep quality, fear, depression) they questioned the reliability and validity of the subscale *impact on daily reality perception*. Because of these shortcomings the authors call into question if this subscale is an integral part of nightmare distress at all (Böckermann et al., 2014).

Although validation studies are important steps in developing new measures or verifying established ones, most of them follow a cross-temporal view, i.e., data are assessed at one particular point in time. For some concepts, this procedure is sufficient. But for concepts that include state aspects, a longitudinal view is necessary (e.g., Reis et al., 2016). In the present case of nightmare distress, related concepts (e.g., sleep quality, dream quality) are mostly assessed retrospectively, i.e., participants should remember how many nightmares (or other kind of dreams such as a nice dream) they had in a particular period (usually a couple of weeks). Meanwhile it is well accepted that these retrospective judgments are often biased (e.g., Schwarz and Sudman, 2012; Monk et al., 2015; Pryss et al., 2018). This especially pertains for events, which are not very salient, i.e., are not strong enough to make it into conscious awareness. Although this may not apply for nightmares because of their disturbing nature accompanied with awakening, sweating, and/or being out of breath, it might apply for other types of dreams, e.g., nice dream, neutral dream. Even for nightmares, retrospective judgments might not be accurate due to long periods that have to be judged. One possible solution are longitudinal designs. The experience sampling method (ESM) with up to several measurements per day offers the opportunity to increase the accuracy within a longitudinal framework (Mehl and Conner, 2012).

Experience sampling offers the possibility to capture participants' everyday life behavior and has the advantage that collected data is more accurate than retrospective self-report data (Conner et al., 2009; Kurtz and Lyubomirsky, 2011; for an example about dream frequency, see Blagrove et al., 2004). Research on nightmare distress already has applied experience sampling designs (e.g., Wood and Bootzin, 1990; Köthe and Pietrowsky, 2001; Blagrove et al., 2004; Lancee and Schrijnemaekers, 2013). However, their usage is still quite rare, despite the potential to advance the field substantially (for a similar recommendation, see Nielsen and Levin, 2007).

Furthermore, using ESM offers ways for the development of measures by advancing the validation and development process by another level – the longitudinal one (for a similar argumentation, see Gillath et al., 2009). Usually, scale development and item selection is done by factor loadings, item difficulty, or stability over time when it comes to traits (usually one or two retests). With ESM designs, we have many more measurement occasions, which offers the possibility to additionally judge the deviations of measurements over time (e.g., Reis et al., 2016). For example, for a measure of state anxiety, a validation step could be to select those items which show large variation, i.e., are capable to assess a large variety of anxiety levels in everyday life. This follows the state logic of changing anxiety levels, validated in a within-person design capturing momentary changes. ESM studies are particularly powerful to investigate these changes. Furthermore, ESM designs offer the possibility to

assess events (e.g., nightmares) alongside the items longitudinally to examine contextual associations (Shiffman et al., 2008). These can be used for the judgment of discriminatory power of items regarding participants' behavior, i.e., predictive validity.

In the present study, we applied such an approach by adding an ESM based level to the cross-sectional assessment via the NDQ. We utilized the longitudinal data to examine the validity of the NDQ. First, we assessed dream quality longitudinally and analyzed whether the items and subscales of the NDQ were able to differentiate between dream qualities. Second, we aggregated longitudinally assessed events and psychological measures to create indicators that are less influenced by recall bias than retrospective judgments. We used these aggregated measures to analyze convergent and discriminant validity of the NDQ. We included variables, which did show some connection to nightmares or bad dreams in the past. For example, it has been shown that nightmares are associated with lower well-being (e.g., Levin and Fireman, 2002), occurrence of life events (e.g., Dunn and Barrett, 1988), alcohol consumption (e.g., Munezawa et al., 2011), and food intake (e.g., Nielsen and Powell, 2015). Third, we analyzed whether the NDQ had predictive validity for dream quality, taking the multilevel structure of the data into account.

MATERIALS AND METHODS

Participants

The sample constitutes a convenience sample from a community in Germany. Research assistants recruited participants by word-of-mouth through friends, relatives, and friends-of-friends resulting in a sample size of $N = 108$. Eight participants only filled in one questionnaire (out of 44 possible ones) during the longitudinal phase and another eight participants failed to fill in the cross-sectional questionnaire. The remaining participants ($N = 92$) were mostly students (93%) with an average age of 22.9 years ($SD = 6.9$, range 17–67). Female participants comprised 71% of the sample (one participant did not disclose his/her sex).

All participants gave written informed consent prior to their participation in accordance with the Declaration of Helsinki, and guidelines of the Department of Psychology, University of Konstanz. Approval by an ethics committee was not necessary because the study did not affect the physical or psychological integrity, the right for privacy, or other personal rights or interests. Data collection was anonymous and no harmful procedures were used. Furthermore, participants were informed that they could withdraw at any time during the study without negative consequences.

Measures

Daily Questionnaire

Participants had to fill in the daily questionnaire two times a day, once in the morning and once in the evening for 22 days. Most variables of interest for the current study were only assessed in the morning. These were critical life events, alcohol consumption, eating behavior, and dream quality. They were reported for the day and night prior to the assessment. The exact wording of the items was: (1) Did you have a critical life event yesterday (yes/no)? accompanied by a short definition what we meant by life event ("A situation or event, which you experienced as disturbing, traumatic, or stressful and which bothered you beyond that situation/event, for example, separation from partner, accident, job conflicts, and so forth."); (2) Did you drink alcohol yesterday? (yes/no); (3) Did you have a feeling of fullness prior to going to bed? (yes/no); (4) How was your dream last night? (nice dream, neutral dream, bad dream without awakening, bad dream with awakening, I cannot remember; see **Table 1**). Well-being was assessed as a state measure at both times of the day (Diener et al., 1999). Participants had to answer the question "How is your current well-being?" [visual analogue scale from 0 = *very bad* to 100 = *very good*]. There were further questions asked, which are not part of this study (e.g., attractiveness, loneliness).

Internet-Based Cross-Sectional Questionnaire After the ESM Part

In the final questionnaire, we assessed sociodemographics (age, sex, occupation), nightmare distress (NDQ) as well as further concepts which are not part of this study (e.g., Extraversion, subclinical Narcissism, Satisfaction with Life).

The Nightmare Distress Questionnaire (NDQ; Belicki, 1992) is a 13-item measure using 5-point Likert-type scales as the response format (10 items with 1 = never to 5 = always; 2 items with 1 = not at all to 5 = a great deal; 1 item with 1 = not at all interested to 5 = extremely interested). The NDQ has been proposed to measure three facets of nightmare distress (Böckermann et al., 2014). These are labeled *general nightmare distress* (NDQ General), *impact on sleep* (NDQ Sleep), and *impact on daily reality perception* (NDQ Daily Reality). We did not instruct participants to consider a specific time frame (e.g., in the last year) for their responses.

E-Diary Procedure

The design of the study followed an experience sampling methodology (ESM; real-time and multiple time point measurements) implementing smartphones. A smartphone app was designed for this project and made freely available through the Google Play Store. Participants could directly download the app anonymously. A back-end server software realized communication with the app as well as the storage of data. When the app was opened for the first time, participants had to provide informed consent and were asked basic demographics once (age, sex, nationality). After this initial stage, the main screen appeared showing the items depending on the time of the day. Before midday the morning items were presented, after midday the evening items. Participants were reminded via text messages or WhatsApp messages to do their ratings. They filled in the items while being in their natural surroundings. The reminders were sent out twice per day for a duration of 22 days. The first daily reminder was sent out during the morning time frame between 8 a.m. and 10 a.m. and the second daily reminder during the evening time frame between 6 p.m. and 9 p.m. The reminders followed a time-contingent sampling approach, meaning they were sent at random times within time frames. The compliance rate was on average 90.3%, i.e., only about 10% of reminders were missed. Missingness for each

measurement occasion was very low ranging from 1.7 to 2.4%. Missingness did not increase or decrease over time as indicated by the correlation between measurement point and percentage of missingness: Spearman $r = -0.09$, $p = 0.56$. After the ESM part of the study, the Internet-based cross-sectional questionnaire was administered. Participation was remunerated by optional entry to a raffle (two gift vouchers for 20€ each) or by course credit (for students). The entire study was run in German.

Statistical Analyses

Our operational definition of a nightmare was a bad dream with awakening (in contrast to a bad dream without awakening). For the item-level analyses, we calculated a generalized linear mixed-effects model (GLMM; Bates et al., 2015). Occasions (level 1) were nested within persons (level 2) and the outcome was the dream quality. Dream quality was transformed into three dummy-coded variables (nice, bad, nightmare). The categories *neutral* and *don't know* did not show substantial differences in nightmare distress and were therefore combined as the reference category. Three logistic GLMMs were calculated, testing the predictive value of the NDQ items in distinguishing the three dream qualities from the neutral reference category. The NDQ items were entered as level 2 predictors into the model and grand-mean centered (cgm; Enders and Tofighi, 2007).

Level 1: logit (Dream quality$_{ti}$) = $\pi_{0i} + e_{ti}$
Level 2: $\pi_{0i} = \beta_{00} + \beta_{01}$ respective NDQ item.cgm$_i$ + r_{0i}

For the person-level analyses, we aggregated the dataset on the person level. For continuous level 1 variables, this resulted in means and mean squared successive differences (MSSD; Ebner-Priemer et al., 2009). For dichotomous level 1 data, frequencies were calculated. MSSDs have the advantage of reflecting the deviation of values across time more accurately than classical standard deviations, because they consider the time sequence. Instead of just using the deviation from the mean, the deviation of a certain value from the preceding value in the time sequence is calculated, incorporating information from the time sequence format.

For the multi-level (person and occasion) analyses, we calculated a GLMM for dichotomous data in line with the item-based analyses. The level 1 predictors were current well-being, alcohol consumption the day before, life event the day before, and the feeling of fullness before going to sleep. The three subscales of the NDQ were entered as level 2 predictors into the model.

Prior to the analyses we person-mean centered all level 1 variables and grand-mean centered all level 2 variables (Enders and Tofighi, 2007). Following the recommended procedure by Curran and Bauer (2011), we reintroduced the person-mean from level 1 centering at level 2. The person-mean centered variable then represents fluctuating state aspects whereas the person-mean itself represents stable trait aspects (cwc = values centered within context, i.e., around each participant's mean; pm = person mean; cgm = centered grand-mean). For the final analysis, we used the following model for each of the three dummy coded dream qualities:

Level 1: logit (Dream quality$_{ti}$) = $\pi_{0i} + \pi_{1i}$ Well-being.cwc$_{ti}$ + e_{ti}
Level 2: $\pi_{0i} = \beta_{00} + \beta_{01}$ Well-being.pm$_i$ + β_{02} NDQ General.cgm$_i$ + β_{03} NDQ Sleep.cgm$_i$ + β_{04} NDQ Daytime Reality.cgm$_i$ + r_{0i}
Level 2: $\pi_{1i} = \beta_{10} + r_{1i}$

RESULTS

To judge data quality, we asked for participants' sex and age at the beginning of the ESM part of the study (assessed via the smartphone app) as well as at the end of the study, 22 days later in the final online questionnaire. Participants' sex corresponded to 100% and age to 99%. Only one participant diverged with a difference of 7 years. Because the rest of the data from this participant was not suspect, we retained this participant in the data set.

The NDQ had satisfactory reliability (Cronbach $\alpha = 0.89$). The subscales suggested by Böckermann et al. (2014) also elicited good to acceptable reliability scores: General distress (NDQ General; 5 items): $\alpha = 0.86$; Impact on sleep (NDQ Sleep; 3 items): $\alpha = 0.66$; Impact on daytime reality perception (NDQ Daytime Reality; 4 items): $\alpha = 0.73$. In contrast to Böckermann et al. (2014), the impact on daytime reality perception subscale had in our case acceptable reliability (Böckermann et al., 2014; $\alpha = 0.51$).

Although our data potentially allowed to differentiate between whether the nightmare was post-traumatic (i.e., due to a life-event) or idiopathic (no known cause), the number of nightmares after a life event was just $n = 15$. Therefore, we did not separate between those two types due to power reasons.

In general, only two participants could not remember any dream at all. All the other participants could remember up to every dream during the 22-day time frame (for dream frequency, see **Table 1**). On average, 10.2 dreams were recalled ($SD = 5.3$). The prevalence rate of recurrent nightmares (more than once a week) was 5% in our sample, which is very much in line with past research (e.g., Schredl, 2010). Furthermore, the correlation between nightmare distress and nightmare frequency was small to moderate ($rs = 0.13$–0.27, see **Table 2**) again in line with past research (e.g., Belicki, 1992). Participants who had at least one nightmare during the study phase ($n = 48$) reported an average of 2.1 nightmares within the 22 days ($SD = 1.16$, range 1–6). In general, we found no sex-specific effects regarding dream frequency [nightmares: $t(89) = 1.09$, $p = 0.278$, $d = 0.25$; nice dream: $t(89) = -1.19$, $p = 0.239$, $d = -0.28$; neutral dream: $t(89) = 0.97$, $p = 0.337$, $d = -0.22$] except for bad dreams. Women had a higher frequency of bad dreams compared to men, 2.3 vs. 1.3, respectively [$t(89) = 2.07$, $p = 0.042$, $d = 0.48$]. Furthermore, we found no age-specific effects regarding the frequency of the different dream qualities (all $rs > -0.132$, all $ps > 0.213$).

Item-Level Analyses

To analyze if the items of the NDQ were associated with the occurrence of the different dream qualities (nice, bad, nightmare), we calculated GLMMs for each NDQ item. If

TABLE 1 | Descriptives of variables under investigation.

Dream quality	Dream frequency (%)
Nice dream	243 (14.4)
Neutral dream	394 (23.4)
Bad dream without awakening	176 (10.4)
Nightmare	103 (6.1)
Don't know	769 (45.6)
Sum	1685 (100%)
Number of days with a critical life events	*N* of participants (%)
0	35 (38.9)
1	20 (22.2)
2	13 (14.4)
3	9 (10.0)
4	5 (5.6)
>4 (max = 9)	8 (8.8)
Sum	90 (100.0)
Number of days where alcohol was consumed	*N* of participants (%)
0	6 (6.7)
1	8 (8.9)
2	5 (5.6)
3	5 (5.6)
4	9 (10.0)
5	12 (13.3)
6	8 (8.9)
7	7 (7.8)
>7 (max = 17)	30 (33.5)
Sum	90 (100.0)
Number of days with food intake before sleep	*N* of participants (%)
0	17 (18.9)
1	16 (17.8)
2	7 (7.8)
3	12 (13.3)
4	8 (8.9)
5	4 (4.4)
6	7 (7.8)
7	6 (6.7)
> 7 (max = 19)	13 (14.3)
Sum	90 (100.0)

an NDQ items measures distress regarding nightmares, then it should show a positive association with nightmares, no association with bad dreams, and a negative (or null) association with nice dreams. As can be seen from **Table 3**, none of the NDQ items only showed associations with nightmare dreams without showing an association with bad dreams as well.

Furthermore, only one of the items (#9) of the NDQ Daytime Reality subscale showed an association with the occurrence of nightmares. All other items of the this subscale failed to show a significant association with nightmares. Counterintuitively, all Daytime Reality subscale items showed *positive* associations with nice dreams, though none was significant.This inconclusive pattern regarding the NDQ subscale Daytime Reality is in line with Böckermann et al. (2014) who stated that this subscale has probably little to do with the occurrence of nightmares.

All items of the NDQ General subscale showed a consistent association with nightmare dreams (see **Table 3**). However, the correlation for item #13 only approached significance, $p < 0.10$. This is also in line with Martínez et al. (2005) and Böckermann et al. (2014) who found that item 13 was problematic, in their case because of poor communalities. All items of the subscale did also show a consistent pattern of positive associations with bad dreams and negative associations with nice dream occurrence.

For the NDQ Sleep subscale, Item #1 and #2 showed associations with nightmare, but Item #4 actually failed. Furthermore, in line with Böckermann et al. (2014), we found that Item 12 was suspicious because it failed to reveal any association with different dream qualities (see **Table 3**).

Person-Level Analyses

Next, we were interested if the NDQ is associated with the frequency of each dream quality as well as other suggested influences on dream quality (e.g., food intake before sleep, alcohol consumption, life events; for descriptives, see **Table 1**). If the NDQ has construct validity, then it should correlate with nightmares and bad dream frequency (positive correlations) as well as mean well-being (negative correlations; Blagrove et al., 2004). Furthermore, we should also find a higher fluctuation of well-being scores (represented by MSSD) due to nightmare distress. Regarding potential daytime influences, nightmare distress should be associated with high food intake before sleep, alcohol consumption, and the occurrence of life events. Intercorrelations of these variables are shown in **Table 2**.

NDQ subscales showed significant and substantial inter-correlations (see Böckermann et al., 2014). The frequency of different dream qualities was unrelated except for a positive correlation between nightmare and bad dream frequency that almost reached statistical significance ($r = 0.20$, $p < 0.10$). Interestingly, frequently having nice dreams does not lower the probability of having a bad dream or nightmare. This supports the assumption that dreams are independent from each other with regard to their quality.

Nightmare frequency was unrelated to trait- and state-levels of well-being, which was surprising. For nice and bad dream frequency, we found significant correlations in the expected directions (nice dreams were *positively* associated with trait well-being, bad dreams *negatively* associated with trait well-being).

Regarding construct validity NDQ General and NDQ Sleep showed convergent as well as discriminant validity by being positively correlated with nightmare and bad dream frequency, and negatively with the mean well-being during the 3-week time frame of data collection. NDQ General and NDQ Sleep were not significantly correlated with nice dream frequency or fluctuation of well-being over time. Descriptively, they did show the expected associations, though. NDQ Daytime Reality failed to show any significant correlations.

To sum up, we found construct validity for NDQ General and NDQ Sleep but not for NDQ Daytime Reality. Furthermore, we found that the NDQ was unable to differentiate between nightmare and bad dream frequency.

TABLE 2 | Results of the person-level analyses (Spearman correlations).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.
1. NDQ general											
2. NDQ sleep	0.64***										
3. NDQ daytime reality	0.61***	0.41***									
4. Nightmare frequency	0.26*	0.27**	0.13								
5. Nice dream frequency	−0.15	−0.05	0.13	0.09							
6. Bad dream frequency	0.33**	0.26*	0.17	0.20†	−0.17						
7. Neutral dream frequency	0.06	0.01	−0.01	0.11	0.01	0.25*					
8. Mean well-being	−0.30**	−0.36***	−0.13	−0.09	0.33**	−0.27**	−0.10				
9. MSSD well-being	0.12	0.19†	0.15	0.17	0.19†	0.27**	−0.04	−0.15			
10. Life event frequency	0.14	0.20†	0.28**	0.01	0.16	0.19†	−0.06	−0.16	0.26*		
11. Alcohol frequency	0.06	0.03	0.03	0.04	0.10	−0.09	0.08	0.17	0.07	−0.12	
12. Food intake frequency	0.05	0.20†	0.29**	−0.06	0.16	0.04	−0.01	−0.06	0.27**	0.13	0.09

*N = 92, ***p < 0.001, **p < 0.01, *p < 0.05, †p < 0.10. NDQ, Nightmare Distress Questionnaire; MSSD, Mean Squared Successive Differences.*

TABLE 3 | Results of the item-based analyses.

	Estimate of the fixed effect coefficient β_{01}		
	Nice dream	Bad dream	Nightmare
NDQ General distress subscale (NDQ General)			
Item 5	−0.27	0.56***	0.49***
Item 6	−0.27	0.37**	0.43**
Item 7	−0.20	0.48***	0.34*
Item 8	−0.18	0.40**	0.46**
Item 13	−0.21	0.29*	0.30†
NDQ Impact on sleep subscale (NDQ Sleep)			
Item 1	0.02	0.31*	0.49**
Item 3	−0.26	0.49**	0.46*
Item 4	−0.08	0.24†	0.23
NDQ Impact on daytime reality perception subscale (NDQ Daytime reality)			
Item 2	0.08	0.32*	0.23
Item 9	0.15	0.25*	0.33*
Item 10	0.15	0.09	0.07
Item 11	0.25	0.12	0.21
Excluded by Böckermann et al., 2014			
Item 12	−0.41	0.28†	0.35†

****p < 0.001, **p < 0.01, *p < 0.05, †p < 0.10. NDQ, Nightmare Distress Questionnaire.*

Multi-Level (Person and Occasion) Analyses

In a further step, we wanted to know if there is an association of the three subscales of the NDQ with the probability of having a certain type of dream. First, alcohol consumption, occurrence of a life event, and feelings of fullness did not show any significant effects on dream quality in any of the analyses (except for a counterintuitive small effect of alcohol consumption on the probability of not having a bad dream) and were therefore discarded to keep the models parsimonious.

As can be seen in **Table 4**, a nice dream was associated with higher well-being the next morning whereas a bad dream and nightmare was associated with significantly lower well-being. Regarding the NDQ, only the NDQ General subscale had any consistent predictive value for dream quality. Higher general nightmare distress was associated with a lower chance for a nice dream, but a higher chance for a bad dream and nightmare (although not significant for a nightmare). NDQ Sleep had no predictive value for any type of dreams and NDQ Daytime Reality

TABLE 4 | Results of the multi-level analyses.

Outcome Predictor	Fixed				Random	
	Coef.	Est.	*SE*	*z*	Coef.	*SD*
Nice dream						
Intercept	β_{00}	−4.38			r_{0i}	1.11
Well-being.cwc	β_{10}	0.02	< 0.01	3.29***	r_{1i}	0.02
Well-being.pm	β_{01}	0.04	0.01	2.37*		
NDQ general	β_{02}	−0.94	0.37	−2.57*		
NDQ sleep	β_{03}	0.39	0.29	1.35		
NDQ daytime reality	β_{04}	0.70	0.29	2.41*		
Bad dream						
Intercept	β_{00}	−1.35			r_{0i}	0.76
Well-being.cwc	β_{10}	−0.02	< 0.01	−2.58**	r_{1i}	0.03
Well-being.pm	β_{01}	−0.01	0.01	−1.26		
NDQ general	β_{02}	0.68	0.23	3.00**		
NDQ sleep	β_{03}	−0.07	0.23	−0.29		
NDQ daytime reality	β_{04}	−0.07	0.23	−0.33		
Nightmare						
Intercept	β_{00}	−3.14			r_{0i}	1.05
Well-being.cwc	β_{10}	−0.04	0.01	−4.40***	r_{1i}	0.02
Well-being.pm	β_{01}	> −0.01	0.01	−0.01		
NDQ general	β_{02}	0.51	0.29	1.75†		
NDQ sleep	β_{03}	0.30	0.28	1.05		
NDQ daytime reality	β_{04}	−0.04	0.28	−0.15		

*Coef., Coefficient from multilevel Equations; Est., Estimate; Well-being.cwc, person-mean centered well-being; Well-being.pm, person mean of well-being reintroduced into the model as level 2 variable. NDQ subscales (level 2) were grand-mean centered. NDQ, Nightmare Distress Questionnaire. ***p < 0.001, **p < 0.01, *p < 0.05, †p < 0.10.*

did show a reversed, counterintuitive value for dream quality, i.e., the higher NDQ Daytime Reality the *higher* was the chance of having a nice dream.

DISCUSSION

In the present methodological study, we analyzed the validity of the NDQ, being one of the most used measures of nightmare distress. To achieve this, we implemented data from an experience sampling design. We assessed dream quality and further related variables over time, investigating the contextual associations as well as their associations with the NDQ. The results can be summarized as follows:

The items from the NDQ General subscale were able to differentiate between dream qualities (negative vs. positive) slightly better (except Item 13) than the items from the NDQ Sleep subscale. Similar to the NDQ Sleep subscale, it showed significant correlations with nightmare and bad dream frequencies, and convergent validity with well-being (only the mean, not the fluctuations over time). Compared to the other NDQ subscales, NDQ General was the best predictor of the different dream qualities in the multi-level view.

The items from the NDQ Sleep subscale were also capable of differentiating between positive and negative dream qualities (except Item #4). The subscale showed significant correlations with nightmare and bad dream frequencies, and convergent validity with well-being (again only the mean, not the fluctuations over time). The subscale was not capable of predicting any kind of dream quality in the multi-level analyses.

Finally, the NDQ Daytime Reality subscale does not seem to be associated with nightmare distress at all. First, the items belonging to that subscale were not clearly capable of differentiating between negative and positive dream qualities in general (except Item #9, but also revealed a counterintuitive positive association with nice dream occurrence). Although this subscale showed substantial correlations with the other two NDQ subscales (NDQ General, NDQ Sleep), it failed to show any significant associations with dream frequencies. Interestingly, this subscale showed significant correlations with the frequency of life events and feeling of fullness frequency, in contrast to the other NDQ subscales.

Nevertheless, because the NDQ General as well as the NDQ Sleep subscale did not show any substantial associations with these variables, it remains unclear if this can be interpreted as a sign of convergent validity for the NDQ Daytime Reality subscale. Furthermore, from the multi-level view, NDQ Daytime Reality had a *positive* effect on the probability of having a nice dream, not, as would have been expected, *negative* dreams (bad dream, nightmare). Although further research is needed here, it seems that NDQ Daytime Reality is probably not an integral concept of nightmare distress (see also Böckermann et al., 2014). If at all, NDQ Daytime Reality might reflect a concept which is indirectly associated with nightmare distress.

To sum up, item-based analyses revealed that the NDQ did not really differentiate between a bad dream and a nightmare. This is supported by the multi-level analyses where the NDQ had similar predictive value for the bad dream and nightmare (descriptively, even higher for the bad dream). Therefore, the NDQ might be rather a measure of *negative* dream distress including bad dreams that are not nightmares.

Furthermore, our analyses suggest perhaps dropping Item 13 from the NDQ General subscale, Item 4 from the NDQ Sleep subscale, as well as dropping the whole NDQ Daytime Reality subscale. In our study, we only found few associations with dream quality and other indicators for this subscale, casting doubt on its validity and usefulness.

Predictors of Negative Dreams

Although past research found associations of negative dreams with well-being (e.g., Levin and Fireman, 2002), occurrence of life events (e.g., Dunn and Barrett, 1988), alcohol consumption (e.g., Munezawa et al., 2011), and food intake (e.g., Nielsen and Powell, 2015), we only found some significant associations with well-being (see **Table 2**). Participants with a higher frequency of nice dreams and lower frequency of bad dreams had higher well-being on average. Bad dream frequency was associated with a higher fluctuation of well-being over time. All the other potential predictors failed to show significant effects. Besides the possibility that indeed H_0 is true, there might be other explanations for these findings. First, alcohol consumption was only assessed in the 3-week time frame, i.e., it might be not representative for the time outside this time frame. Furthermore, the alcohol intake of the night before was assessed the next day. This measure has been shown to be less accurate than real-time assessment (Monk et al., 2015). Second, the definition of life events was very broad, beginning with minor conflicts with the partner to severe life events such as the death of a beloved person. Focusing on severe life events might have shown effects. Third, the effect of food intake onto dreams is in general a rather weak finding mostly of anecdotal origin (Nielsen and Powell, 2015). In our study, we did not find any effects for feelings of fullness, except for NDQ Daytime reality, but its validity is unclear.

Limitations

Although we collected over 3,500 data points from 92 participants, 44 retests, and a mean compliance rate of 90.3%, the design was slightly underpowered to detect small to medium effects (ICC = 0.3, α = 5%, power = 80%, conservative power calculation based on the recommendation by Twisk (2006), (p. 123ff), 80% power reached for correlations larger than 0.16). Nevertheless, convergent validity requires substantial correlations. Therefore, the low power for small effects only reduces the exploratory power of divergent validity where weak to null correlations are expected.

Furthermore, our results are limited by the fact that our sample was relatively young and consists mainly of women. Moreover, we had a non-clinical sample. In practice, the NDQ might be mostly used to screen for distress in patients with nightmares (nightmare diagnoses ICD-10: F51.5). Associations between the NDQ and the prospectively assessed items might be different in actual patients.

Future Directions

It is interesting that the frequencies of the different dream qualities were almost unrelated. This supports the assumption that a certain dream type on a particular day does not influence the occurrence of a certain dream type in the following night (such as dream-lag effects; e.g., Henley-Einion and Blagrove, 2014), i.e., they seem to be isolated events with minimal "spillover" effects, if any. Future research might address this in more detail.

Furthermore, in line with Böckermann et al. (2014), we found that distress was produced not only by nightmares, but also by bad dreams (see item-based analyses in **Table 3**). The awakening, which distinguishes nightmares from bad dreams, did not elicit any differences in distress. This could be due to two reasons: First, the NDQ might really be a global measure of "negative sleep distress." Second, participants who are asked about nightmare distress retrospectively might not be capable to differentiate between distresses elicited by nightmares as compared to bad dreams. Because the questions in the NDQ explicitly focus on the frequency of nightmare-related aspects (e.g., falling to sleep again, negative impact on well-being) and not bad dreams *per se*, we would rather think that the second reasoning is true, i.e., because of the retrospective remembering of nightmare events, participants are not capable to differentiate between bad dreams and nightmares anymore. Future research could investigate the differentiation further to try to discern distress from bad dreams and nightmares.

Because prevalence rates of nightmares in the population are comparatively high, future research could try to assess situation-dependent state-aspects of nightmare distress in the morning after a nightmare took place using an experience sampling design with an event-based sampling procedure. After several of these events, a mean of these NDQ state scale measurements can be calculated which might be a better predictor of nightmare-related aspects than the classical trait-based NDQ (for a similar discussion about the dimensional structure of state- and trait-aspects, see Schimmack et al., 2000).

AUTHOR CONTRIBUTIONS

SS and TK conceived and designed the study, acquired the data, and wrote the manuscript. SS analyzed and interpreted the data.

ACKNOWLEDGMENTS

We thank Sören Gottwals and Johanna Gruzla for their support in conducting this study and Deborah Canales for her useful comments on the present paper.

REFERENCES

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Belicki, D., and Belicki, K. (1982). Nightmares in a university population. *Sleep Res.* 11:116.

Belicki, K. (1992). The relationship of nightmare frequency to nightmare suffering with implications for treatment and research. *Dreaming* 2, 143–148. doi: 10.1037/h0094355

Blagrove, M., Farmer, L., and Williams, E. (2004). The relationship of nightmare frequency and nightmare distress to well-being. *J. Sleep Res.* 13, 129–136. doi: 10.1111/j.1365-2869.2004.00394.x

Böckermann, M., Gieselmann, A., and Pietrowsky, R. (2014). What does nightmare distress mean? Factorial structure and psychometric properties of the nightmare distress questionnaire (NDQ). *Dreaming* 24, 279–289. doi: 10.1037/a0037749

Conner, T. S., Tennen, H., Fleeson, W., and Barrett, L. F. (2009). Experience sampling methods: a modern idiographic approach to personality research. *Soc. Person. Psychol. Compass* 3, 292–313. doi: 10.1111/j.1751-9004.2009.00170.x

Cuddy, M. A., and Belicki, K. (1992). Nightmare frequency and related sleep disturbance as indicators of a history of sexual abuse. *Dreaming* 2, 15–22. doi: 10.1037/h0094344

Curran, P. J., and Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annu. Rev. Psychol.* 62, 583–619. doi: 10.1146/annurev.psych.093008.100356

Diener, E., Suh, E. M., Lucas, R. E., and Smith, H. L. (1999). Subjective well-being: three decades of progress. *Psychol. Bull.* 125, 276–302. doi: 10.1037/0033-2909.125.2.276

Dunn, K., and Barrett, D. (1988). Characteristics of nightmare subjects and their nightmares. *Psychiatr. J. Univ. Ottawa* 13, 91–93.

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., and Trull, T. J. (2009). Analytic strategies for understanding affective (in) stability and other dynamic processes in psychopathology. *J. Abnorm. Psychol.* 118, 195–202. doi: 10.1037/a0014868

Enders, C. K., and Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* 12, 121–138. doi: 10.1037/1082-989X.12.2.121

Gillath, O., Hart, J., Noftle, E. E., and Stockdale, G. D. (2009). Development and validation of a state adult attachment measure (SAAM). *J. Res. Person.* 43, 362–373. doi: 10.1016/j.jrp.2008.12.009

Henley-Einion, J. A., and Blagrove, M. T. (2014). Assessing the day-residue and dream-lag effects using the identification of multiple correspondences between dream reports and waking life diaries. *Dreaming* 24, 71–88. doi: 10.1037/a0036329

Janson, C., Gislason, T., De Backer, J., Plaschke, P., Björnsson, E., Hetta, J., et al. (1995). Prevalence of sleep disturbances among young adults in three European countries. *Sleep* 18, 589–597.

Köthe, M., and Pietrowsky, R. (2001). Behavioral effects of nightmares and their correlations to personality patterns. *Dreaming* 11, 43–52. doi: 10.1023/A:1009468517557

Kurtz, J. L., and Lyubomirsky, S. (2011). "Positive psychology," in *Handbook of Research Methods for Studying Daily Life*, eds M. R. Mehl and T. S. Conner (New York, NY: Guilford), 553–568.

Lancee, J., and Schrijnemaekers, N. C. M. L. (2013). The association between nightmares and daily distress. *Sleep Biol. Rhytms* 11, 14–19. doi: 10.1111/j.1479-8425.2012.00586.x

Levin, R. (1994). Sleep and dreaming characteristics of frequent nightmare subjects in a university population. *Dreaming* 4, 127–137. doi: 10.1037/h0094407

Levin, R., and Fireman, G. (2002). Nightmare prevalence, nightmare distress, and self-reported psychological disturbance. *Sleep* 25, 205–212.

Martínez, M. P., Miró, E., and Arriaza, R. (2005). Evaluation of the distress and effects caused by nightmares: a study of the psychometric properties of the nightmare distress questionnaire and the nightmare effects survey. *Sleep Hypn.* 7, 29–41.

Mehl, M. R., and Conner, T. S. (2012). *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford.

Monk, R. L., Heim, D., Qureshi, A., and Price, A. (2015). "I have no clue what i drunk last night" using smartphone technology to compare in-vivo

and retrospective self-reports of alcohol consumption. *PLoS One* 10:e0126209. doi: 10.1371/journal.pone.0126209

Munezawa, T., Kaneita, Y., Osaki, Y., Kanda, H., Ohtsu, T., Suzuki, H., et al. (2011). Nightmare and sleep paralysis among Japanese adolescents: a nationwide representative survey. *Sleep Med.* 12, 56–64. doi: 10.1016/j.sleep.2010.04.015

Nielsen, T., and Levin, R. (2007). Nightmares: a new neurocognitive model. *Sleep Med. Rev.* 11, 295–310. doi: 10.1016/j.smrv.2007.03.004

Nielsen, T., and Powell, R. A. (2015). Dreams of the rarebit fiend: food and diet as instigators of bizarre and disturbing dreams. *Front. Psychol.* 6:47. doi: 10.3389/fpsyg.2015.00047

Ohayon, M. M., Morselli, P. L., and Guilleminault, C. (1997). Prevalence of nightmares and their relationship to psychopathology and daytime functioning in insomnia subjects. *Sleep* 20, 340–348. doi: 10.1093/sleep/20.5.340

Picard-Deland, C., Carr, M., Paquette, T., Saint-Onge, K., and Nielsen, T. (2018). Sleep spindle and psychopathology characteristics of frequent nightmare recallers. *Sleep Med.* 50, 113–131. doi: 10.1016/j.sleep.2017.10.003

Pryss, R., Probst, T., Schlee, W., Schobel, J., Langguth, B., Neff, P., et al. (2018). Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus – stress associations based on the TrackYourTinnitus mobile platform*. *Int. J. Data Sci. Anal.* in press. doi: 10.1007/s41060-018-0111-4

Reis, D., Arndt, C., Lischetzke, T., and Hoppe, A. (2016). State work engagement and state affect: similar yet distinct concepts. *J. Vocat. Behav.* 93, 1–10. doi: 10.1016/j.jvb.2015.12.004

Salvio, M. A., Wood, J. M., Schwartz, J., and Eichling, P. S. (1992). Nightmare prevalence in the healthy elderly. *Psychol. Aging* 7, 324–325. doi: 10.1037/0882-7974.7.2.324

Schimmack, U., Oishi, S., Diener, E., and Suh, E. M. (2000). Facets of affective experiences: a framework for investigations of trait affect. *Person. Soc. Psychol. Bull.* 26, 655–668. doi: 10.1177/0146167200268002

Schredl, M. (2010). Nightmare frequency and nightmare topics in a representative German sample. *Eur. Arch. Psychiatry Clin. Neurosci.* 260, 565–570. doi: 10.1007/s00406-010-0112-3

Schwarz, N., and Sudman, S. (2012). *Autobiographical Memory and the Validity of Retrospective Reports*. Berlin: Springer Science & Business Media.

Shen, C., Wang, J., Ma, G., Zhu, Q., He, H., Ding, Q., et al. (2016). Waking-hour cerebral activations in nightmare disorder: a resting-state functional magnetic resonance imaging study. *Psychiatry Clin. Neurosci.* 70, 573–581. doi: 10.1111/pcn.12455

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Ann. Rev. Clin. Psychol.* 4, 1–32. doi: 10.1146/annurev.clinpsy.3.022806.091415

Twisk, J. W. R. (2006). *Applied Multilevel Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511610806

Wood, J. M., and Bootzin, R. R. (1990). The prevalence of nightmares and their independence from anxiety. *J. Abnorm. Psychol.* 99, 64–68. doi: 10.1037/0021-843X.99.1.64

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

OPINION
published: 18 December 2018
doi: 10.3389/fnins.2018.00971

Mobile Data Collection: Smart, but Not (Yet) Smart Enough

Alexander Seifert[1]*, *Matthias Hofer*[1,2] *and Mathias Allemand*[1,3]

[1] University Research Priority Program "Dynamics of Healthy Aging, " University of Zurich, Zurich, Switzerland, [2] Department of Communication and Media Research, University of Zurich, Zurich, Switzerland, [3] Department of Psychology, University of Zurich, Zurich, Switzerland

Keywords: mobile data, experience sampling, real-life study, big data, smartphone

BACKGROUND

Mobile data collection with smartphones—which belongs to the methodological family of ambulatory assessment, ecological momentary assessment, and experience sampling—is a method for assessing and tracking people's ongoing thoughts, feelings, behaviors, or physiological processes in daily life using a smartphone (Mehl and Conner, 2012; Miller, 2012; Trull and Ebner-Priemer, 2013; Harari et al., 2016). The primary goal of this method is to collect in-the-moment or close-to-the-moment *active* data (i.e., subjective self-reports) and/or *passive* data (e.g., data collected from smartphone sensors) directly from people in their daily lives. The collection and assessment of such data is possible because smartphones are widely available and come with the computational power and sensors needed to obtain information about their owners' daily lives. Researchers in the fields of social science (e.g., Raento et al., 2009), psychology (e.g., Miller, 2012; Harari et al., 2016), and neuroscience (e.g., Schlee et al., 2016; Ladouce et al., 2017) use smartphones to collect data about personality processes and dynamics (Allemand and Mehl, 2017; Beierle et al., 2018a; Stieger et al., 2018; Zimmermann et al., 2018), daily cognitive behaviors (Aschwanden et al., 2018), social support behaviors (Scholz et al., 2016), momentary thoughts (Demiray et al., 2017), couple interactions (Horn et al., 2018), physical activity (Gruenenfelder-Steiger et al., 2017), and moods and emotions (Erbas et al., 2018).

Using smartphones for data collection provides a snapshot of individuals' everyday perceptions, experiences, and interactions with their environments. The use of mobile devices for the assessment of individuals' daily lives is not a new research method (e.g., Fahrenberg et al., 1996). However, because smartphones have now become so widespread throughout the population, are low in cost, and are equipped with sensor technology and ready for data collection through apps (Miller, 2012; Cartwright, 2016; Harari et al., 2016; Beierle et al., 2018a), we are now living in an interesting time for smart mobile data collection. Despite much progress, based on our experiences and discussions with experts in the field, we see the potential for further development of this method.

SMART MOBILE DATA COLLECTION

Mobile data collection with smartphones is growing rapidly in popularity due to its many advantages. One such advantage is that the findings are ecologically valid because they are collected during people's day-to-day lives and capture behaviors and experiences in real environments outside of research laboratories (Wrzus and Mehl, 2015). *Real-time* reports (i.e., active data) and sensor data (i.e., passive data) are measured in the moment and are therefore less prone to memory bias than are retrospective assessments (Redelmeier and Kahneman, 1996). By capturing real-time data about when and where an action takes place, the method provides important information about the dynamics of real-life patterns (Hektner et al., 2007). A smartphone allows researchers to capture such data by installing random, continuous, or event-based alarms to ask participants for

OPEN ACCESS

Edited by:
Myra Spiliopoulou, Otto-von-Guericke Universität Magdeburg, Germany

Reviewed by:
Henning Peters, Neuroimaging Center (TUM-NIC), Germany
Johannes Schobel, University of Ulm, Germany

***Correspondence:**
Alexander Seifert
alexander.seifert@uzh.ch

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *13 September 2018*
Accepted: *04 December 2018*
Published: *18 December 2018*

Citation:
Seifert A, Hofer M and Allemand M (2018) Mobile Data Collection: Smart, but Not (Yet) Smart Enough. Front. Neurosci. 12:971. doi: 10.3389/fnins.2018.00971

their responses to questions or events during the day. Intensive repeated measurements of one participant capture *within-person* information, which represents the behaviors and experiences of a single individual. In contrast, *between-person* information demonstrates variability between individuals. Collecting within-person information allows for the study of the mechanisms and processes that underlie behavior, and this can be contrary to between-person information (Hamaker, 2012). For example, a study by Stawski et al. (2013) showed that processing speed is important for understanding between-person differences in working memory, whereas attention switching is of greater importance to within-person variations. Therefore, it can be argued that the proper study of the dynamic nature of psychological processes requires repeated observations within individuals (Conner et al., 2009). Smartphones are ideal tools for collecting such data.

Real-life data measurements are also rich in *contextual information,* as mobile data collection allows for the combination of self-reports or observer-reports (i.e., active data) and objective assessments (i.e., passive data) of activities, movements, social interactions, bodily functions, and biological markers, using the sensors that are built into smartphones (Ebner-Priemer et al., 2013). For example, it is possible to collect self-reports (e.g., individuals' feelings of social inclusion) and simultaneously to record acoustic sound clips of conversation to collect the objective patterns of participants' actual proximity to and interaction with others (e.g., Mehl et al., 2001).

Finally, as *measurement devices*, smartphones are both powerful and widespread in the population. This enables data analysis in real time and the opportunity to run machine learning approaches within the devices, allowing for large, individualized, dynamic, and intensive real-life studies (Raento et al., 2009; Bleidorn and Hopwood, 2018). Because most participants already have their own smartphones, an app is the only thing they need to install to participate in a study (Miller, 2012). This gives researchers the opportunity to conduct studies with large samples (Dufau et al., 2011).

SMARTER MOBILE DATA COLLECTION IN THE FUTURE

In our research, we identified some of the challenges accompanying mobile data collection with smartphones. In addition to discerning six challenging areas, we offer some suggestions for dealing with these challenges in the future. The first challenge relates to collecting data in real-life environments. Collecting smart data in daily life may result in the validation of existing theories, some of which may relate to behaviors and phenomena outside the realm of day-to-day life. However, this requires that researchers develop theories that reflect the multiple factors and dynamics of the real-life context that may influence the individual. Additionally, real-life data should not be collected simply because it is possible to do so, with conclusions about the theoretical significance of the data being drawn afterwards. Instead, we should develop and discuss the potential of real-life theories that consider both the within-person and between-person effects and the real-life context.

The second challenging area relates to real-time measurements. In data collection, *real-time* also means *right on time*; in other words, researchers have to carefully determine whether they are collecting data about the most relevant variables at the most appropriate moments and at ideal time intervals. To do so, they must first know when to collect data and when behaviors, thoughts, or changes are likely to occur. This question is crucial in mobile data collection, because conclusions about fluctuations, variability, and dynamics need to stem from a sound theoretical rationale or from the behavior patterns of the target participant (e.g., Wright and Hopwood, 2016). For instance, smartphone sensor technology and machine learning can help researchers by detecting the time points of events within a participant, by learning when events normally occur, or by learning the dependency of other subjective or objective variables upon events (e.g., Albert et al., 2012).

The third challenging area concerns within-person data. Typical smartphone studies collect data with great fidelity and generate large quantities of observations, placing the approach clearly within the domain of "big data" and requiring its associated advanced analytic techniques (Yarkoni, 2012; Fan et al., 2014). Working with big data requires highly technical expertise that researchers outside the field of computational science do not normally have. Resources must be organized, and after collecting the data, skills in advanced statistical analyses, including longitudinal structural equation modeling (Little, 2013), dynamic structural equation modeling (Asparouhov et al., 2017), multilevel modeling (Bolger and Laurenceau, 2013), and machine learning (e.g., Bleidorn and Hopwood, 2018), are required. As a result, an interdisciplinary research approach involving researchers interested in collecting data with smartphones and experts familiar with those forms of data collection, management, and analysis is crucial. Such endeavors should be supported by funding organizations and academic career programs, enabling the full potential of mobile data collection with smartphones to be achieved.

As a fourth challenging area, we identify the contextual information that can be collected with smartphone sensor data (i.e., passive data), as researchers have to consider the different forms, intervals, and amounts of sensor data (e.g., GPS data, app use, and accelerometer data). When collecting passive data continuously over multiple days, researchers need to consider more than just the data itself; they must also be able to interpret what the measurements indicate and convert the data into psychologically meaningful variables, such as sociability or mobility patterns (e.g., Mehl et al., 2006; Harari et al., 2016). Although this task is fundamental to the research, it often requires new skills of researchers and new approaches within the technology—approaches that ideally automatically aggregate passive smartphone-sensor-based data. For example, when collecting sound files containing conversation, it would be very helpful to automatically detect the spoken words of a target person (e.g., Mehl et al., 2001), detect contextual information (e.g., Lu et al., 2012), or interpret GPS data in terms of mobility patterns (e.g., Ryder et al., 2009). For such requirements, preliminary solutions do exist (e.g., Barry et al., 2006; White et al., 2011), but much more development and validation work is needed before we can achieve automatic, preprocessed, and

validated smartphone-sensor data that can be combined with other types of data collection.

The fifth challenging area relates to the smartphone device itself. Mobile data collection with smartphones requires more technical preparation and greater technical confidence and skills, on the side of both the researcher and participant, than is required in classic paper-and-pencil studies. Daily technical hassles such as malfunctioning software and hardware, low smartphone batteries, and operation systems crashing during ongoing studies cost time and resources. Therefore, we highly recommend including an explicit time buffer and anticipating a higher than usual drop-out rate in smartphone studies to compensate for potential technical problems and challenges (for more information on technical issues, please see Mehl and Conner, 2012; Miller, 2012; Harari et al., 2016). Although the technical side of mobile data collection with smartphones is likely to become more reliable over time, more validation studies are required in this area and more ready-made valid apps are needed. When using smartphones for data collection within specific population groups, it is also important to consider the unique needs of the target group. For example, when working with older adults, it can be helpful to reflect participants' potential lack of smartphone skills by adapting briefings on smartphone/app use (Seifert et al., 2017).

The final, though certainly not least important, challenging area is that of data security and ethical issues. Collecting mobile data has revived past concerns about data protection and the ethical use of data. Using mobile devices for data collection, including tracking behavior and lifestyle patterns, introduces a unique dimension to individual participant protection. When collecting intensive profiles of individuals, which is the main research method within mobile data collection with smartphones, anonymization is nearly impossible. Therefore, traceable real-life data requires an intensive consideration of ethical and legal approval, the safeguarding of participant privacy, and the establishment of data security and data privacy (Harari et al., 2016; Marelli and Testa, 2018). As an example, Beierle et al. (2018b) conceived a privacy model for mobile data collection apps. Zook et al. (2017) present ten simple rules for responsible big data research, concluding that ethical and data protection issues should not prevent research but that it is vital to ensure "that the work is sound, accurate, and maximizes the good while minimizing harm" (Zook et al., 2017, p. 8). When using participants' own smartphones, it is also important that researchers acquire participants' consent to share self-recorded data with researchers (Gustarini et al., 2016). In a quantitative population survey among persons over 50 years of age, Seifert et al. (2018) found that more than the half of this demographic group is willing to share self-recorded data with researchers, regardless of participants' age, gender, education, technology affinity, or perceived health. The sharing and use of participants' own self-recorded data may require new models of participant involvement, with the goal of creating a trusted relationship between the data providers and researchers working with the data (Beierle et al., 2018b; Seifert et al., 2018).

CONCLUSIONS

Mobile data collection with smartphones offers unique and innovative opportunities for studying human beings and processes in real life and real time. This approach offers researchers the opportunity to collect real-time reports of participants in their natural environment and within their individual dynamics and life contexts with the help of a regular smartphone. However, the approach also brings many challenges that provide interesting avenues for future developments. To date, mobile data collection with smartphones is already very smart, but we see the potential for even smarter mobile data collection in the future.

AUTHOR CONTRIBUTIONS

All authors worked on this paper from conception to final approval and share the same opinion.

ACKNOWLEDGMENTS

MH thanks the Swiss National Science Foundation for funding (No. PY00PI_17485/1). MA thanks the UZH Digital Society Initiative (DSI) from the University of Zurich and the Swiss National Science Foundation (No. 162724) for funding.

REFERENCES

Albert, M. V., Kording, K., Herrmann, M., and Jayaraman, A. (2012). Fall classification by machine learning using mobile phones. *PLOS ONE* 7:e36556. doi: 10.1371/journal.pone.0036556

Allemand, M., and Mehl, M. R. (2017). "Personality assessment in daily life: a roadmap for future personality development research," in *Personality Development Across the Lifespan*, ed. J. Specht (London: Elsevier Academic Press), 437–454.

Aschwanden, D., Luchetti, M., and Allemand, M. (2018). Are open and neurotic behaviors related to cognitive behaviors in daily life of older adults? *J. Pers.* doi: 10.1111/jopy.12409. [Epub ahead of print].

Asparouhov, T., Hamaker, E. L., and Muthén, B. (2017). Dynamic latent class analysis. *Struct. Equ. Modeling* 24, 257–269. doi: 10.1080/10705511.2016.1253479

Barry, S. J., Dane, A. D., Morice, A. H., and Walmsley, A. D. (2006). The automatic recognition and counting of cough. *Cough* 2:8. doi: 10.1186/1745-9974-2-8

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., et al. (2018a). "TYDR: Track your daily routine. Android app for tracking smartphone sensor and usage data," in *Proceedings of the 5th International Conference on Mobile Software Engineering and Systems - MOBILESoft '18; 2018 May 27–28* (Gothenburg: ACM Press), 72–75.

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., et al. (2018b). Context data categories and privacy model for mobile data collection apps. *Procedia Comput. Sci.* 134, 18–25. doi: 10.1016/j.procs.2018.07.139

Bleidorn, W., and Hopwood, C. J. (2018). Using machine learning to advance personality assessment and theory. *Pers. Soc. Psychol. Rev.* 1:1088868318772990. doi: 10.1177/1088868318772990

Bolger, N., and Laurenceau, J. P. (2013). *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*. New York, NY: Guilford Press.

Cartwright, J. (2016). Technology: Smartphone science. *Nature* 531, 669–671. doi: 10.1038/nj7596-669a

Conner, T. S., Tennen, H., Fleeson, W., and Barrett, L. F. (2009). Experience sampling methods: a modern idiographic approach to personality research. *Soc. Personal. Psychol. Compass*. 3, 292–313. doi: 10.1111/j.1751-9004.2009.00170.x

Demiray, B., Mischler, M., and Martin, M. (2017). Reminiscence in everyday conversations: a naturalistic observation study of older adults. *J. Gerontol. B Psychol. Sci. Soc. Sci*. doi: 10.1093/geronb/gbx141. [Epub ahead of print].

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., et al. (2011). Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. *PLoS ONE* 6:e24974. doi: 10.1371/journal.pone.0024974

Ebner-Priemer, U. W., Koudela, S., Mutz, G., and Kanning, M. (2013). Interactive multimodal ambulatory monitoring to investigate the association between physical activity and affect. *Front. Psychol*. 3:596. doi: 10.3389/fpsyg.2012.00596

Erbas, Y., Ceulemans, E., Kalokerinos, E. K., Houben, M., Koval, P., Pe, M. L., et al. (2018). Why I don't always know what I'm feeling: the role of stress in within-person fluctuations in emotion differentiation. *J. Pers. Soc. Psychol*. 115, 179–191. doi: 10.1037/pspa0000126

Fahrenberg, J., and Myrtek, M., eds. (1996). *Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies*. Seattle, WA: Hogrefe and Huber.

Fan, J., Han, F., and Liu, H. (2014). Challenges of big data analysis. *Natl. Sci. Rev*. 1, 293–314. doi: 10.1093/nsr/nwt032

Gruenenfelder-Steiger, A. E., Katana, M., Martin, A. A., Aschwanden, D., Koska, J. L., Kündig, Y., et al. (2017). Physical activity and depressive mood in the daily life of older adults. *GeroPsych* 30, 119–129. doi: 10.1024/1662-9647/a000172

Gustarini, M., Wac, K., and Dey, A. K. (2016). Anonymous smartphone data collection: Factors influencing the users' acceptance in mobile crowd sensing. *Pers. Ubiquitous Comput*. 20, 65–82. doi: 10.1007/s00779-015-0898-0

Hamaker, E. L. (2012). "Why researchers should think 'within-person': A paradigmatic rationale," in *Handbook of Research Methods for Studying Daily Life*, eds. M. R. Mehl and T. S. Conner (New York, NY: Guilford), 43–61.

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., and Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspect. Psychol. Sci*. 11, 838–854. doi: 10.1177/1745691616650285

Hektner, J. M., Schmidt, J. A., and Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks, CA: Sage Publications.

Horn, A. B., Samson, A. C., Debrot, A., and Perrez, M. (2018). Positive humor in couples as interpersonal emotion regulation: a dyadic study in everyday life on the mediating role of psychological intimacy. *J. Soc. Pers. Relat*. 1–21. doi: 10.1177/0265407518788197

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., and Ietswaart, M. (2017). Understanding minds in real-world environments: toward a mobile cognition approach. *Front. Hum. Neurosci*. 10:694. doi: 10.3389/fnhum.2016.00694

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: The Guilford Press.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., et al. (2012). "StressSense: detecting stress in unconstrained acoustic environments using smartphones," in: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing – UbiComp '12; 2012 Sept 5–8* (Pittsburgh, PA: ACM Press), 351.

Marelli, L., and Testa, G. (2018). Scrutinizing the EU general data protection regulation. *Science* 360, 496–498. doi: 10.1126/science.aar5419

Mehl, M. R., and Conner, T. S. (eds) (2012). *Handbook of Research Methods for Studying Daily Life*. New York, NY: Guilford Publications, Inc.

Mehl, M. R., Gosling, S. D., and Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *J. Pers. Soc. Psychol*. 90, 862–877. doi: 10.1037/0022-3514.90.5.862

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., and Price, J. H. (2001). The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behav. Res. Methods Instrum. Comput*. 33, 517–523. doi: 10.3758/BF03195410

Miller, G. (2012). The smartphone psychology manifesto. *Perspect. Psychol. Sci*. 7, 221–237. doi: 10.1177/1745691612441215

Raento, M., Oulasvirta, A., and Eagle, N. (2009). Smartphones: an emerging tool for social scientists. *Sociol. Methods Res*. 37, 426–454. doi: 10.1177/0049124108330005

Redelmeier, D. A., and Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain* 66, 3–8. doi: 10.1016/0304-3959(96)02994-6

Ryder, J., Longstaff, B., Reddy, S., and Estrin, D. (2009). "Ambulation: a tool for monitoring mobility patterns over time using mobile phones," in *Proceedings of the 2009 International Conference on Computational Science and Engineering; 2009 Aug 29–31* (Vancouver, BC: IEEE), 927–931. doi: 10.1109/CSE.2009.312

Schlee, W., Pryss, R. C., Probst, T., Schobel, J., Bachmeier, A., Reichert, M., et al. (2016). Measuring the moment-to-moment variability of tinnitus: the trackyourtinnitus smart phone app. *Front. Aging Neurosci*. 8:294. doi: 10.3389/fnagi.2016.00294

Scholz, U., Stadler, G., Ochsner, S., Rackow, P., Hornung, R., and Knoll, N. (2016). Examining the relationship between daily changes in support and smoking around a self-set quit date. *Health Psychol*. 35, 514–517. doi: 10.1037/hea0000286

Seifert, A., Christen, M., and Martin, M. (2018). Willingness of older adults to share mobile health data with researchers. *GeroPsych* 31, 41–49. doi: 10.1024/1662-9647/a000181

Seifert, A., Schlomann, A., Rietz, C., and Schelling, H. R. (2017). The use of mobile devices for physical activity tracking in older adults' everyday life. *Digital Health* 3, 1–12. doi: 10.1177/2055207617740088

Stawski, R. S., Sliwinski, M. J., and Hofer, S. M. (2013). Between-person and within-person associations among processing speed, attention switching, and working memory in younger and older adults. *Exp. Aging. Res*. 39, 194–214. doi: 10.1080/0361073X.2013.761556

Stieger, M., Nißen, M., Rüegger, D., Kowatsch, T., Flückiger, C., and Allemand, M. (2018). PEACH, a smartphone- and conversational agent-based coaching intervention for intentional personality change: study protocol of a randomized, wait-list controlled trial. *BMC Psychol*. 6:43. doi: 10.1186/s40359-018-0257-9

Trull, T. J., and Ebner-Priemer, U. (2013). Ambulatory assessment. *Annu. Rev. Clin. Psychol*. 9, 151–176. doi: 10.1146/annurev-clinpsy-050212-185510

White, J., Thompson, C., Turner, H., Dougherty, B., and Schmidt, D. C. (2011). WreckWatch: Automatic traffic accident detection and notification with smartphones. *Mobile Netw. Appl*. 16, 285–303. doi: 10.1007/s11036-011-0304-8

Wright, A. G., and Hopwood, C. J. (2016). Advancing the assessment of dynamic psychological processes. *Assessment* 23, 399–403. doi: 10.1177/1073191116654760

Wrzus, C., and Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *Eur. J. Pers*. 29, 250–271. doi: 10.1002/per.1986

Yarkoni, T. (2012). Psychoinformatics: new horizons at the interface of the psychological and computing sciences. *Curr. Dir. Psychol. Sci*. 21, 391–397. doi: 10.1177/0963721412457362

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., et al. (2018). Integrating structure and dynamics in personality assessment: first steps toward the development and validation of a personality dynamics Diary. *PsyArXiv*. doi: 10.31234/osf.io/5zcth

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., et al. (2017). Ten simple rules for responsible big data research. *PLoS Comput. Biol*. 13:e1005399. doi: 10.1371/journal.pcbi.1005399

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with the authors.

frontiers in Neuroscience

PERSPECTIVE
published: 18 December 2018
doi: 10.3389/fnins.2018.00945

Feasibility of Linking Molecular Genetic Markers to Real-World Social Network Size Tracked on Smartphones

Rayna Sariyska[1], Eva-Maria Rathner[1], Harald Baumeister[1] and Christian Montag[1,2]**

[1] Institute of Psychology and Education, Ulm University, Ulm, Germany, [2] MOE Key Lab for Neuroinformation, The Clinical Hospital of Chengdu Brain Science Institute, University of Electronic Science and Technology of China, Chengdu, China

The study of individual differences in human social behavior has a long tradition in (personality) psychology focusing on traits such as extraversion linked to vividness and assertiveness. The study of molecular genetic underpinnings of individual differences in social behavior produced many genetic association studies with only few genetic variants, robustly associated with individual differences in personality. One possible reason for non-replication of findings might be the different inventories used to assess human social traits. Moreover, self-report methods to assess personality and social behavior might be problematic due to their susceptibility to different biases such as social desirability or poor abilities in self-reflection. We stress the importance of including recorded behavior to understand the molecular genetic basis of individual differences in personality and linked social traits. We present preliminary data linking oxytocin genetics to individual differences in social network size derived from smartphones. Here, the genetic variation rs2268498, located in the adjacent area of the promoter of the gene coding for the oxytocin receptor (OXTR), was linked to the number of active contacts and incoming calls, tracked on the smartphone for 12 days (note that these results became a bit weaker when age was controlled for). Although the present empirical findings should only be seen as a proof of concept study, this work demonstrates the feasibility to combine molecular genetic variables with real world behavior. If this approach keeps its promises, the field of personality research might experience a boost in psychometric quality in the near future.

Keywords: *Personality Neuroscience*, molecular genetics, oxytocin, oxytocin genetics, extraversion, smartphones, *Psychoinformatics*, digital phenotyping

OPEN ACCESS

Edited by:
Winfried Schlee,
University of Regensburg, Germany

Reviewed by:
Elouise Koops,
University Medical Center Groningen, Netherlands
Brent Winslow,
Design Interactive, United States

****Correspondence:***
Rayna Sariyska
Rayna.Sariyska@uni-ulm.de
Christian Montag
mail@christianmontag.de

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *21 September 2018*
Accepted: *29 November 2018*
Published: *18 December 2018*

Citation:
Sariyska R, Rathner E-M, Baumeister H and Montag C (2018) Feasibility of Linking Molecular Genetic Markers to Real-World Social Network Size Tracked on Smartphones. Front. Neurosci. 12:945. doi: 10.3389/fnins.2018.00945

INTRODUCTION

Disentangling individual differences in personality and intelligence represents an old quest, going back to the days of Sir Francis Galton, who was an early advocate of the use of twin studies (Montag and Hahn, 2018). Currently, abundant research is available demonstrating that individual differences in the mentioned areas are shaped by both nature and nurture. Per rule of thumb about 0.50 on the genetic and 0.50 on the environmental side impact individual differences in human traits as carved out in a large study reviewing 2,748 twin studies published between 1958 and 2012 (Polderman et al., 2015).

A logical next step from this branch of research would be to estimate the heritability of individual differences in a given trait such as personality through the localization of distinct areas on the human genome linked to individual differences in traits such as extraversion or neuroticism. This kind of research has started over 20 years ago being either pursued via the candidate gene approach or genome wide scan association studies (for an overview see Montag and Reuter, 2014). Until today the study of the molecular genetic basis of personality struggles with many problems, perhaps the greatest struggle is to still see only few genetic variants to be robustly associated with personality traits (for a recent overview on genome wide scan studies see Sanchez-Roige et al., 2018). One of the problems clearly has been underpowered, small sample-size studies (see for an overview also Munafò and Flint, 2011). Therefore, a recent attempt is noteworthy, that came up with reproducible gene-personality associations, but needed to include more than 329,000 participants from a United Kingdom biobank to observe 15 SNPs being robustly linked to neuroticism (Luciano et al., 2018). Of note, from our perspective this does not mean that the candidate gene approach is not able to produce robust associations (although others see this differently, e.g., Jern et al., 2017). For example, a highly cited meta-analysis observed that the interaction between the prominent 5-HTTLPR polymorphism and adverse environmental effects on negative emotionality seems to be stable (Karg et al., 2011). Recent work presented a new promising research paradigm in the context of the candidate gene approach, namely linking genetic variations to individual differences in personality in independent samples stemming from different ethnic groups, probably hinting at globally valid effects (Montag et al., 2017b; Sindermann et al., 2018).

In sum, different routes to the study of the molecular genetic basis of personality might ultimately be successful, but without doubt the "hunt" for genetic variants underlying personality is still challenging. This clearly is also due to (a) the polygenetic nature of personality, potentially influenced by several hundreds or even thousands of genetic variations all with small to tiny effects (see also recent advances stressed by Plomin and von Stumm, 2018) and (b) the type of personality assessment used in a respective study. Moreover, often not the same inventories and/or personality assessments are applied, making the comparison of results across genetic association studies even more problematic. In addition, most of the studies assess traits "only" via self-report, ergo problems such as social desirable answers (Van de Mortel, 2008) or not being able to remember previous events correctly (Stone and Shiffman, 2002; Montag et al., 2015a) might bias the data. While the polygenetic nature of personality needs consistent research efforts on a large scale, the limited psychometric quality on which personality research is often based, jeopardizes a whole research area.

Therefore, we aim to present in this short communication *preliminary data* on a new way to assess personality and, thus, to conduct research in the field of *Personality Neuroscience*. We already stress at this point that the presented empirical data of this work should be seen as preliminary, because the sample size is not sufficient to produce a stable outcome. On the other hand collecting the present data took more than one and half years with the recruiting of more than 100 participants providing us with insights into their objectively measured smartphone behavior and molecular genetic variables. Therefore, the present work should be understood as a study testing the feasibility to combine molecular genetic information with real-world behavior, tracked on smartphones, giving insights into individual differences in extraversion-linked smartphone variables.

In earlier works it was demonstrated that in particular call variables (Montag et al., 2014; Stachl et al., 2017), the use of social messengers such as WhatsApp (Montag et al., 2015b), but probably also the here investigated size of a person's network are linked to extraversion (see the importance to assess age in this context, Roberts et al., 2008). The latter assumption is based on existing literature, reporting a link between high extraversion and the number of "friends" and memberships in different groups on Facebook (Ross et al., 2009; Amichai-Hamburger and Vinitzky, 2010). Extraversion itself is a personality trait closely linked to gregariousness, but also assertiveness, to name a few (Costa and McCrae, 1992). As the size of the social network of a person might be linked to the oxytocinergic system (Pearce et al., 2017; for problems with this work see Jern et al., 2017), the present study focused on the investigation of a polymorphism on the oxytocin receptor (OXTR) gene and individual differences in social network size. For the present work we hypothesized that the prosocial TT variant of the OXTR gene, linked to lower autistic traits in both Germany and China (Montag et al., 2017b), higher empathy (Christ et al., 2016), higher abilities in face recognition (Melchers et al., 2013) and processing of social information (Melchers et al., 2015), would be linked to having also a higher number of (active) telephone contacts in the smartphone and more active call behavior.

MATERIALS AND METHODS

Participants

Smartphone and genetic data was available from N = 117 participants (77 females), mostly with a student background. The average age of participants was 23.04 years (SD = 7.32) and 76.9% reported having A level as their highest educational qualification. Most of the participants were recruited at Ulm University and signed an informed consent prior to participation in the study. They received university credits or monetary compensation for their participation. The study was approved by the local ethics committee of Ulm University, Ulm, Germany.

Materials

The application *Insights* (an Android-based smartphone application, developed by Christopher Kannen[1]) was installed on participants' phones either by the examiner or the participants themselves. This application records different variables such as the number of calls per day (incoming, outgoing, missed), the use of different applications such as YouTube or how active a person is (distance per day measured using the GPS function on the phone) etc. In the current study the number of contacts (names

[1]https://www.ckannen.com

saved in the phone book, as well as the total phone numbers) and the average number of contacts one is in touch with per day through calls/actively used contacts per day (referred to as "active contacts"[2]) were used to measure the size of the social network. Furthermore, the average number and duration of calls per day (including the incoming, outgoing and missed calls) were considered as an additional measure. Twelve days of recordings were used to build an average of the tracked variables.

The genotyping was conducted at Ulm University. DNA was extracted from cell material via buccal swabs. DNA purification was conducted by means of the MagNa Pure 96 system (Roche Diagnostics) and genotyping via a Light Cycler Cobas z480 (Roche Diagnostics, real-time quantitative PCR and subsequently high resolution melting; Primer Assays by TibMolBiol) and a mass spectrometer MassARRAY (Agena Bioscience / Sequenom). Participants were genotyped for rs2268498, a functional single nucleotide polymorphism (SNP) on the OXTR gene (Reuter et al., 2017), positioned on chromosome 3p25.

Participants filled in a short version of the Trait-Self Description Inventory (TSDI, for the German version see Olaru et al., 2015), consisting of 42 items rated on a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree). Only the personality characteristic extraversion was used in the analyses of the present study. Cronbach's Alpha was $\alpha = 0.77$.

Statistical Analyses

The distribution of the variables was examined by assessing the skewness and kurtosis of the variables (Miles and Shevlin, 2001). Since all smartphone variables and age deviated from the normal distribution, non-parametric tests were applied. The CC and CT genotypes of rs2268498 were combined to a C+ group and compared to the TT genotype (C− group), according to our hypothesis. Since both age and gender have been linked to recorded smartphone use (see studies by Montag et al., 2014; Stachl et al., 2017), these associations were also tested in the present study. Spearman's correlations were used to examine the link between age and the investigated smartphone variables. Mann–Whitney *U* test was applied to compare male and female participants with respect to the investigated smartphone variables and to assess the association between the rs2268498 genotypes and the smartphone variables. Where there was a need to control for age, the respective dependent variables were normalized using Blom rank-based transformation (Solomon and Sawilowsky, 2009) and an ANCOVA was conducted.

[2]We computed the variable "active contacts" by running through/inspecting the call list of every participant per day. All numbers (from incoming, outgoing, and missed calls) were looked up in the contact list and saved to an active contact list. If there were duplicates, these were deleted. Also different phone numbers of one contact person were summed up to one active contact. At the end a mean over the 12 days of recordings was computed per participant.

RESULTS

The distribution of rs2268498 genotypes did not deviate from the Hardy-Weinberg equilibrium ($X^2 = 0.22$, $p = 0.64$). $N = 27$ participants were CC-carriers, $n = 61$ were carriers of the CT-genotype and $n = 29$ of the TT-genotype. According to our grouping 29 participants (TT or C−) were tested against $n = 88$ C+ carriers (CC + CT).

In **Table 1** the descriptive statistics of the investigated variables are presented (including the median due to the non-normal distribution of the variables).

The correlation analysis demonstrated that the variables *active contacts* (rho = 0.24, $p < 0.01$), *calls count* (rho = 0.23, $p < 0.05$), *outgoing calls* (rho = 0.20, $p < 0.05$), *and incoming calls* (rho = 0.28, $p < 0.01$) were significantly linked to age.

Next, gender differences were tested by means of a Mann–Whitney *U* test. With respect to the smartphone variables, males demonstrated higher values in *active contacts* ($Z = -2.744$, $p < 0.01$), *calls count* ($Z = -3.317$, $p < 0.01$), *incoming calls* (-3.782, $p < 0.01$), *outgoing calls* ($Z = -2.977$, $p < 0.01$) and *call duration in minutes* ($Z = -2.221$, $p < 0.05$).

The results of a Mann–Whitney *U* test demonstrated that the TT-genotype (C- group) was linked to a significantly higher number of *active contacts* ($Z = -2.313$, $p = 0.02$) and significantly higher number of *incoming calls* ($Z = -2.298$, $p = 0.02$) (**Figure 1**). Due to the significant association between those variables and age, an ANCOVA with Blom-transformed variables was conducted where age was included as a covariate. The results with respect to the variable *active contacts* [$F(1,114) = 3.890$, $p = 0.05$] barely missed significance. The same was true for the variable *incoming calls* after age was controlled for [$F(1,114) = 3.428$, $p = 0.07$]. Moreover, no significant interactions between gender and rs2268498 on the smartphone variables *incoming calls* and *active contacts* could be observed, when gender was entered as a second independent variable. However, we point to the fact that searching for a gene by gender interaction is not

TABLE 1 | Descriptive statistics for the investigated variables.

	Total contact names	Total phone numbers	Active contacts mean	Calls count	Incoming calls	Outgoing calls	Missed calls	Call duration min.
Mean	205.06	219.23	1.15	2.15	0.45	1.33	0.36	7.50
Median	180.00	190.00	0.83	1.33	0.25	0.67	0.25	4.10
SD	115.05	124.47	1.13	2.38	0.52	1.72	0.36	9.86
Min.	20	21	0.00	0.00	0.00	0.00	0.00	0.00
Max.	700	773	8.33	15.50	3.58	10.67	1.67	51.24

N = 117, SD = standard deviation; Min. = minimum; Max. = maximum.

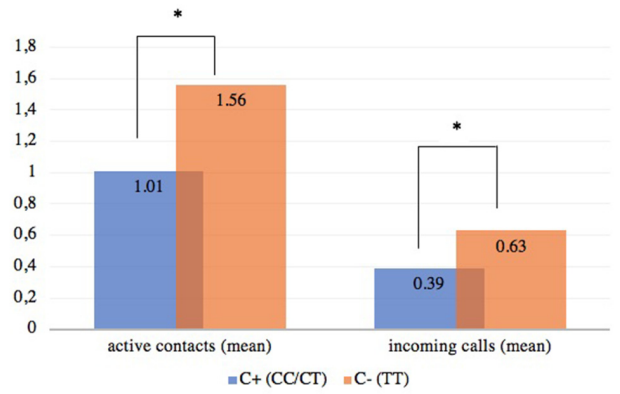


FIGURE 1 | Association between the rs2268498 genotypes and the smartphone variables *active contacts* and *incoming calls*. *N*(C+) = 88, *n*(C–) = 29, *$p < 0.05$. Please note that the depicted significance is derived by the means of the Mann-Whitney *U* test. The means of the variables on the *Y* axis (and not the mean ranks) are presented here for reasons of clarity.

meaningful in our study because the cell sizes were rather small (e.g., the number of male TT-carriers was 13).

Since in total eight smartphone variables were examined, we applied the Bonferroni method as a multiple-comparisons correction. The significance threshold was then $p = 0.05/8 = 0.006$ and none of the previously shown associations remained significant. However, please note that the Bonferroni correction is a very strict correction with the consequence of low power in statistical testing and a number of other disadvantages (we refer to Bender and Lange, 2001; the authors advise to use the Bonferroni correction when the number of tests is less than five). Additionally, since we set up a *directed* hypothesis on the relationship between the smartphone variables and the rs2268498, and did not test a random/large number of smartphone variables, we think that the results from the Bonferroni correction might be too strict and need to be interpreted with caution. In sum, we find it important to report the results of the present study (being in line with a large body of literature on this SNP), but again stress that the present findings should be understood as preliminary. From our perspective, the findings demonstrate the feasibility of linking molecular genetic markers with real-world variables and the here presented findings should be "only" seen as an illustration of this.

Extraversion ($M = 4.37$, $SD = 0.91$) was positively linked to all investigated smartphone variables. The correlations varied between rho = 0.20 and rho = 0.40 ($p < 0.05$). The rs2268498 genotypes and extraversion were not significantly linked.

DISCUSSION

The present work aimed to prove the feasibility to combine molecular genetic information in a meaningful way with real-world behavior, here size of the active social network tracked directly from the smartphone. We do not want to overstress the present results, because our sample is too small to claim general validity of our findings. Aside from this, the observed genetic association fits very well with the literature, again demonstrating that the TT-variant of rs2268498 is linked to higher prosocial behavior/prosocial abilities (Melchers et al., 2015, 2017), here in the light of a larger social network mirrored in the variable number of *active contacts*. Note that other genetic variations of the OXTR gene have been also investigated in the context of social neuroscience (Ebstein et al., 2012; Kumsta and Heinrichs, 2013), therefore other candidates on this gene clearly would have been interesting targets in the realm of the present work. Given that rs2268498 is one of the few, where functionality is likely/has been demonstrated (Reuter et al., 2017) and also in line with the rather straight forward findings so far (as cited), we focused on this single SNP.

Aside from the genetic link to this smartphone variable, the present study reveals several important notes for researchers interested in this new discipline coined *Psychoneuroinformatics* (Montag et al., 2016; see also Yarkoni, 2012; Markowetz et al., 2014 for an introduction into the term *Psychoinformatics*). Of note, other researchers speak in the realm of this new field of *digital phenotyping* (Onnela and Rauch, 2016; Insel, 2017), probably best achieved via methods of *Psychoinformatics*. First of all, an advantage of the present research approach to study the biological underpinnings of personality/sociality traits is the inclusion of information beyond self-report. E.g., if you ask a person how large his or her social network is, you might get biased data. Using smartphone variables such as the present ones gives you an exact estimate also with the advantage that one gets insights into the actual size of both the social network *per se*, but also the *active* social network. The importance to distinguish between these concepts (active vs. passive or complete social network) becomes visible, because in our work an association appeared only between rs2268498 and the active social network size. Using smartphone applications as the present one will also enable researchers to conduct more easily longitudinal research, also in the area of *Personality Neuroscience*.

Although real-world behavior is of great relevance to be included in future neuroscientific works (see evidence for the feasibility to combine MRI data with real-word data in Montag et al., 2017a), several problems arise. First, for the moment it is not that easy to recruit a large number of participants for biologically/neuroscientifically oriented works, investigating individual differences in human behavior, since applications have to be installed on smartphones or related devices together with gathering biomarkers. This naturally limits the inclusion of thousands of participants, as done in the impressive work by Luciano et al. (2018). In particular, in molecular genetics small sample sizes such as the present one represent a problem. In addition, researchers will need to find a standard on how often and how long variables from the Internet of Things need to be tracked to get stable insights into a person's behavior. Or more generally spoken, the psychometric quality of predicting personality by real-world behavior tracked on smartphones need yet to be established. Additionally, more research is needed on the question if and how personality or the situational context might affect one's smartphone use, and, in turn, affect the examined

associations. First studies demonstrated divergent findings on this topic, with some studies reporting a positive link between smartphone use and social engagement (e.g., attending gatherings with friends and colleagues) (Kim et al., 2016), while others demonstrated using mobile smart devices less for online content or social activities when in social situations such as at a restaurant with friends or in an intimate moment with a partner (Vorderer et al., 2016). See also the new work by Dwyer et al. (2018) showing that smartphones reduce enjoyment of face-to-face interactions. Kushlev et al. (2019) even reported that smartphones reduce smiles between strangers. However, please note that several of the here mentioned studies used self-report data, where the answers might be biased through (a) information recall difficulties (e.g., frequency of smartphone use) or (b) social desirability (when participants need to report how often they use their phones in social situations, they might adapt their responses in accordance with social norms). Finally, problems regarding multiple testing arise. A trait such as extraversion impacts on many features of the smartphone. Therefore, it is very difficult to hypothesize on which exact variable on a smartphone the effect of a SNP, best linked to self-reported personality, can be observed. Please note that due to technical reasons ("sandbox principle") it was not possible to take a look at activities inside social network applications such as WhatsApp. Moreover, tracking content in WhatsApp would raise further ethical concerns and might also lower recruitment success given the very intimate nature of one's own content. However, since WhatsApp works with the telephone numbers, saved on our phones, we believe that the variable "contacts" examined in the present study also represents a reasonable approximate of a person's WhatsApp contacts.

In sum, we stress at the end of this article again, that the presented findings should be understood as being illustrative of a new approach to do studies in the field of molecular genetic association studies.

DATA AVAILABILITY STATEMENT

Datasets are available on request: The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

CM, RS, E-MR, and HB designed the present study. CM drafted the first version of the introduction and discussion, whereas RS drafted the first version of the method and result section. Both RS and E-MR were responsible for data-gathering and processing of smartphone variables. All authors worked over the manuscript again and approved its final draft.

FUNDING

The position of CM is funded by a Heisenberg grant awarded to him by the German Research Foundation (DFG, MO 2363/3-2).

REFERENCES

Amichai-Hamburger, Y., and Vinitzky, G. (2010). Social network use and personality. *Comput. Hum. Behav.* 26, 1289–1295. doi: 10.1016/j.chb.2010.03.018

Bender, R., and Lange, S. (2001). Adjusting for multiple testing—when and how? *J. Clin. Epidemiol.* 54, 343–349. doi: 10.1016/S0895-4356(00)00314-0

Christ, C. C., Carlo, G., and Stoltenberg, S. F. (2016). Oxytocin receptor (OXTR) single nucleotide polymorphisms indirectly predict prosocial behavior through perspective taking and empathic concern. *J. Pers.* 84, 204–213. doi: 10.1111/jopy.12152

Costa, P. T., and McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Dwyer, R. J., Kushlev, K., and Dunn, E. W. (2018). Smartphone use undermines enjoyment of face-to-face social interactions. *J. Exp. Soc. Psychol.* 78, 233–239. doi: 10.1016/j.jesp.2017.10.007

Ebstein, R. P., Knafo, A., Mankuta, D., Chew, S. H., and San Lai, P. (2012). The contributions of oxytocin and vasopressin pathway genes to human behavior. *Horm. Behav.* 61, 359–379. doi: 10.1016/j.yhbeh.2011.12.014

Insel, T. R. (2017). Digital phenotyping: technology for a new science of behavior. *JAMA* 318, 1215–1216. doi: 10.1001/jama.2017.11295

Jern, P., Verweij, K. J., Barlow, F. K., and Zietsch, B. P. (2017). Reported associations between receptor genes and human sociality are explained by methodological errors and do not replicate. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9185–E9186. doi: 10.1073/pnas.1710880114

Karg, K., Burmeister, M., Shedden, K., and Sen, S. (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* 68, 444–454. doi: 10.1001/archgenpsychiatry.2010.189

Kim, Y., Wang, Y., and Oh, J. (2016). Digital media use and social engagement: how social media and smartphone use influence social activities of college students. *Cyberpsychol. Behav. Soc. Netw.* 19, 264–269. doi: 10.1089/cyber.2015.0408

Kumsta, R., and Heinrichs, M. (2013). Oxytocin, stress and social behavior: neurogenetics of the human oxytocin system. *Curr. Opin. Neurobiol.* 23, 11–16. doi: 10.1016/j.conb.2012.09.004

Kushlev, K., Hunter, J. F., Proulx, J., Pressman, S. D., and Dunn, E. (2019). Smartphones reduce smiles between strangers. *Comput. Hum. Behav.* 91, 12–16. doi: 10.1016/j.chb.2018.09.023

Luciano, M., Hagenaars, S. P., Davies, G., Hill, W. D., Clarke, T. K., Shirali, M., et al. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* 50, 6–11. doi: 10.1038/s41588-017-0013-8

Markowetz, A., Błaszkiewicz, K., Montag, C., Switala, C., and Schlaepfer, T. E. (2014). Psycho-informatics: big data shaping modern psychometrics. *Med. Hypotheses* 82, 405–411. doi: 10.1016/j.mehy.2013.11.030

Melchers, M., Montag, C., Felten, A., and Reuter, M. (2015). The oxytocin receptor gene and social perception. *Soc. Neurosci.* 10, 345–353. doi: 10.1080/17470919.2015.1008646

Melchers, M., Montag, C., Markett, S., Niazy, N., Groß-Bölting, J., Zimmermann, J., et al. (2017). The OXTR gene, implicit learning and social processing: does empathy evolve from perceptual skills for details? *Behav. Brain Res.* 329, 35–40. doi: 10.1016/j.bbr.2017.04.036

Melchers, M., Montag, C., Markett, S., and Reuter, M. (2013). Relationship between oxytocin receptor genotype and recognition of facial emotion. *Behav. Neurosci.* 127:780. doi: 10.1037/a0033748

Miles, J., and Shevlin, M. (2001). *Applying Regression and Correlation: A Guide for Students and Researchers*. Newcastle upon Tyne: Sage.

Montag, C., Błaszkiewicz, K., Lachmann, B., Andone, I., Sariyska, R., Trendafilov, B., et al. (2014). Correlating personality and actual phone usage. *J. Individ. Differ.* 35, 158–165. doi: 10.1027/1614-0001/a000139

Montag, C., Błaszkiewicz, K., Lachmann, B., Sariyska, R., Andone, I., Trendafilov, B., et al. (2015a). Recorded behavior as a valuable resource for diagnostics in mobile phone addiction: evidence from psychoinformatics. *Behav. Sci.* 5, 434–442. doi: 10.3390/bs5040434

Montag, C., Błaszkiewicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., et al. (2015b). Smartphone usage in the 21st century: who is active on WhatsApp? *BMC Res. Notes* 8:331. doi: 10.1186/s13104-015-1280-z

Montag, C., Duke, É., and Markowetz, A. (2016). Toward psychoinformatics: computer science meets psychology. *Comput. Math. Methods Med.* 2016:2983685. doi: 10.1155/2016/2983685

Montag, C., and Hahn, E. (2018) "Nature-nurture debate," in *Encyclopedia of Personality and Individual Differences*, eds V. Zeigler-Hill and T. Shackelford (Heidelberg: Springer-Verlag).

Montag, C., and Reuter, M. (2014). Disentangling the molecular genetic basis of personality: from monoamines to neuropeptides. *Neurosci. Biobehav. Rev.* 43, 228–239. doi: 10.1016/j.neubiorev.2014.04.006

Montag, C., Markowetz, A., Blaszkiewicz, K., Andone, I., Lachmann, B., Sariyska, R., et al. (2017a). Facebook usage on smartphones and gray matter volume of the nucleus accumbens. *Behav. Brain Res.* 329, 221–228. doi: 10.1016/j.bbr.2017.04.035

Montag, C., Sindermann, C., Melchers, M., Jung, S., Luo, R., Becker, B., et al. (2017b). A functional polymorphism of the OXTR gene is associated with autistic traits in Caucasian and Asian populations. *Am. J. Med. Genet. Part B* 174, 808–816. doi: 10.1002/ajmg.b.32596

Munafò, M. R., and Flint, J. (2011). Dissecting the genetic architecture of human personality. *Trends Cogn. Sci.* 15, 395–400. doi: 10.1016/j.tics.2011.07.007

Olaru, G., Witthöft, M., and Wilhelm, O. (2015). Methods matter: testing competing models for designing short-scale big-five assessments. *J. Res. Personality* 59, 56–68. doi: 10.1016/j.jrp.2015.09.001

Onnela, J. P., and Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 41, 1691. doi: 10.1038/npp.2016.7

Pearce, E., Wlodarski, R., Machin, A., and Dunbar, R. I. (2017). Variation in the β-endorphin, oxytocin, and dopamine receptor genes is associated with different dimensions of human sociality. *Proc. Natl. Acad. Sci. U.S.A.* 114, 5300–5305. doi: 10.1073/pnas.1700712114

Plomin, R., and von Stumm, S. (2018). The new genetics of intelligence. *Nat. Rev. Genet.* 19, 148–159. doi: 10.1038/nrg.2017.104

Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* 47, 702–709. doi: 10.1038/ng.3285

Reuter, M., Montag, C., Altmann, S., Bendlow, F., Elger, C., Kirsch, P., et al. (2017). Functional characterization of an oxytocin receptor gene variant (rs2268498) previously associated with social cognition by expression analysis in vitro and in human brain biopsy. *Soc. Neurosci.* 12, 604–611. doi: 10.1080/17470919.2016.1214174

Roberts, S. G., Wilson, R., Fedurek, P., and Dunbar, R. I. M. (2008). Individual differences and personal social network size and structure. *Personal. Individ. Differ.* 44, 954–964. doi: 10.1016/j.paid.2007.10.033

Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., and Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Comput. Hum. Behav.* 25, 578–586. doi: 10.1016/j.chb.2008.12.024

Sanchez-Roige, S., Gray, J. C., MacKillop, J., Chen, C. H., and Palmer, A. A. (2018). The genetics of human personality. *Genes Brain Behav.* 17, e12439. doi: 10.1111/gbb.12439

Sindermann, C., Luo, R., Zhang, Y., Kendrick, K. M., Becker, B., and Montag, C. (2018). The COMT Val158Met polymorphism and reaction to a transgression: findings of genetic associations in both chinese and german samples. *Front. Behav. Neurosci.* 12:148. doi: 10.3389/fnbeh.2018.00148

Solomon, S. R., and Sawilowsky, S. S. (2009). Impact of rank-based normalizing transformations on the accuracy of test scores. *J. Modern Appl. Stat. Methods* 8, 448–462. doi: 10.22237/jmasm/1257034080

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., et al. (2017). Personality traits predict smartphone usage. *Eur. J. Personal.* 31, 701–722. doi: 10.1002/per.2113

Stone, A. A., and Shiffman, S. (2002). Capturing momentary, self-report data: a proposal for reporting guidelines. *Ann. Behav. Med.* 24, 236–243. doi: 10.1207/S15324796ABM2403_09

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Aust. J. Adv. Nursing* 25, 40.

Vorderer, P., Krömer, N., and Schneider, F. M. (2016). Permanently online–Permanently connected: explorations into university students' use of social media and mobile smart devices. *Comput. Hum. Behav.* 63, 694–703. doi: 10.1016/j.chb.2016.05.085

Yarkoni, T. (2012). Psychoinformatics: new horizons at the interface of the psychological and computing sciences. *Curr. Dir. Psychol. Sci.* 21, 391–397. doi: 10.1177/0963721412457362

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 21 March 2019
doi: 10.3389/fnins.2019.00248

Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students

Jeremy F. Huckins[1], Alex W. daSilva[1], Rui Wang[2], Weichen Wang[2], Elin L. Hedlund[1], Eilis I. Murphy[1], Richard B. Lopez[1], Courtney Rogers[1], Paul E. Holtzheimer[3,4], William M. Kelley[1], Todd F. Heatherton[1], Dylan D. Wagner[5], James V. Haxby[1] and Andrew T. Campbell[2]*

[1] Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, United States, [2] Department of Computer Science, Dartmouth College, Hanover, NH, United States, [3] National Center for PTSD, White River Junction, VT, United States, [4] Department of Psychiatry, Dartmouth–Hitchcock Medical Center, Lebanon, NH, United States, [5] Department of Psychology, The Ohio State University, Columbus, OH, United States

As smartphone usage has become increasingly prevalent in our society, so have rates of depression, particularly among young adults. Individual differences in smartphone usage patterns have been shown to reflect individual differences in underlying affective processes such as depression (Wang et al., 2018). In the current study, a positive relationship was identified between smartphone screen time (e.g., phone unlock duration) and resting-state functional connectivity (RSFC) between the subgenual cingulate cortex (sgCC), a brain region implicated in depression and antidepressant treatment response, and regions of the ventromedial/orbitofrontal cortex (OFC), such that increased phone usage was related to stronger connectivity between these regions. This cluster was subsequently used to constrain subsequent analyses looking at individual differences in depressive symptoms in the same cohort and observed partial replication in a separate cohort. Similar analyses were subsequently performed on metrics of circadian rhythm consistency showing a negative relationship between connectivity of the sgCC and OFC. The data and analyses presented here provide relatively simplistic preliminary analyses which replicate and provide an initial step in combining functional brain activity and smartphone usage patterns to better understand issues related to mental health. Smartphones are a prevalent part of modern life and the usage of mobile sensing data from smartphones promises to be an important tool for mental health diagnostics and neuroscience research.

Keywords: depression, mental health, smartphone, screen time, fMRI, resting-state, circadian rhythm

OPEN ACCESS

Edited by:
Rüdiger Christoph Pryss, Ulm University, Germany

Reviewed by:
Brent Winslow, Design Interactive, United States
Elouise Koops, University Medical Center Groningen, Netherlands

****Correspondence:***
Jeremy F. Huckins
jeremy.f.huckins@dartmouth.edu

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *16 September 2018*
Accepted: *04 March 2019*
Published: *21 March 2019*

Citation:
Huckins JF, daSilva AW, Wang R, Wang W, Hedlund EL, Murphy EI, Lopez RB, Rogers C, Holtzheimer PE, Kelley WM, Heatherton TF, Wagner DD, Haxby JV and Campbell AT (2019) Fusing Mobile Phone Sensing and Brain Imaging to Assess Depression in College Students. Front. Neurosci. 13:248. doi: 10.3389/fnins.2019.00248

INTRODUCTION

Smartphone usage has become nearly ubiquitous in daily life at a time when depression rates are concurrently rising, particularly among college students. Smartphones contain a variety of sensors that can allow researchers to passively measure various behaviors of the phone's user. Previous research has linked smartphone usage to self-reported depressive symptoms (Matar and Jaalouk, 2017; Twenge et al., 2018; Wang et al., 2018). In parallel, depressive symptoms have been linked to

brain connectivity using resting-state functional connectivity (RSFC) MRI (Greicius et al., 2007). The current manuscript has multiple goals. First, is to provide a proof-of-concept for linking passive mobile smartphone sensing technologies to brain connectivity measures that have also been linked to self-reported depressive symptoms. Second is to replicate these initial findings in a separate cohort. Third, is to identify preliminary links between a key behavior inferred from sensing (e.g., smartphone screen time or circadian rhythm consistency) and brain connectivity metrics. Fourth, is to briefly describe a variety of methods which could be used to combine results across these various data types in the future.

Depression Assessment

Depressive disorders affect over 300 million people worldwide and is currently ranked as the single largest contributor to global disability (Ustün et al., 2004; World Health Organization, 2018). Despite this, the diagnosis of depression has remained largely unchanged; further, a reliable means of identifying individuals at risk of becoming depressed remains absent. Psychology, psychiatry and neuroscience have long relied up self-reported surveys and in-person interviews to measure symptoms, diagnose mental health disorders and identify appropriate treatment strategies (Horwitz et al., 2016). As a result of staggering fiscal and personal costs inflicted at both individual and societal levels, clinicians and researchers set out to redefine the way mental disorders are conceptualized in hopes of creating innovative identification and prevention strategies. The aforementioned aims have been synthesized in a research framework known as RDoC (Research Domain Criteria). RDoC's objective is to incorporate information across all planes of analysis ranging from cellular level data to person level self-report survey data to provide of a holistic picture of mental disorders (NIMH). A core principle within the RDoC framework is the notion that neuroscience will inform future psychiatric classification schemes; in other words, aid in moving toward the establishment of a neural biomarker for depression. Thus, of great importance is understanding the complete range of human behavior (and neurological functioning) from typical to atypical (Insel et al., 2010). The Patient Health Questionnaire (PHQ, with two, four, eight and nine question versions) is a reliable, short survey which has been validated in clinical settings and can be used to assess self-reported symptoms of depression that cause significant impairment and subjective distress (Kroenke et al., 2001, 2009a,b; Cameron et al., 2008), an approach in keeping within the RDoC research framework, seeking to explain individual variance in symptoms across domains, constructs, and units of analysis. Future methods to accurately diagnose depression may hold promise with the inclusion of techniques that capitalize on the passive collection of behavioral data through mobile sensors (e.g., smartphones).

Passive Sensing

Passive sensing using mobile smartphone technology allows for the assessment of daily activities by the smartphone user without continual effort on their part. This increases the frequency with which data can be collected and is less vulnerable to self-report bias, which is often a problem in prompted surveys (Rosenman et al., 2011; Ben-Zeev et al., 2015). Smartphone ownership has increased steadily over the last decade, with over 75% of the United States population owning one (Smith, 2017). In parallel, depression rates have increased over the last decade (Twenge et al., 2018). While it is unlikely that smartphone ownership by itself has prompted increased rates of depression, has perhaps facilitated increased access to and usage of social network platforms (Kross et al., 2013). Prevalence of both smartphone ownership and depression rates are often reported as being higher in college-age students (Eisenberg et al., 2013; Nielsen.com, 2016). Screen time, e.g., the amount of time that the screen is unlocked and being used is a relatively simple metric to calculate that has been previously related to depressive symptoms by multiple groups through either passive sensing or self-reported surveys (Twenge et al., 2018; Wang et al., 2018). Screen time and unlock duration will be used interchangeably henceforth.

Depression has been linked to a variety of metrics available from smartphone sensing applications including amount of stationary time, GPS patterns, phone usage and conversation patterns, among others (Burns et al., 2011; Canzian and Musolesi, 2015; Saeb et al., 2015; Mehrotra et al., 2017; Wang et al., 2018). The higher amplitude circadian rhythms as measured by accelerometer are associated with reduced chances of major depressive disorder and other negative mental health outcomes (Lyall et al., 2018). Saeb et al. (2015) determined that circadian movement (regularity in 24-h patterns), mobility between favorite locations and location variance were all negatively correlated with depressive symptoms, while phone usage was positively correlated with depressive symptoms. Using smartphone passive sensing, distance between locations visited and a routine index, or the reliability of the locations visited on a day-to-day basis were related to depressive symptoms (Canzian and Musolesi, 2015). Links between features such as location category (home, car, office etc.) and depression, with further accuracy in prediction when adding context, such as if the individual is alone, with other people (particularly friends) or current physical exertion status (Burns et al., 2011). Self-reported happiness has been linked to decreased phone usage in the subsequent hour (Mehrotra et al., 2017). While several groups have started to characterize traits linked to depression, phone usage and circadian rhythms are the ones that are most prominent in the current literature.

Resting-State Functional Connectivity

Blood-oxygenation-level dependent (BOLD) functional magnetic resonance imaging (fMRI) is a non-invasive way to study activity in the human brain. Changes in BOLD signal are highly correlated with changes in neuronal activity in the local area, particularly local field potentials (Logothetis et al., 2001). RSFC measures the relationship between the time-courses of different regions, often by using the correlation of the time-series. While connectivity across the whole brain, or "functional connectome" is fairly similar across individuals, there are small individual

differences in connectivity between individuals which can be reliably observed across time. There are a variety of factors which may potentially influence RSFC, including genetics, experiences across the lifetime and current physiological and emotional state (Shehzad et al., 2009; Birn et al., 2013; Patriat et al., 2013; Zuo et al., 2014; Poldrack et al., 2015; Richiardi et al., 2015; Sinclair et al., 2015).

Depression and Neuroimaging

Resting-state functional connectivity has been used successfully to distinguish between healthy controls and depressed individuals, even going so far as to distinguish between subtypes of depressed individuals (Greicius et al., 2007; Berman et al., 2013; Kaiser et al., 2015; Drysdale et al., 2016). Task-based studies of self-referential processing have revealed that the sgCC is preferentially involved in processing valenced self-referential information (Moran et al., 2006; Somerville et al., 2006). Additionally, this region has been associated with antidepressant treatment response, and an area proximal to this has been used as a site of deep-brain stimulation for treatment-resistant depression (Mayberg et al., 2005; Holtzheimer, 2012).

Combing RSFC and Mobile Smartphone Passive-Sensing Technology

There are a wide-variety of approaches that can be taken when combining high-dimensional data from multiple modalities. We wanted to answer the following question: do smartphone sensing features previously identified as being related to depression show correlations with RSFC from a region previously identified to have aberrant connectivity in depressed individuals? A targeted approach was used, selecting screen time with mobile smartphone (e.g., unlock duration), a feature previously shown to be linked to depressive symptoms (Saeb et al., 2015; Twenge et al., 2018; Wang et al., 2018) and a brain area, the subgenual cingulate cortex (sgCC) which has previously been identified as having aberrant RSFC in depressed individuals, and more recently has been used as a target for deep brain stimulation for treatment resistant depression (Mayberg et al., 2005; Greicius et al., 2007; Holtzheimer, 2012). Furthermore, if there are regions identified in the passive-sensing unlock duration analysis and RSFC analysis, do these regions also show similar connectivity patterns when looking at the same correlations with brief surveys of self-reported depressive symptoms (PHQ-2, 4 and 8)? We expect that they would. Alternatively, depression may be a summation of multiple factors and may be better understood by interrogating passive-sensing mobile technology and neuroimaging than self-reported scales. As a secondary analysis, other passive-sensing features similar to those previously reported by other groups to be indicative of depression were explored, specifically, circadian rhythms in both movement and number of locations visited. Keeping within the RDoC matrix, a variety of units of analysis including brain connectivity with fMRI (physiological), passive-sensing of phone usage (behavioral) and both computer-based and phone-based depression scales (self-report) were assessed.

MATERIALS AND METHODS

Study Design

In the current study two separate cohorts of first-year undergraduate students were enrolled and analyzed separately for test-retest comparison. Individuals were enrolled in three study components: neuroimaging, smartphone sensing/EMA and online surveys. Three modified versions of the PHQ-9 were used: PHQ-2/4/8. PHQ-8 is the same as PHQ-9 with the suicide ideation question removed. This question was removed before administration because the survey results are not monitored in real-time. PHQ-4 is a four-question survey which includes two questions from the PHQ-8 and two from the GAD-7 as to assess both depressive and anxiety related symptoms (Kroenke et al., 2009a,b). They are used because of their brief form. They may miss some of the nuances that the other inventories pick up on but have been found to have high internal reliability (Cronbach's Alpha > 0.8) and are correlated with diagnoses of clinically relevant depression (Cameron et al., 2008; Khubchandani et al., 2016). PHQ-2 is used as a super-brief form of the PHQ-8 that is slightly more specific to depressive symptoms by excluding the GAD-related questions (Arroll et al., 2010).

Individuals completed an online survey to assess study eligibility (safe for MRI per Dartmouth Brain Imaging Center guidelines, no contraindications that would lead to MRI signal loss, and owned an Android or iOS smartphone compatible with StudentLife). If an individual was eligible and interested in participating in the study, she or he completed a battery of online surveys, including the PHQ-8 through REDCap (Harris et al., 2009). Individuals were then scanned during the academic term and had the StudentLife application (Wang et al., 2014) installed on their phone at or near the time of scanning. In Cohort 1, StudentLife data was collected from the time of scanning until the end of the term. In Cohort 2, StudentLife data was collected from the time of scanning and data collection is currently ongoing but the data presented here is only from their first term in college.

StudentLife

A smartphone application, StudentLife is used in the current study to collect a variety of data about smartphone usage and mood from participants. The application is installed on a participant's phone (iOS or Android) and collects data from the GPS, microphone, accelerometer and lock/unlock status among others. Data from StudentLife is uploaded to a secure server whenever a participant is both using WiFi and charging their phone, which they were encouraged to do daily. Data from these sensors are processed on the server to create variables that assesses the day-to-day and week-by-week impact of workload on stress, sleep, activity, mood, sociability, mental well-being, and academic performance of students (Wang et al., 2014). The workflow of the current study includes data collected through StudentLife, MRI scanning sessions and self-reported surveys (**Figure 1**). Unlock duration is a measurement of time that the phone is unlocked and the screen is on, calculated as the time between the user unlocking the phone and the user either manually relocking the phone or autolocking due

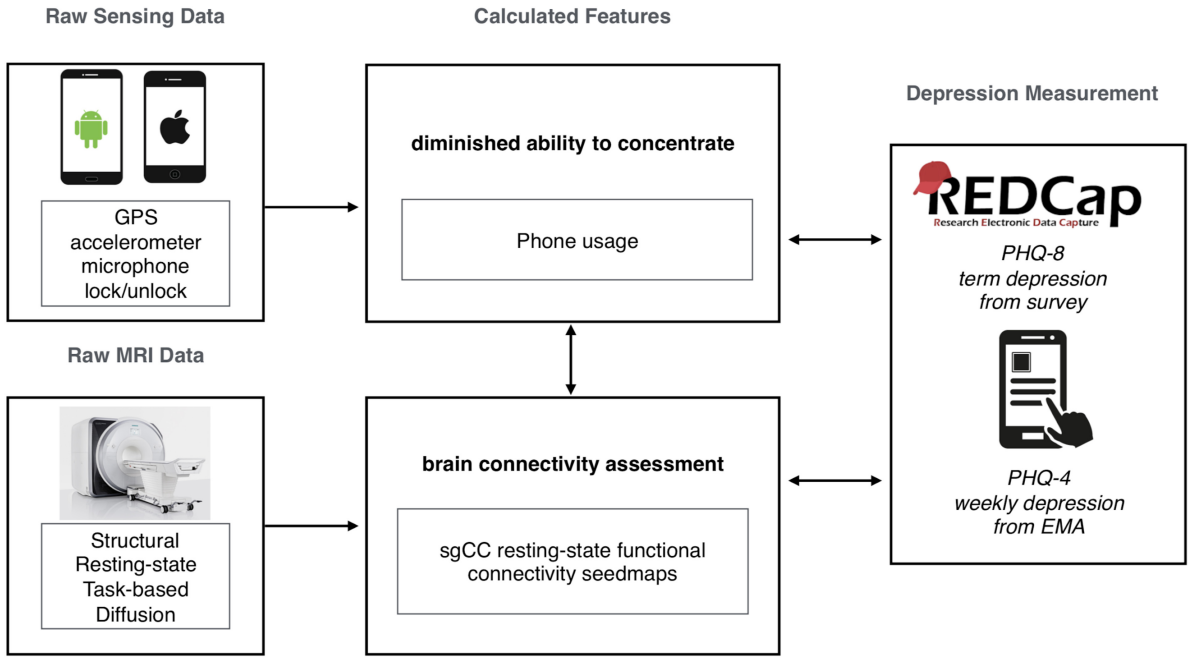


FIGURE 1 | Summary graphic of the study workflow in the current study, showing raw data collection from both smartphones (StudentLife, passive sensing) and MRI (resting-state functional connectivity, sgCC seed-based analysis). Calculated features were selected based on previous research. Survey data was collected with both online (REDCap, PHQ-8) and smartphone (StudentLife, Ecological Momentary Assessments, PHQ-2/4) sources.

to disuse (iOS default of 30 seconds, Android default vary by manufacturer). Notification and system services do not influence the measurement of unlock duration. While not an absolute measurement of phone usage it is the closest approximation implemented in StudentLife. In Cohort 1, unlock duration (phone usage) was continually sampled, providing coverage 100% of the time. This was decreased in Cohort 2 to help conserve battery usage. In Cohort 2, phones were remotely triggered every 10 min, sampling 1 min every 10 min period (minimum 10% temporal coverage), unless conversation was detected during the 1-min sampling period, in which case sampling was extended up to 3 min for a maximum of 30% temporal coverage.

Ecological Momentary Assessments

Students were prompted once a week within the StudentLife application during the term to complete a few short surveys as Ecological Momentary Assessments EMA, one of which was PHQ-4 (Shiffman et al., 2008). In the current study PHQ-4 was collected weekly as an EMA PHQ-4 is a modified, shorter version of the PHQ-8 which in four questions provides a glimpse of depressive and anxious symptoms (two questions related to each, with the two depression questions comprising the PHQ-2).

Calculation of Circadian Similarity

As part of the StudentLife app, many feature estimates are calculated for each of the following time-epochs: 9 am – 6 pm (day), 6 pm – 12 am (evening), 12 am – 9 am (night). Accordingly, the relative occurrences of behaviors within each epoch can be estimated and analyzed alongside their daily totals as features. Similarity of day-to-day variation in these feature values across these three time periods were calculated using intra-class correlation, or ICC (Shrout and Fleiss, 1979) which was slightly modified to still run with missing values, by changing mean and summation operations to the equivalent NaN operator in MATLAB. Only individuals with more than 20 days of data for a given feature were included.

Several motion features such as time spent walking, biking, running, or in car are calculated, there is some variance in how they are calculated between Android and iOS. The feature with the most similarity across platforms, which allows for the retention of the greatest number of subjects is the feature "time still," which is a relatively simple metric which is calculated by how much time the phone is still or not moving. This was broken into three time-epochs as mentioned above and the similarity of activity cycles (or lack thereof) across days was calculated using ICC and termed Circadian Stillness Similarity.

Previous research has focused on frequency of visits to known places and the interaction with depression. Within the constraints of the currently processed data, these features could not be calculated exactly, but instead the number of unique locations visited during each time-epoch was calculated (Wang et al., 2014). The reliability of how many locations a person visited through the three epochs each day was calculated with ICC and termed Circadian Location Number Similarity.

Subjects

Subjects were first-year undergraduate students recruited from the Dartmouth College community. Cohort 1 included 151 subjects (94 female, mean age = 19.59, std = 1.69, range = 18–28)

which were all scanned during their first year at Dartmouth and followed for the subsequent academic term. Cohort 2 included 106 subjects (75 female, mean age = 18.25, std = 0.63, range = 18–22) which were all scanned during the first academic term of their first year at Dartmouth. In Cohort 2, one subject was removed from the study for having an incompatible phone and one MRI session was stopped due to not reporting a permanent top retainer.

See **Table 1** for a summary of the number of individuals included in each analysis, grouped by Cohort. Subjects were only included in each analysis if they met the minimum number of time-points for smartphone-based StudentLife data and each analysis and had RSFC that passed quality control (see RSFC analysis methods section below for further details). Subjects had normal or corrected-to-normal visual acuity. The Committee for the Protection of Human Subjects at Dartmouth College approved this study. Each subject provided written informed consent in accordance with guidelines set by the above-mentioned committee and received either course credit or monetary compensation for participating in the study.

RSFC Data Collection

Apparatus

Cohort 1 imaging was performed on a Philips Intera Achieva 3-Tesla scanner (Philips Medical Systems, Bothell, WA, United States). Cohort 2 imaging was performed on a Siemens MAGNETOM Prisma 3-Tesla scanner (Siemens Medical Solutions, Malvern, PA, United States). Data for both cohorts was collected using a 32-channel phased array head coil. During scanning, participants viewed a white fixation cross on a black background projected on a screen positioned at the head end of the scanner bore, which participants viewed through a mirror mounted on top of the head coil.

Cohort 1 Imaging

Anatomic images were acquired using a high-resolution 3-D magnetization-prepared rapid gradient echo sequence (MP-RAGE; 160 sagittal slices; TE, 4.6 ms; TR, 9.9 ms; flip angle, 8°; voxel size, 1 × 1 × 1 mm). Resting-state functional images were collected using T2*-weighted fast field echo, echo planar functional imaging sensitive to BOLD contrast (TR = 2500 ms; TE = 35 ms; flip angle = 90°; 3 × 3 mm in-plane resolution; sense factor of 2). Functional scanning was performed in one or two runs; during each run, 240 brain volumes (36 slices, 3.5 mm slice thickness, 0.5 mm skip between slices) were acquired, allowing complete brain coverage. As such, each participant completed between 10 and 20 min of RSFC scanning.

TABLE 1 | Summary of the number of subjects in each analysis.

	Cohort 1	Cohort 2
Total scanned	151	106
RSFC data (Passed QC)	145	93
PHQ-8	65	89
PHQ-4 (>= 1-Day)	84	89
PHQ-2 (>= 1-Day)	84	89
Unlock duration (>= 20-Days)	77	89

Cohort 2 Imaging

Anatomic images were acquired using a high-resolution 3-D magnetization-prepared rapid gradient echo sequence (MP-RAGE; 192 sagittal slices; TE, 2.32 ms; TR, 2300 ms; flip angle, 8°; voxel size, 1 × 1 × 1 mm) with a Grappa 2 acceleration factor. Resting-state functional images were collected using T2*-weighted fast field echo, echo planar functional imaging sensitive to BOLD contrast (TR = 1190 ms; TE = 32 ms; flip angle = 63°; 2.4 × 2.4 mm in-plane resolution; SMS factor of 4). Functional scanning was performed in one or two runs; during each run, 605 volumes (46 slices, 3 mm slice thickness, no skip between slices) were acquired, allowing complete brain coverage. As such, each participant completed 12 or 24 min of RSFC scanning. Initial data acquisition and conversion to BIDS for cohort 2 was facilitated by the ReproIn specification and tools (ReproNim project NIH-NIBIB P41 EB019936) and organized into BIDS format with datalad (Gorgolewski et al., 2016; Halchenko et al., 2017).

RSFC Analyses

All processing was performed using a standard previously published processing stream (Power et al., 2014) with two exceptions: frame-displacement (FD) threshold was set to 0.25 mm (instead of 0.2 mm) and 36 motion parameters (instead of 24) were used for motion regression. Functional images were preprocessed to reduce artifacts, including: (i) slice-timing correction, (ii) rigid body realignment to correct for head movement within and across runs, (iii) within-run intensity normalization such that the intensity of all voxels and volumes achieved a mode value of 1000 scale with 10 units equal to ∼1% signal change, (iv) transformation to a standardized atlas space (3 mm isotropic voxels) based on (Talairach and Tournoux, 1988), (v) frame censoring, (vi) nuisance regression (excluding censored frames), (vii) interpolation, and (viii) bandpass filtering ($0.009 < f < 0.08$Hz) following Power et al. (2014) and using exactly the same processing stream as Huckins et al. (2019). Final correlation calculations between time-courses were calculated based upon *uncensored* frames. Preprocessing steps i-v were completed using custom scripts which call 4dfp Tools[1]. Steps specific to resting-state functional-connectivity processing (vi-x) were completed using custom MATLAB (Version R2012b, by MathWorks, Natick, MA, United States) scripts.

Nuisance Regressors

To control for motion, a Volterra expansion (Friston et al., 1996) with 36 motion parameters was used. This expansion includes motion, motion squared, motion at the previous two frames, and motion in the previous two frames squared. Tissue-based nuisance regressors were calculated by taking the mean signal across voxels within each of the following individual masks from FreeSurfer[2] (Dale et al., 1999; Desikan et al., 2006): an eroded

[1]ftp://imaging.wustl.edu/pub/raichlab/4dfp_tools/

[2]http://surfer.nmr.mgh.harvard.edu

(up to 4x) ventricular mask for the cerebrospinal fluid, an eroded white matter mask for the white matter signal, and a whole-brain mask for global signal. When eroded masks included no voxels, lesser erosions were progressive considered until a mask with qualifying voxels was identified. This occurred infrequently for white-matter masks while erosions of 1 were often used for CSF masks. The first derivative for each tissue regressor, as calculated by the difference from the current from to the previous frame, was also included.

Volume Censoring and Data Retention

Movement of the head from one volume to the next (FD) was calculated by the sum of the absolute values of the differentiated realignment values (*x, y, z*, pitch, roll, yaw) at each time-point (Power et al., 2012). A frame displacement threshold of 0.25mm was used. Volumes with motion above the frame displacement threshold were identified and replaced after multiple regressions but prior to frequency filtering. Spectral decomposition of the uncensored data was performed and used to reconstitute (stage vii: interpolation) data at censored time-points. The frequency content of uncensored data was calculated with a least squares spectral analyses for non-uniformly sampled data (Mathias et al., 2004) based upon the Lomb-Scargle periodogram (Lomb, 1976). Segments of data with less than 5 contiguous volumes below the FD threshold were flagged for censoring. Functional runs were only included in the final analysis if the run contained 50 or more uncensored frames. Only subjects with at least 5 min of uncensored data across runs were included in the current study. Consistent with Power et al. (2014), only uncensored volumes were used when calculating temporal correlations.

Neurosynth Analysis and Subgenual Cingulate Cortex Seedmaps

To identify an unbiased sgCC seed to create voxelwise functional seed maps, an automated meta-analysis was performed using Neurosynth for the term "subgenual" (Yarkoni et al., 2011). sgCC seed maps were created from a 4mm spherical seed placed at 0, 25, −10 (MNI coordinates), which was the peak of the term "subgenual" as of February 17th, 2017 and are centered around BA 25. The mean time-course from this seed was correlated with the time-course from every voxel within the brain. These seed maps, i.e., maps of resting-state connectivity from the subgenual region, were produced for each individual that passed quality control (more than 5 min of uncensored frames, see above for more details).

Combining Data

Since the version of the StudentLife application used in the current study generates 182 features automatically, and with RSFC it is possible to generate thousands of features, it is necessary to minimize the number of features compared given the relatively small size of the Cohorts ($N < 100$). To minimize the number of features inspected, unlock duration was the only feature inspected given its simplicity to calculate and previously identified relationship with PHQ-8 (Wang et al., 2018). While many features were automatically calculated, unlock duration (e.g., screen time) was first targeted as a simple feature both to calculate and to conceptualize as it can be considered a proxy for total phone screen time.

For all surveys analyzed here, one time-point was sufficient for a subject to be included in the current analyses. If there were multiple responses to ecological momentary assessments (EMAs, e.g., surveys prompted by the application) over the course of the term those responses were averaged. Individuals were included in the passive sensing unlock duration analysis if they had 20 days of quality data with more than 16 h of quality unlock duration data for each day that was included.

Group Analyses and Statistics

Subgenual cingulate cortex seedmaps from Cohort 1 were correlated with unlock duration sampled from smartphone usage with the StudentLife application. For each analysis, the degrees of freedom was N-2, with N being the number of subjects which is listed in **Table 1**. Results from the unlock duration and sgCC correlational analysis from Cohort 1 were volume corrected to account for multiple comparisons using AFNI's 3dClustSim ACF function. Results from the sgCC/unlock duration analysis were used to restrict the regions investigated in further analyses. Given the proof-of-concept and exploratory nature of the current work, clusters are marked as having passed volume-correction or not.

Visualization

All results were transformed into MNI space (Montreal Neurological Institute) and mapped onto the Conte69 template for volume-based slices or inflated surfaces for visualization (Van Essen et al., 2012). Group results were visualized in Connectome Workbench Version 1.1.1 (Marcus et al., 2010).

RESULTS

Self-Reported Depression Measures

Depression symptomatology severity was assessed pre-scan with an online survey using PHQ-8 and during the term using the StudentLife application to administer the PHQ-4 (which contains the PHQ-2). PHQ-8 distributions were similar between Cohort 1 and Cohort 2 (mean = 4.77, 4.52; SEM = 0.58, 0.47, respectively). Depression severity (as categorized by Kroenke et al., 2001) revealed that in both Cohorts a large portion of individuals had minimal depressive symptoms (56.9 and 62.9%, respectively), leaving roughly 40 percent of individuals with a range of depressive symptoms (**Supplementary Table S1**). PHQ-4 distributions where also similar between Cohort 1 and Cohort 2 (mean = 2.52, 2.09; SEM = 0.24, 0.18, respectively). PHQ-2 distributions where also similar between Cohort 1 and Cohort 2 (mean = 0.77, 0.80; SEM = 0.13, 0.10, respectively). Density figures for all self-reported depression symptoms can be found in the **Supplementary Figure S1**.

Passive Sensing Features Correlated With sgCC Connectivity

In Cohort 1 exploratory whole-brain analyses of the correlation between unlock duration and sgCC seedmaps identified a

large cluster (584 voxels, 15,768mm^3) in the ventromedial prefrontal cortex with a positive linear relationship (**Figure 2** and **Supplementary Figure S2A**). This cluster extended from the ventral striatum to medial frontal orbitofrontal cortex (OFC) and dorsally to medial prefrontal cortex. Information about subpeaks within this cluster can be found in **Table 2**. To determine if these results replicated in Cohort 2, the cluster identified in Cohort 1 was used as a mask and voxels which showed a significant positive relationship between unlock duration and sgCC connectivity in Cohort 2 were identified. This analysis identified a cluster with the peak located at −6, 51, −18 (MNI coordinates, peak T = 2.94, voxel extent = 42, volume-corrected to $p < 0.05$) (**Supplementary Figure S2B**).

Two features estimating the reliability of day-to-day activity patterns, including phone motion measured as how long the phone is still at three different time epochs throughout the day and the number of locations an individual visits per time epoch were subsequently analyzed. Circadian Stillness Similarity derived from phone stillness across the three daily time-epochs did not identify any significant regions in Cohort 1 (N = 77) after volume-correction within the prefrontal mask from unlock duration used in other analyses in the main text. Cohort 2 (N = 89) did, however, identify a small cluster (MNI = 12, 45, −12; t = 2.73; 31 voxels) in right medial OFC which was negatively correlated with circadian similarity. In other words, individuals with daily movements patterns that were more similar had less connectivity between sgCC and medial OFC. Similar results were observed for Circadian Location Number Similarity, where no clusters passed volume correction in Cohort 1, but a small cluster (MNI = −9, 45, −12; t = −2.54; 31 voxels) was found in left medial OFC (not shown given similarity with **Supplementary Figure S5**). Between the two analyses there were 7 voxels which overlapped.

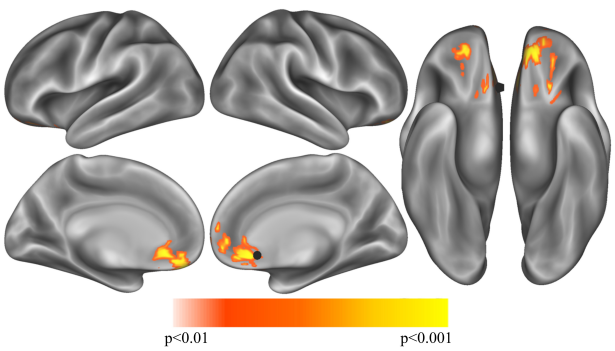


FIGURE 2 | Exploratory analysis correlation sgCC RSFC seedmaps correlated with mean unlock duration identified a cluster with a positive relationship to unlock duration in the ventromedial prefrontal cortex ($p < 0.01$, volume corrected using ACF to $p < 0.001$) shown on inflated lateral (top left), medial (bottom left) and ventral (right) cortical surfaces. The sgCC seed is represented as a black 10 mm sphere, larger than the 4 mm sphere used to create the seedmaps for visualization purposes.

TABLE 2 | Exploratory analysis correlation sgCC RSFC seedmaps correlated with mean unlock duration (smartphone screen time) identified one cluster in the ventromedial prefrontal cortex ($p < 0.01$, volume corrected using AFNI's ACF to $p < 0.001$, $k > 449$, voxel extent = 548).

Best estimate of region	X	Y	Z	T
Caudate	−15	21	−9	4.29
Caudate	12	21	−9	3.64
Anterior sgCC	6	33	−12	3.34

Peaks were identified with xjview 9.6, showing 3 maximia within this cluster, at least 8 mm apart.

Self-Reported Depression Symptoms Correlated With sgCC Connectivity

Previous research (Wang et al., 2018) identified a relationship between depressive symptoms and unlock duration. To determine if depressive symptoms and unlock duration had overlap in the brain connectivity (seed based subgenual RSFC) regressions for both computer-based pre-screening (PHQ-8), phone based post-scanning (PHQ-2/4 as EMA) were performed. Results from each of these analyses were masked with the cluster identified in Cohort 1's sgCC/unlock duration analysis.

PHQ-8 computer-based surveys correlated with sgCC connectivity maps identified clusters with a positive relationship with sgCC connectivity in both Cohorts and identified a cluster which overlapped between the two. Cohort 1 revealed one cluster at which passed volume-correction −21, 42, −12 (peak T = 3.19, voxel extent = 63, volume corrected to $p < 0.05$), 24, 51, −9 (peak T = 2.55, voxel extent = 15, did not pass volume correction) (**Figure 3** and **Table 3**). In the PHQ-8 analysis of Cohort 2, results were further masked by the cluster which

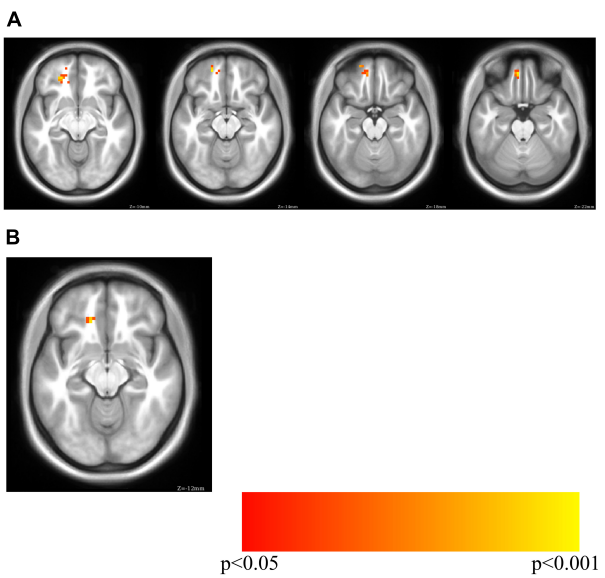


FIGURE 3 | PHQ-8 regression for sgCC connectivity seedmaps for **(A)** Cohort 1 (MNI Z of –10 to –22 in steps of 4) and **(B)** overlap between Cohort 1 and Cohort 2 (MNI Z of –12). Cohort 1 PHQ-8 results were masked with the volume-corrected cluster identified in the Cohort 1 phone usage analysis (unlock duration) and Cohort 2 PHQ-8 results were masked with the PHQ-8 results from Cohort 1.

TABLE 3 | Results for the correlation of sgCC RSFC seedmaps with PHQ-8, masked by phone screen time results.

Best estimate of region	X	Y	Z	T	Extent
Cohort 1					
Left OFC	−21	42	−12	3.19	63
	−18	51	−15	3.09	Subpeak
	−6	48	−21	3.04	Subpeak
Right OFC*	24	51	−9	2.55	15
	18	42	−12	2.19	Subpeak
Overlap between cohorts					
Left OFC	−15	33	−12	2.98	8

*Overlap between Cohort 1 and Cohort 2 for sgCC RSFC seedmaps correlated with PHQ-8 (Bottom). Subpeaks are at least 8 mm apart. *signifies that cluster didn't pass volume correction. Cohort 1 PHQ-8 results were masked with the cluster identified in the Cohort 1 phone usage analysis (unlock duration) and Cohort 2 PHQ-8 results were masked with the PHQ-8 results from Cohort 1.*

passed volume-correction in the Cohort 1 PHQ-8 analysis (63 voxels), identifying 1 significant cluster in Cohort 2, located at −15, 33, −12 (peak $T = 2.98$, voxel extent = 8, volume corrected to $p < 0.05$). In addition to identifying a cluster with overlap between the both Cohorts for the PHQ-8 analysis, qualitative visual inspection suggests proximal cortical regions in both cohorts meeting a voxelwise threshold of $p < 0.05$, with regions proximal to the mask having overlap at a threshold of $p < 0.05$ and increased overlap, including right OFC at a more liberal threshold of $p < 0.1$.

PHQ-4 EMAs correlated with sgCC connectivity maps identified peaks in Cohort 1 and 2, but there was no overlap in the clusters between the Cohorts (**Supplementary Figure S3** and **Table S2**). In Cohort 1 no significant clusters were identified when PHQ-4 was masked with Cohort 1 unlock duration. As Cohort 1 didn't identify any regions which passed volume-correction, there was no overlap of significant volume-corrected regions between Cohort 1 and Cohort 2 for PHQ-2 (**Supplementary Figure S4**). As such, Cohort 2 results were masked with the Cohort 1 unlock duration cluster which identified one significant cluster with the peak at −15, 30, −12 (peak $T = 3.71$, voxel extent = 41, volume corrected to $p < 0.05$). Two clusters were identified that didn't pass volume correction were also identified at −9, 51, −18 (peak $T = 2.87$, voxel extent = 28, volume correction *ns*) and 24, 39, −15 (peak $T = 1.87$, voxel extent = 9, volume correction *ns*).

PHQ-4 includes two anxiety questions, so the subsequent analysis was restricted to the two questions related to depressive symptoms which comprise the PHQ-2. As Cohort 1 didn't identify any regions which passed volume-correction, there was no overlap of significant volume-corrected regions between Cohort 1 and Cohort 2 for PHQ-2. As such, Cohort 2 results were masked with the Cohort 1 Unlock Duration cluster which identified 1 cluster which passed volume correction, with the peak at −18, 30, −12 (peak $T = 3.81$, voxel extent = 60, volume corrected to $p < 0.05$). One cluster was identified that didn't pass volume correction with peak at −9, 42, −27 (peak $T = 2.93$, voxel extent = 40, *ns*).

Overlap Across Analyses

Given the similarity of regions found across the PHQ analyses in Cohort 2, the overlap between the results of PHQ 2/4 masked by the Cohort 1 unlock duration was investigated, with 39 voxels out of the 41 voxels identified in the PHQ-2 analysis overlapping with the PHQ-4 analysis. The overlap between Cohort 2 PHQ-2, 4 and 8 identified 11 voxels, which are located around the peaks of the PHQ-8 analysis.

DISCUSSION

The current manuscript is provided as a proof-of-concept example of how passive smartphone metrics, active smartphone-based surveys of mental health and computer-based surveys of mental health with brain connectivity measures can be linked. Specifically, RSFC between the subgenual cingulate cortex, a region previously implicated in depression, and nearby ventral prefrontal regions, was strongly related to unlock duration, such that more connectivity was associated with more screen time, which has been implicated as being related to self-reported depressive symptoms. The link between RSFC and individual differences has long been established but extending that and combining it with an individual's behavior inferred from smartphone sensors provides exciting new directions. While the results presented here are a relatively simple analysis of complex, highly dimensional data, methods are discussed which could be used in the future to combine these highly multivariate and complex datasets in exciting ways.

Phone-related screen time, defined here as the amount of time a phone is unlocked, or unlock duration, has previously been shown to be related to self-reported depression levels (Twenge et al., 2018; Wang et al., 2018). An exploratory analysis in Cohort 1 of the correlation between unlock duration and sgCC seedmaps identified a large cluster which extended from the anterior caudate to medial frontal OFC and dorsally to medial prefrontal cortex, a result which was replicated in Cohort 2 with a smaller voxel extent, even though the sampling rate for screen time was greatly reduced, reducing our sensitivity to pick up individual differences in phone usage for this cohort. Next, to determine if depressive symptoms showed a similar pattern of connectivity between sgCC and ventral prefrontal cortex the cluster from Cohort 1's unlock duration analysis was used as a mask with PHQ-8, a commonly used survey to assess depressive symptoms in the general population. Two small clusters of overlap were identified in the left OFC, one of them neighboring voxels that were identified to replicate in the unlock duration analysis between the Cohorts. While these clusters are not large and would not necessarily survive volume correction on their own, observing similar regions across Cohorts and analyses suggests that there is a link between depressive symptoms and related behaviors and sgCC-OFC connectivity, particularly left OFC that should be further investigated. The PHQ-4, which contains two depression questions and two anxiety questions, did not show the same robust relationship across both Cohorts, with no voxels overlapping, although Cohort 2 identified a cluster in the left OFC which overlapped with results

observed with PHQ-8 in both Cohorts. Connectivity between the sgCC seed (BA 25), located at 0, 25, −10 and the left OFC region around −15, 33, −12 shows a consistent relationship between self-reported depressive symptoms and screen time, which has previously been associated with depression. Increased connectivity between sgCC, a region involved in processing of valenced information about the self (Moran et al., 2006) and OFC, which is involved in valuation and reward processing has been linked increased depressive symptoms and *screen time* across both Cohorts. Similar results were observed with PHQ-2, which only contains the two questions directly related to mood. It seems quite plausible that regions involved in valence processing related to the concept of self and a more general reward valuation processing region would have increased connectivity in individuals with higher depressive symptoms.

Individuals in Cohort 2 with daily movement routines which were more similar from day-to-day exhibited less connectivity between sgCC and medial OFC. This is the opposite direction of a correlation that unlock duration and PHQ depression surveys identified, which is expected in light of results by Lyall et al. (2018), where individuals that exhibited activity patterns with reliable rest/activity cycles were less likely to be depressed. Similarly, individuals with more similarity in locations visited, meaning consistent day-to-day schedules had less connectivity between sgCC and medial OFC, which in the current study is associated with lower depression levels. The current work used very large time epochs and could be investigated in more depth with future modifications to the StudentLife application and feature generation pipeline to perform finer grained analyses. Similarly, extending StudentLife to calculate frequently visited locations such as Burns et al. (2011) could prove fruitful. In summary, in the current dataset the regularity in the number of locations visited (as measured by GPS) and regularity in the time that the phone is not moving are both negatively correlated with connectivity between the sgCC and medial OFC.

We have shown that RSFC of the brain, as measured with MRI, in two separate Cohorts of individuals, with two separate MRI's and two separate versions of the StudentLife application and three separate passive-sensing feature show similarity in the results observed. The cluster identified with the unlock duration analysis covered an extent similar to that of the limbic network previously identified (Yeo et al., 2011; Choi et al., 2012). Due to the constraints we imposed on the analysis, all of the subsequent results were within this area, but noticeably, many of the results were proximal to the left OFC, which is also a member of a set of nodes which are commonly activated during reward processing and can form their own preferentially coupled system (Huckins et al., 2018) and is identified as a peak of the term "reward" in reverse-inference meta-analyses using Neurosynth (Yarkoni et al., 2011).

LIMITATIONS AND FUTURE DIRECTIONS

The current work is a first-pass at analyzing longitudinal multi-cohort, multimodality data and has several limitations. There are several ways in which future research may provide a more comprehensive survey of the relationships between the diverse set of features provided from passive smartphone sensing, functional brain connectivity measures and self-reported measures of depression or other mental health metrics. The relatively small number of clinically depressed individual in the current sample weighs the results heavily on the RSFC and passive-sensing features from those individuals. Test-retest within the moderately sized samples allows for identification of factors with reliable cross-cohort replicability in RSFC both and passive-sensing features. Ideally, similar sensing features could be collected across many sites, allow for identification and characterization of depressive subtypes that span across passive-sensing and RSFC as has been done by Drysdale et al. (2016) with RSFC and survey data. Diagnosis of depression by neuroimaging techniques such as RSFC MRI could potentially be cost prohibitive in a medical setting. With that said, the medical costs associated with untreated depression accounts for $26.1 billion per year with a total economic loss about $83 billion in just the United States alone (Greenberg et al., 2003). As noted in the current Cohorts, roughly 40 percent of participants had mild depressive symptoms or worse as measured.

In the current study, particularly Cohort 2 in which data quality was actively monitored, a relatively large portion of individuals from those scanned was retained (see **Table 1**). The sample sizes used here would have been considered relatively large several years ago. Increased sample sizes in the current study would help future analyses given the large number of features from both passive mobile smartphone sensing and RSFC. An outstanding question is if long-term changes in depressive symptoms can be better predicted by RSFC or smartphone sensing metrics at the initiation of the study or if changes in either of these over time parallel depressive symptoms. Ideally to assess this a large number of individuals would be tracked over multiple years. In the second Cohort our working group aims to track them over multiple years while eventually increasing the number of individuals enrolled. Furthermore, including multiple sites, as the ABCD study does (Volkow et al., 2017), would increase applicability to a wider population. Multiple research sites are currently collecting MRI data, self-reported surveys and smartphone sensing metrics. An unresolved issue is what, exactly, is the optimal approach to analyze the huge amounts of multivariate data produced by these methods.

Application Changes Between Cohorts

In the current study, unlock duration data collection changed between the cohorts. In Cohort 1, unlock duration was continually sampled, while in Cohort 2 unlock duration was adaptively sampled between 10 and 30% of the time. This change was instituted to optimize battery life, a primary limitation to users being willing to keep the StudentLife app on their phone. By decreasing the amount of time sampled from 100% to 10–30%, our ability to accurately estimate unlock duration may decrease slightly as evidenced by an observed decrease in peak effect (T-value) and voxel extent. As with all passive and active smartphone features, the ability to collect data must be weighed against the invasiveness to the user experience, either through app prompts or decreased battery life and phone speed.

Feature Selection and Calculation

In the current study, initial analyses focused on unlock duration as a proxy for general phone usage then investigated the similarity of individuals circadian rhythms from day to day and how each of these was related to brain connectivity from a region known to be involved in depression and many cognitive functions. Unlock duration on its own in very unlikely to be an optimal feature to predict depression and this is where generating and testing a variety of higher-level features may prove fruitful. Identifying changes in features from day-to-day or week-to-week may increase predictability, such as an increase in unlock duration could be associated with increases in depression within an individual (Wang et al., 2018). Variability or stability of passive-sensing features may also be able to predict individual differences in depression. Ideally, a template of passive-sensing features for non-depressed individuals could be created and deviance from this template could be calculated as a sort of depression-index or propensity score. This high-level feature could then be linked to deviance of brain connectivity patterns from non-depressed individuals. Critically, future work should select features that reflect not just phone usage and other standard passive-sensing metrics, but build upon the current sensing literature related to depression (Burns et al., 2011; Canzian and Musolesi, 2015; Saeb et al., 2015; Lyall et al., 2018; Wang et al., 2018) and calculate higher level features which are likely to better reflect nuances in behavioral differences across individuals.

Temporal Factors Related to School

The demands of the academic term provide a generally applicable path of stress which is shaped over the term. Avoiding, or potentially purposefully collecting MRI data during finals, which may be particularly stressful, or during popular social weekends may lead to changes in stress levels, sleep patterns and other variables which could alter connectivity patterns and self-reported behavioral data that would have otherwise been observed. In the study herein, attempts were made to scan before finals and avoid well-known "party weekends." Future studies may be able to capitalize on temporal differences in stress and depression levels by scanning at these peak times of stress or sleep deprivation and comparing that data to less stressful times, such as the beginning of the term.

Functional Differences and Alignment Across Individuals

Resting-state functional connectivity shows robust and relatively reliable connectivity across large groups of individuals across methods (Yeo et al., 2011; Gordon et al., 2016). Meanwhile there are individual differences in the cortical extent of large-scale functional regions across individuals and even the network membership of these regions can vary (Gordon et al., 2017). Furthermore, critical to identifying group and individual differences is acquiring a large quantity of high-quality data (Gratton et al., 2018). Defining networks on an individual basis will likely help in the pursuit of the individual differences in brain connectivity that underlie depression. Variability in RSFC has been observed at the functional parcel level, but what about at finer resolutions? While a departure of traditional anatomical alignment methods, hyperalignment is a method which attempts to align brain based on similar response patterns in high-dimensional space (Guntupalli et al., 2016). While this method originated using time-locked dynamic stimuli such as a movie, it has recently been applied to RSFC as connectivity hyperalignment (CHA), which revealed both coarse-scale, areal structure as previously observed, along with fine-scale structure which was previously inaccessible. Applying CHA to RSFC data will hopefully allow for increased ability to discern individual differences in depression and other mental-health metrics.

Voxelwise Resting-State Functional Connectivity

A relatively simple first-pass method is to target specific region and feature pairs. If there are *a priori* hypotheses related to the topic of interest it may be possible to look at connectivity from one region using seed maps or between a small number of regions and relate them to specific passive-sensing features. As shown here this is plausible but even correlating seed maps with 1 sensing variable leads to potential multiple comparisons issues based on the 50,000+ voxels in the brain using a 3 mm^3 voxel size. Recent statistical simulations have suggested an increased false-positive rate associated with older versions of 3dClustSim, a function of AFNI (Cox et al., 2017). Indeed, the authors of 3dClustSim now suggest using a different algorithm with the same program, the autocorrelation function (ACF) with a high p-value threshold per voxel to minimize the possibility of false-positives. In some datasets, at lower p-value thresholds ACF requires a much larger voxel-extent than the old version of 3dClustSim. The increased voxel-extent may make it less likely to identify smaller functional regions in a whole-brain regression using a lower per-voxel p-value threshold ($p < 0.05$). This evolution of methods decreases the rate of the false-positives which is critical but requires a larger expected functional region, a very strong effect size or a very large number of participants. Across all possible methods presented here there are a variety of factors which should be taken into consideration to decrease false positive rates. Having a large number of subjects to draw data will increase the portion of the population sampled.

If possible having two distinct Cohorts to analyze then looking for overlap in results between the Cohorts would decrease false positives due to random sampling, Cohort specific variance, and further increase the total size of individuals sampled. The above factors apply to most any study. With passive smartphone mobile sensing there are many features which can be measured or computed based on the intersection of multiple features. For example, "phone unlock duration" is a very simple metric, which measures the time that the smartphone was unlocked. This can be further broken down into location specific features, such as "phone unlock duration at dorm" or "phone unlock duration at study places" by looking at the intersection of location on a geo-tagged campus and "phone unlock duration." Given the large number of initial features that can be calculated, along with the nearly endless number of meta-features that could potentially be generated, making sure that the feature

is relatively straightforward to calculate and interpret should be at the forefront of anyone analyzing passive-mobile phone sensing features. Features that are difficult to calculate or interpret could easily be embedded with unforeseen confounds. Furthermore, such features should be validated to make sure they are measuring the effect or phenomena they are supposed to in an accurate manner.

Typically, only features with sensing data from many days should be used to get a more stable estimate of that features' value. While putting a sensing application of many students' phones may seem like a plausible method for maximizing data collect, there are a variety of factors which can lead to reduced data collection, potentially rendering an individual's sensing data unusable. Phone operating system (OS) updates can often change application permission or render the sensing application completely useless. To avoid this beta testing should be done as early as possible and new versions of the application that are compatible with the latest OS pushed to participants. Participant non-compliance or attrition is another important factor to consider. Individuals may delete the application, limit its permissions within the OS or otherwise limit the researcher's' ability to accurately measure data. Clearly, it is the individual's choice to continue to participate in any study, particularly one where large amounts of data are being collected (anonymously) on their habits. It may be difficult for the researcher to determine if the individual has deleted the application or simply not uploaded their data in while. Finally, a rate of attrition is expected in all longitudinal studies and some individuals may simply decide that they do not wish to continue their participation in the study.

Whole-Brain and Network-Based Connectivity

A possible method to deal with the large number of comparisons related to voxelwise or whole-brain connectivity is to simply look at connectivity between a set of predefined regions or parcellation (Power et al., 2011; Yeo et al., 2011; Poldrack et al., 2015; Gordon et al., 2016; Huckins et al., 2018). Connectivity between each pair of regions can be correlated with the sensing feature of interest. Unfortunately, many of the commonly used parcellations have many nodes, which increases the total number of comparisons in a non-linear manner as the number of nodes increases. The number of comparisons can soon approach the number of comparisons evident when using voxelwise seed maps without methods such as voxel extent to appropriately correct for the associated multiple comparisons.

A simple but perhaps relatively unsophisticated sophisticated method is to calculate mean connectivity within a functional system or network. The system or network would be determined off of data driven approach such community detection using a random walk technique like InfoMap (Rosvall and Bergstrom, 2008) or regions identified as being part of a coherent functional system using another method or even searching Neurosynth.org for a term of interest. In this approach, the mean of all Fisher r-to-z transformed correlation values between nodes of interest is calculated. For example, mean connectivity within the Cingulo-Opercular network would be calculated between all nodes or parcels belonging to that network. Between-network or system connectivity can also be calculated by taking the mean of all pairwise connections between the two networks of interest. This can greatly reduce the number of total connections observed, thus reducing the multiple comparisons problem mentioned under the whole-brain connectivity section. One drawback to this method is that it is not selective about which connections it is using in the calculation – specifically, that it may be and probably is including connections that are not physiologically or psychologically relevant.

A plausible may to reduce the number of connections by selections ones that are likely to be "real," such that information may actually travel through that connection on the neural level, even if not on a first-order or even second-order synapse. Multiple approaches have been taken to identify meaningful connections. Within or between networks there are likely to be positive and negative correlations, which then somewhat cancel out. One could take the absolute value of each connection before averaging across the network, but this would introduce bias in any connections with a distribution of correlation values that included positive and negative values. Values of correlation, or connectivity measures in the brain vary by orders of magnitude. Identifying a multiscale network backbone that accounts for important connections within and between communities, regardless of the connectivity strength would be a method to decrease the number of connections analyzed. One way of identifying the network backbone is to use the *z*-value from each connection as the weight, or amount of information that could travel between the two brain regions that the connectivity was estimated from. A group did just this (Serrano et al., 2009), identifying connections which are statistically relevant across multiple scales of connectivity, work which has been extended non-parametrically (Foti et al., 2011). By identifying the network backbone for each individual (Huckins et al., 2019), it may be plausible to identify a variety of subcategories or continuums of depression along which different symptom severities fall for each individual, along with passive smartphone monitoring will allow for greater insight into interactions of behavioral, self-report and physiological RDoC matrix criteria.

Wrangling High Dimensional Data

A variety of techniques can be used to extract information from data that are both longitudinal and high-dimensional; that is, situations where the data are collected from participants at multiple time points and the number of covariates begins to approach, or even surpasses the number of subjects in the dataset (Wang et al., 2012; Cheng et al., 2014; Zipunnikov et al., 2014; Chu et al., 2016).

As has been mentioned repeatedly above, both with resting-state and passive smartphone sensing there are a large quantity of features and analyses that can be generated. In the current study we chose features that were reasonable based on previous data but are unlikely to be the optimal features that describe the relationship between depression, passive mobile sensing and brain connectivity. Multiple approaches could be taken with data

from both sources. One approach which would greatly decrease the number of features that were necessary including trying to create a singular propensity metric, or biomarker of depression for both the resting-state fMRI data and a separate one for the sensing data then observing the relationship between the two. Alternatively, data reduction techniques such as independent component analysis could be applied to each group then the relationship between them could be measured. Many researchers have taken a "risk" or "propensity" score approach, where they generate models which contain predictive variables (gender, substance use, family history) pertinent to the outcome of interest and use the propensity score as a regressor when doing analyses at the group or individual difference level (Stuart, 2010; Hansen et al., 2012). This could be applied to smartphone data, but only once appropriate sensor features, and model have been calculated. By creating a unitary risk feature multiple comparison issues can be greatly mitigated. Data reduction techniques that account for variance that is common between two data modalities such as joint ICA, parallel ICA and CCA-Joint ICA, which has been implemented for combining high-dimension data across fMRI and genetic data (FusionICA, available from http://mialab.mrn.org/software/fit/).

Unresolved Questions About Directionality and Timing

In the current sample, resting-state fMRI data is from 1 time-point while mobile smartphone sensing data is dynamic and data is collected over a longer period of time. An unresolved question is if changes in fMRI data across multiple sessions reflects or predict changes in smartphone usage. Likely a more sensitive measure would be to do the reverse – using changes in smartphone usage, which is continuously monitored, to predict when there may be changes in brain connectivity as measured by fMRI. Changes in depressive symptoms have been successfully predicted with passive smartphone features (Wang et al., 2018), and may be useful for signaling when an individual should be referred to clinical services or brought in for a subsequent fMRI session. Longitudinal penalized functional regression is a method designed to deal with multiple timepoints of both exposure and outcomes (Goldsmith et al., 2012) which may help provide insight into the temporal association between brain connectivity, depression and phone usage.

Moderating Factors of RSFC

Resting-state functional connectivity has repeatedly been shown to be relatively stable across individuals and time, displaying similar network structure across thousands of individuals. While similar network structure and connectivity patterns are observed between sites, preprocessing methods, and Cohorts, differences between individuals are observed across individual differences in personality, affect and current mood have been related to alterations in RSFC. Furthermore, individual differences in the network structure on an individual level have been observed. Properly mapping individual differences in networks across the cortex would allow for better cross-subject alignment. The network assignment of particular regions may in itself be linked to depressive symptoms, while lining up networks would allow for the proper comparison of networks across individuals. Additionally, the current state physiological state an individual is in, such as food satiety or caffeination status can influence their mood (Rogers and Lloyd, 1994) and has also been shown to influence an individual's brain connectivity (Poldrack et al., 2015). While there are a variety of factors that can influence RSFC, reliable individual differences across brain disorders have been observed in previous studies and here. As the predictive accuracy of RSFC or other neuroimaging methods increases the field may move closer to using MRI as a biomarker of depression, as has been done with physical pain (Atlas et al., 2010; Wager et al., 2013).

CONCLUSION

In summary, the current work identified proof-of-concept relationships between RSFC of the brain, passive mobile smartphone sensing features (unlock duration and circadian similarity of stillness and number of location visited), web-based self-reported surveys of depressive symptoms (PHQ-8) and mobile smartphone based ecological momentary assessments of depressive symptoms (PHQ-4). The results observed here extend previous work which relates the amount of time spent using a phone is with depressive symptoms. Further, these symptoms, both before and after time-of-scanning (PHQ-8 and PHQ-2/4, respectively), show a relationship with connectivity between areas implicated in depression, reward and processing of valenced self-relevant material. Importantly, these initial results predominantly replicate across the two separate cohorts and similar results are observed across three passive sensing features, increasing the applicability and scope of the findings herein. Although the current results do not elucidate causality in the relationship between phone usage metrics, depression and brain connectivity, future work should aim to do so, especially given recent changes to public policy, with professional groups such as the American Academy of Pediatrics providing suggesting screen-time limits and policy and investor groups calling on media device makes such as Apple and other phone makers. Previous research was extended, with results that replicate across multiple MRI scanners and cohorts all while combining data from a while variety of sources. The analyses done here are by no means comprehensive and we hope that the findings of this study and future research methods proposed herein are useful to a wide-range of researchers. Ultimately continuation and extensions of this research has the potential to provide important insights into mental health, as well as inform psychological treatments and other interventions.

AUTHOR CONTRIBUTIONS

JH, PH, WK, TH, and AC designed the study. Data collection was performed by JH, CR, RL, Ad, EM, and EH. All analyses were conduced by JH. All authors provided manuscript feedback and guidance.

FUNDING

This work was supported by NIMH nos. 5R01MH059282-12 and 5R01DA022582-10.

ACKNOWLEDGMENTS

We would like to acknowledge Yaroslav O. Halchenko for assistance with MRI data conversion and BIDS organization, Terry Sackett for assistance with MRI data collection and the editor and reviewers for their helpful feedback. Pre-print available for reference (Huckins et al., 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2019.00248/full#supplementary-material

REFERENCES

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., et al. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann. Fam. Med.* 8, 348–353. doi: 10.1370/afm.1139

Atlas, L. Y., Bolger, N., Lindquist, M. A., and Wager, T. D. (2010). Brain mediators of predictive cue effects on perceived pain. *J. Neurosci.* 30, 12964–12977. doi: 10.1523/JNEUROSCI.0057-10.2010

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., and Campbell, A. T. (2015). Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. *Psychiatr. Rehabil. J.* 38, 218–226. doi: 10.1037/prj0000130

Berman, M. G., Yourganov, G., Askren, M. K., Ayduk, O., Casey, B. J., Gotlib, I. H., et al. (2013). Dimensionality of brain networks linked to life-long individual differences in self-control. *Nat. Commun.* 4:1373. doi: 10.1038/ncomms2374

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., et al. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage* 83, 550–558. doi: 10.1016/j.neuroimage.2013.05.099

Burns, M. N., Begale, M., Duffecy, J., Gergle, D., Karr, C. J., Giangrande, E., et al. (2011). Harnessing context sensing to develop a mobile intervention for depression. *J. Med. Internet Res.* 13:e55. doi: 10.2196/jmir.1838

Cameron, I. M., Crawford, J. R., Lawton, K., and Reid, I. C. (2008). Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Brit. J. Gen. Pract.* 58, 32–36. doi: 10.3399/bjgp08X263794

Canzian, L., and Musolesi, M. (2015). "Trajectories of depression," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing – UbiComp '15* (New York, NY: ACM Press), 1293–1304. doi: 10.1145/2750858.2805845

Cheng, M.-Y., Honda, T., Li, J., and Peng, H. (2014). Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *Ann. Stat.* 42, 1819–1849. doi: 10.1214/14-AOS1236

Choi, E. Y., Yeo, B. T. T., and Buckner, R. L. (2012). The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.* 108, 2242–2263. doi: 10.1152/jn.00270.2012

Chu, W., Li, R., and Reimherr, M. (2016). Feature screening for time-varying coefficient models with ultrahigh-dimensional longitudinal data. *Ann. Appl. Stat.* 10, 596–617. doi: 10.1214/16-AOAS912

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., and Taylor, P. A. (2017). fMRI clustering and false-positive rates. *Proc. Nat. Acad. Sci. U.S.A.* 114, E3370–E3371. doi: 10.1073/pnas.1614961114

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194. doi: 10.1006/nimg.1998.0395

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., et al. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* 23, 28–38. doi: 10.1038/nm.4246

Eisenberg, D., Hunt, J., and Speer, N. (2013). Mental health in American colleges and universities: variation across student subgroups and across campuses. *J. Nerv. Ment. Dis.* 201, 60–67. doi: 10.1097/NMD.0b013e31827ab077

Foti, N. J., Hughes, J. M., and Rockmore, D. N. (2011). Nonparametric sparsification of complex multiscale networks. *PLoS One* 6:e16431. doi: 10.1371/journal.pone.0016431

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S., and Turner, R. (1996). Movement-related effects in fMRI time-series. *Magn. Reson. Med.* 35, 346–355. doi: 10.1002/mrm.1910350312

Goldsmith, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2012). Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J. R. Stat. Soc. Ser. C* 61, 453–469. doi: 10.1111/j.1467-9876.2011.01031.x

Gordon, E. M., Laumann, T. O., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Dosenbach, N. U. F., et al. (2017). Individual-specific features of brain systems identified with resting state functional correlations. *NeuroImage* 146, 918–939. doi: 10.1016/j.neuroimage.2016.08.032

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex* 26, 288–303. doi: 10.1093/cercor/bhu239

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., et al. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* 98:439-452.e. doi: 10.1016/j.neuron.2018.03.035

Greenberg, P. E., Kessler, R. C., Birnbaum, H. G., Leong, S. A., Lowe, S. W., Berglund, P. A., et al. (2003). The economic burden of depression in the United States: how did it change between 1990 and 2000? *J. Clin. Psychiatry* 64, 1465–1475. doi: 10.4088/JCP.v64n1211

Greicius, M. D., Flores, B. H., Menon, V., Glover, G. H., Solvason, H. B., Kenna, H., et al. (2007). Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol. Psychiatry* 62, 429–437. doi: 10.1016/j.biopsych.2006.09.020

Guntupalli, J. S., Hanke, M., Halchenko, Y. O., Connolly, A. C., Ramadge, P. J., and Haxby, J. V. (2016). A model of representational spaces in human cortex. *Cereb. Cortex* 26, 2919–2934. doi: 10.1093/cercor/bhw068

Halchenko, Y., Hanke, M., Poldrack, B., Debanjum, Alteva, G., Gors, J., et al. (2017). *datalad/datalad 0.9.1*. Available at: https://zenodo.org/record/1000098#.XH8o4skzaM8

Hansen, R. A., Dusetzina, S. B., Ellis, A. R., Stürmer, T., Farley, J. F., and Gaynes, B. N. (2012). Risk of adverse events in treatment-resistant depression: propensity-score-matched comparison of antidepressant augment and switch strategies. *Gen. Hosp. Psychiatry* 34, 192–200. doi: 10.1016/j.genhosppsych.2011.10.001

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. doi: 10.1016/j.jbi.2008.08.010

Holtzheimer, P. E. (2012). Subcallosal cingulate deep brain stimulation for treatment-resistant unipolar and bipolar depression. *Arch. Gen. Psychiatry* 69, 150–158. doi: 10.1001/archgenpsychiatry.2011.1456

Horwitz, A. V., Wakefield, J. C., and Lorenzo-Luaces, L. (2016). "History of depression," in *The Oxford Handbook of Mood Disorders*, eds R. J. DeRubeis and D. R. Strunk (Oxford: Oxford University Press).

Huckins, J. F., Adeyemo, B., Miezin, F. M., Power, J. D., Gordon, E. M., Laumann, T. O., et al. (2019). Reward-related regions form a preferentially coupled system at rest. *Hum. Brain Mapp.* 40, 361–376. doi: 10.1002/hbm.24377

Huckins, J. F., daSilva, A. W., Wang, R., Wang, W., Hedlund, E. L., Murphy, E. I., et al. (2018). Fusing mobile phone sensing and brain imaging to assess depression in college students: a proof-of-concept study. *bioRxiv* [Preprint]. doi: 10.1101/276568

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., et al. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 167, 748–751. doi: 10.1176/appi.ajp.2010.09091379

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., and Pizzagalli, D. A. (2015). Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. *JAMA Psychiatry* 72, 603–611. doi: 10.1001/jamapsychiatry.2015.0071

Khubchandani, J., Brey, R., Kotecki, J., Kleinfelder, J., and Anderson, J. (2016). The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch. Psychiatric Nurs.* 30, 457–462. doi: 10.1016/j.apnu.2016.01.014

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Int. Med.* 16, 606–613.

Kroenke, K., Spitzer, R. L., Williams, J. B. W., and Löwe, B. (2009a). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics* 50, 613–621. doi: 10.1176/appi.psy.50.6.613

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B. W., Berry, J. T., and Mokdad, A. H. (2009b). The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114, 163–173. doi: 10.1016/j.jad.2008.06.026

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., et al. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS One* 8:e69841. doi: 10.1371/journal.pone.0069841

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* 39, 447–462. doi: 10.1007/BF00648343

Lyall, L. M., Wyse, C. A., Graham, N., Ferguson, A., Lyall, D. M., Cullen, B., et al. (2018). Association of disrupted circadian rhythmicity with mood disorders, subjective wellbeing, and cognitive function: a cross-sectional study of 91 105 participants from the UK Biobank. *Lancet Psychiatry* 5, 507–514. doi: 10.1016/S2215-0366(18)30139-1

Marcus, D. S., Fotenos, A. F., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2010). Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. *J. Cogn. Neurosci.* 22, 2677–2684. doi: 10.1162/jocn.2009.21407

Matar, B. J., and Jaalouk, D. (2017). Depression, anxiety, and smartphone addiction in university students- a cross sectional study. *PLoS One* 12:e0182239. doi: 10.1371/journal.pone.0182239

Mathias, A., Grond, F., Guardans, R., Seese, D., Canela, M., and Diebner, H. H., (2004). Algorithms for spectral analysis of irregularly sampled time series. *J. Stat. Softw.* 11, 1–26. doi: 10.18637/jss.v011.i02

Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., et al. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron* 45, 651–660. doi: 10.1016/j.neuron.2005.02.014

Mehrotra, A., Tsapeli, F., Hendley, R., and Musolesi, M. (2017). MyTraces: Investigating correlation and causation between users' emotional states and mobile phone interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 1–21. doi: 10.1145/3130948

Moran, J. M., Macrae, C. N., Heatherton, T. F., Wyland, C. L., and Kelley, W. M. (2006). Neuroanatomical evidence for distinct cognitive and affective components of self. *J. Cogn. Neurosci.* 18, 1586–1594. doi: 10.1162/jocn.2006.18.9.1586

Nielsen.com (2016). *Millennials are Top Smartphone Users*. Available at: https://www.nielsen.com/us/en/insights/news/2016/millennials-are-top-smartphone-users.html

Patriat, R., Molloy, E. K., Meier, T. B., Kirk, G. R., Nair, V. A., Meyerand, M. E., et al. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated. *NeuroImage* 78, 463–473. doi: 10.1016/j.neuroimage.2013.04.013

Poldrack, R. A., Laumann, T. O., Koyejo, O., Gregory, B., Hover, A., Chen, M.-Y., et al. (2015). Long-term neural and physiological phenotyping of a single human. *Nat. Commun.* 6:8885. doi: 10.1038/ncomms9885

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E., (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., et al. (2011). Functional network organization of the human brain. *Neuron* 72, 665–678. doi: 10.1016/j.neuron.2011.09.006

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341. doi: 10.1016/j.neuroimage.2013.08.048

Richiardi, J., Altmann, A., Milazzo, A.-C., Chang, C., Chakravarty, M. M., Banaschewski, T., et al. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science* 348, 1241–1244. doi: 10.1126/science.1255905

Rogers, P. J., and Lloyd, H. M. (1994). Nutrition and mental performance. *Proc. Nutr. Soc.* 53, 443–456. doi: 10.1079/PNS19940049

Rosenman, R., Tennekoon, V., and Hill, L. G. (2011). Measuring bias in self-reported data. *Int. J. Behav. Healthc. Res.* 2:320–332. doi: 10.1504/IJBHR.2011.043414

Rosvall, M., and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proc. Nat. Acad. Sci. U. S. A.* 105, 1118–1123. doi: 10.1073/pnas.0706851105

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., et al. (2015). Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *J. Med. Internet Res.* 17:e175. doi: 10.2196/jmir.4273

Serrano, M. A., Boguñá, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proc. Nat. Acad. Sci. U. S. A.* 106, 6483–6488. doi: 10.1073/pnas.0808904106

Shehzad, Z., Kelly, A. M. C., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., et al. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229. doi: 10.1093/cercor/bhn256

Shiffman, S., Stone, A. A., and Hufford, M. R. (2008). Ecological momentary assessment. *Ann. Rev. Clin. Psychol.* 4, 1–32.

Shrout, P. E., and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86, 420–428. doi: 10.1037/0033-2909.86.2.420

Sinclair, B., Hansell, N. K., Blokland, G. A. M., Martin, N. G., Thompson, P. M., Breakspear, M., et al. (2015). Heritability of the network architecture of intrinsic brain functional connectivity. *NeuroImage* 121, 243–252. doi: 10.1016/j.neuroimage.2015.07.048

Smith, A. (2017). *Record Shares of Americans Now Own Smartphones, have Home Broadband*. Available at: http://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology/

Somerville, L. H., Heatherton, T. F., and Kelley, W. M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat. Neurosci.* 9, 1007–1008. doi: 10.1038/nn1728

Stuart, E. A. (2010). Matching methods for causal inference: a review and a look forward. *Stat. Sci.* 25, 1–21. doi: 10.1214/09-STS313

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxis Atlas of the Human Brain*. New York, N.Y: Thieme Medical Publishers Inc.

Twenge, J. M., Joiner, T. E., Rogers, M. L., and Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. *Clin. Psychol. Sci.* 6, 3–17. doi: 10.1177/2167702617723376

Ustün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., and Murray, C. J. L. (2004). Global burden of depressive disorders in the year 2000. *Brit. J. Psychiatry? J. Ment. Sci.* 184, 386–392.

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., and Coalson, T. (2012). Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases. *Cereb. Cortex* 22, 2241–2262. doi: 10.1093/cercor/bhr291

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., et al. (2017). The conception of the ABCD study: from substance use to a broad NIH collaboration. *Devel. Cogn. Neurosci.* 32, 4–7. doi: 10.1016/j.dcn.2017.10.002

Wager, T. D., Atlas, L. Y., Lindquist, M. A., Roy, M., Woo, C.-W., and Kross, E. (2013). An fMRI-based neurologic signature of physical pain. *N. Engl. J. Med.* 368, 1388–1397. doi: 10.1056/NEJMoa1204471

Wang, L., Zhou, J., and Qu, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* 68, 353–360. doi: 10.1111/j.1541-0420.2011.01678.x

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., et al. (2014). "StudentLife," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing – UbiComp '14 Adjunct*, (New York, N Y: ACM Press), 3–14. doi: 10.1145/2632048.2632054

Wang, R., Wang, W., daSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., et al. (2018). Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1–26. doi: 10.1145/3191775

Wang, R., Wang, W., DaSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., et al. (n.d.). Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proc. ACM Interact. Mob.Wearable Ubiquitous Technol.* 2:43.

World Health Organization (2018). *WHO Depression Fact Sheet*. Available at: https://www.who.int/news-room/fact-sheets/detail/depression.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 1125–1165. doi: 10.1152/jn.00338.2011

Zipunnikov, V., Greven, S., Shou, H., Caffo, B. S., Reich, D. S., and Crainiceanu, C. M. (2014). Longitudinal high-dimensional principal components analysis with application to diffusion tensor imaging of multiple sclerosis. *Ann. Appl. Stat.* 8, 2175–2202. doi: 10.1214/14-AOAS748

Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 25 June 2019
doi: 10.3389/fpubh.2019.00157

Toward Personalized Tinnitus Treatment: An Exploratory Study Based on Internet Crowdsensing

Jorge Simoes[1], Patrick Neff[1,2], Stefan Schoisswohl[1], Jan Bulla[1,3], Martin Schecklmann[1], Steve Harrison[4], Markku Vesala[4], Berthold Langguth[1] and Winfried Schlee[1]*

[1] Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany, [2] University Research Priority Program "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland, [3] Department of Mathematics, University of Bergen, Bergen, Norway, [4] Tinnitus Hub, Hemsworth, United Kingdom

Introduction: Chronic tinnitus is a condition estimated to affect 10–15% of the population. No treatment has shown efficacy in randomized clinical trials to reliably and effectively suppress the phantom perceptions, and little is known why patients react differently to the same treatments. Tinnitus heterogeneity may play a central role in treatment response, but no study has tried to capture tinnitus heterogeneity in terms of treatment response.

Research Goals: To test if the individualized treatment response can be predicted using personal, tinnitus, and treatment characteristics.

Methods: A survey conducted by the web platform Tinnitus Hub collected data of 5017 tinnitus bearers. The participants reported which treatments they tried and the outcome of the given treatment. Demographic and tinnitus characteristics, alongside with treatment duration were used as predictors of treatment outcomes in both an univariate as well as a multivariate regression setup. First, simple linear regressions were used with each of the 13 predictors on all of 25 treatment outcomes to predict how much variance could be explained by each predictor individually. Then, all 13 predictors were added together in the elastic net regression to predict treatment outcomes.

Results: Individual predictors from the linear regression models explained on average 2% of the variance of treatment outcome. "Duration of treatment" was the predictor that explained, on average, most of the variance, 6.8%. When combining all the predictors in the elastic net, the model could explain on average 16% of the deviance of treatment outcomes.

Discussion: By demonstrating that different aspects predict response to various treatments, our results support the notion that tinnitus heterogeneity influences the observed variability in treatment response. Moreover, the data suggest the potential of personalized tinnitus treatment based on demographic and clinical characteristics.

Keywords: tinnitus, heterogeneity, crowdsensing, smart device, personalized treatment

OPEN ACCESS

Edited by:
Claudio Eccher, Fondazione Bruno Kessler, Italy

Reviewed by:
Laszlo Balkanyi, Retired, Solna, Sweden
Pierpaolo Cavallo, University of Salerno, Italy

****Correspondence:***
Jorge Simoes
jorge.simoes@ukr.de

Specialty section:
This article was submitted to Digital Health, a section of the journal Frontiers in Public Health

Received: *08 February 2019*
Accepted: *29 May 2019*
Published: *25 June 2019*

Citation:
Simoes J, Neff P, Schoisswohl S, Bulla J, Schecklmann M, Harrison S, Vesala M, Langguth B and Schlee W (2019) Toward Personalized Tinnitus Treatment: An Exploratory Study Based on Internet Crowdsensing. Front. Public Health 7:157. doi: 10.3389/fpubh.2019.00157

1. INTRODUCTION

Tinnitus is a condition characterized by an auditory perception, usually in the form of ringing or hissing, for which there is no corresponding external source (1). The prevalence of tinnitus has been estimated between 10 and 15% in the adult population (2, 3). From those, one fifth will require clinical intervention (4). Additionally, the mean annual cost of illness was estimated at 6.8 billion euros globally (5). On the individual level, tinnitus may be accompanied by comorbidities such as insomnia, anxiety and depression, constituting a high burden to patients (6). Current clinical guidelines recommend that clinicians target those potential comorbidities, and although no treatment has shown efficacy in randomized clinical trials to reliably and effectively suppress the phantom perceptions, it is clear that various treatment options result in different degree of improvements—most likely because of the underlying heterogeneity of the etiology and pathophysiology of tinnitus (2, 7, 8). The clinical guidelines also recommend different management strategies for tinnitus, including, but not limited to, psycho-education, counseling, cognitive behavior therapy, hearing aids when assessed as necessary and sound therapy (2). Importantly, the current clinical understanding is that certain treatments may not be suitable/effective for all, and clinicians should recommend treatments to patients in an individual basis (8). Thus, albeit the low evidence levels for treatments on a group level, these same treatments may be beneficial in specific cases on the individual level.

From a clinical perspective, bothersome, chronic, and subjective tinnitus is a common and challenging form of tinnitus (2, 6). However, this form of tinnitus might be highly heterogeneous. In recent years, the notion of tinnitus as a complex, multi-faceted condition gained traction (9). For that reason, researchers and clinicians have drawn their attention to the different ways of tinnitus manifestation, including its etiology (e.g., sound blast, persistent loud noise exposure, whiplash, etc.), phenotype (e.g., type of sound perceived, laterality of the sound perception, presence of hearing loss, etc.), and accompanying comorbidities (e.g., insomnia, depression, anxiety, etc.). Such heterogeneity constitutes a complex puzzle that challenges both researchers and clinicians in their understanding of the pathophysiology of tinnitus and in the development of new treatments (1). Importantly, tinnitus heterogeneity may account for the low success rates of clinical trials at the group level, as well as why certain individuals respond positively to specific treatments (8, 10).

Noteworthy efforts to capture tinnitus heterogeneity include the studies from Langguth et al. (11), Tyler et al. (12) and Van den Berge et al. (13). Overall, the studies showed modest results without a clear delineation of tinnitus subtypes. However, those studies were limited due to sample size and/or homogeneous samples recruited from specialized tinnitus clinics. It is yet unclear how representative samples from tertiary clinics represent the whole tinnitus population; thus, we consider a broader data sample necessary to capture a yet unexplored facet of tinnitus heterogeneity (14).

Crowdsourced health research studies have been proposed as a mean to circumvent the difficulties experienced during patient's recruitment, such as the increased costs of adding participants to a study and the homogeneous sample representation from tertiary clinics (15). Crowdsourcing can be defined as the collaborative collection of data in which individuals and/or institutions participate voluntarily (15, 16). When the data is collected through mobile devices, such as smartphones, tablets, or wearable devices, the term crowdsensing is commonly used (17). The number of policy makers, health providers and academics using such technologies increased drastically in the last decade due to the ubiquity of mobile and sensing devices (18). Especially in tinnitus research, crowdsensing has been substantially used (14, 17, 19, 20). Importantly, such technologies may yield new insights about phenomena hardly accessible to traditional settings.

To the best of our knowledge, no study tried to capture tinnitus' heterogeneity using crowdsensing technology, especially in terms of treatment response. Our study aims to fill that research gap. We collected crowdsensed data from an online tinnitus self-help platform to explore tinnitus heterogeneity avoiding the aforementioned limitations during data collection, namely the reduced sample size and/or homogeneous patient representation. First, we investigated whether tinnitus heterogeneity could be expressed not only

TABLE 1 | Sample size of each treatment.

Treatment	*n*
Self Sound Stimulation	1,562
Supplements and Herbal	1,157
Antidepressants	785
Hearing Aid	681
Acunpuncture	621
Masker	503
Chiropractor	489
Homeopathic	425
Psychologist	388
Cognitive Behavior Therapist	371
Tinnitus Retraining Therapy	370
Steroids	346
Off-label Medication	312
Psychiatrist	298
Neurofeedback / Meditation	270
Books / self help	254
Gabaergic medication	237
Notched Music	223
Soundcure	144
Acoustic Neuromodulation	120
Neuromonics	95
Low Level Laser Therapy	65
Retigabbine	53
Hyperbaric Oxygen Therapy	46
Transcranial Magnetic Stim.	45

TABLE 2 | Sample's demographic and tinnitus characteristics.

Predictor	Levels	*n*	Percentage
Gender	Male	1,712	58.8%
	Female	1,181	40.5%
	Other	21	0.7%
Age	Under 18	13	0.4%
	18–24	162	5.6%
	25–34	364	12.5%
	35–44	427	14.7%
	45–54	606	20.8%
	55–64	869	29.8%
	65–74	405	13.9%
	75 +	58	2.0%
	Prefer not to say	10	0.3%
Tinnitus onset	Less than 3 months	147	5.0%
	4–6 months	156	5.4%
	6–12 months	293	10.1%
	1–2 years	427	14.7%
	2–3 years	359	12.3%
	3–5 years	347	11.9%
	5–10 years	388	13.3
	10–20 years	339	11.6%
	20 + years	458	15.7%
Noise reactiveness	Sounds have no affect	587	20.1%
	Some sounds make it a lot worse	627	21.5%
	Some sounds make it somewhat worse	354	12.1%
	Some sounds make it better and some make it worse	725	24.9%
	Some sounds make it somewhat better	212	7.3%
	Some sounds make it a lot better	113	3.9%
	NA	296	10.2%
Hyperacusis	No	1,006	34.5%
	Mildly	795	27.3%
	Moderately	776	26.6%
	Severely	291	10.0%
	NA	96	3.3%
Somatic	No	1,643	56.4%
	Yes	1,056	36.2%
	NA	215	7.4%
Jaw and neck problems	Problems with Jaw	261	9.0%
	Problems with Neck	503	17.3%
	Problems with Jaw and Neck	407	14.0%
	NA	1,743	59.8%
Hearing loss	Mild hearing Loss	1,265	43.4%
	Moderate hearing loss	400	13.7%
	Severe hearing loss	152	5.2%
	NA	1,097	37.6%
Laterality of hearing loss	Both ears	699	24.0%
	One ear	1,119	38.4%
	NA	1,096	37.6%
Tinnitus frequency	Low (<1 kHz)	152	5.2%
	Mid (1–3kHz)	151	5.2%
	Mid high (3–8 kHz)	525	18.0%
	Very high (8 kHz +)	350	12.0%
	Several dis in Hearing	77	2.6%
	Unsure	563	19.3%
	Na	1,096	37.6%
Perception of tinnitus	One ear	688	23.6 %
	Both ears	1,031	35.4 %
	More in the brain	204	7 %
	In the ears and brain	952	32.6 %
	Not sure	39	1.3 %
Perception of tinnitus during the day	Does not change at all	774	26.6 %
	Fluctuates, no pattern	1,369	46.9 %
	Fluctuates, better in the morning	131	4.5 %
	Fluctuates, better in the evening	626	21.4 %
	NA	14	0.4 %

in terms of phenotype, etiology and comorbidities as has previously been done, but also in terms of treatment response. To investigate this hypothesis, we modeled each predictor (i.e., tinnitus characteristics and demographics) individually as an independent variable on single linear regressions with treatment outcomes for 25 different treatments as dependent variables. Second, we investigated whether tinnitus heterogeneity could predict treatment response from demographic factors and tinnitus characteristics. We operationalized this hypothesis by combining all predictors in a statistical model to predict the outcome of treatments.

2. METHODS

Data for our sample were collected by Tinnitus Hub. Founded in 2015 by SH and MV, the Tinnitus Hub operates "Tinnitus Talk" (www.tinnitustalk.com), created in 2011, the largest online, anglophone self-help platform for tinnitus patients. The survey took place between February 8th and March 13th of 2016. Members of the forum received a link to the digital survey. We collected information of 5017 participants, from those 2916 reported trying at least one treatment and thus were included in the data set for the final analysis. It was not possible to obtain written informed consent from the users of Tinnitus Talk, but the "Terms and Rules" of the website informed the users that the collected data will be analyzed for scientific purposes. All the data were saved anonymously. A similar dataset was used in a former study (14).

Personal and tinnitus information was collected from participants of the survey alongside questions about which tinnitus-related treatments were tried and were used as independent values in our statistical models. In total, 13

factors were included in our analysis (**Table 2**). Additionally, participants were asked to rate how effective a given treatment was in reducing the distress and/or suppressing the noise perception, and the duration of the treatment retrospectively (1: "this treatment made my tinnitus much worse," 2: "this treatment made my tinnitus mildly worse," 3: "this treatment had no effect on my tinnitus," 4: "this treatment made my tinnitus slightly better," and 5: "this treatment made my tinnitus much better").

Our analysis included the outcome of 25 different treatments and used as dependent variables in our statistical model. Participants consented to have their anonymous data used for scientific research. Simple linear regressions were performed for individual predictors (i.e., demographics and tinnitus characteristics, and treatment duration) on treatment outcomes (i.e., dependent variable). Regressions were weighted based on the number of treatments that patients tried and *p*-values were adjusted for multiple comparisons using Hommel correction (21, 22). Collinearity was assessed with the variance inflation factor (VIF). The VIF is the ratio of variance in a model with multiple predictors, divided by the variance of a model with one predictor alone (23). The high VIF values in our models indicated that models containing all 13 demographic factors and tinnitus characteristics as predictors would contain high collinearity. To address this issue, we used elastic net regularization (24). Elastic net accounts for collinearity by penalizing the coefficients in the model either by shrinking their values or by setting them to 0 (24). We ran a n-fold cross validated elastic net to estimate the optimal lambda (i.e., one of the penalizing coefficients from elastic net) over 11 different alpha values ranging from 0 (i.e., RIDGE regression) to 1 (i.e., Lasso Regression). For this analysis, the predictors encoded as factors were converted into dummy variables as a prerequisite from the statistical software. We

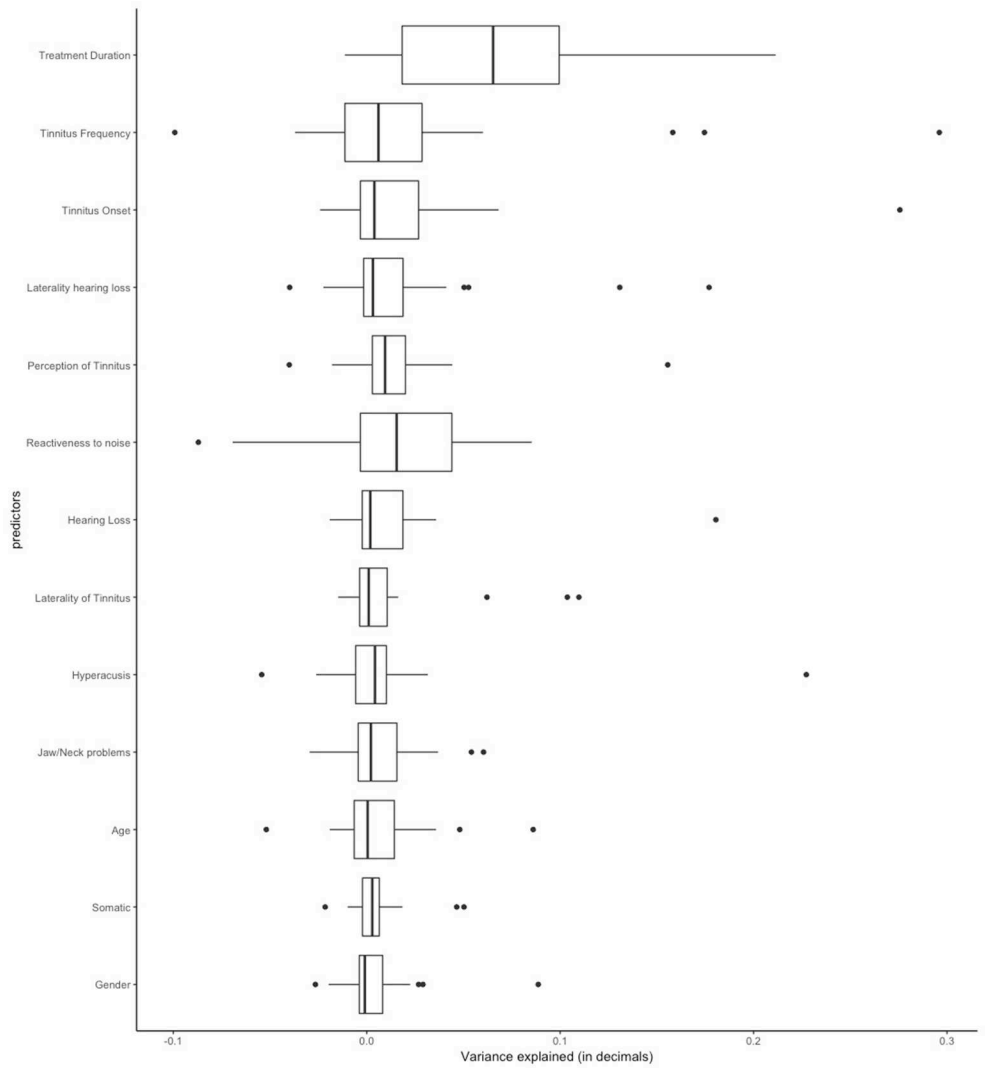


FIGURE 1 | Amount of variance explained in the linear regression models by each predictor across all different 25 treatments.

selected the models with minimized mean squared error for our final analysis.

All statistical analysis was conducted with R statistical software (25), alongside the "tidyverse" package (26). Power analysis were calculated using the "effsize" package (27) and the elastic net was performed by the "GLMnet" package (24). Non-parametric tests were used when statistical assumptions of parametric tests were not met. *P*-values below 0.05 were considered statistically significant.

3. RESULTS

Table 1 shows the frequency of each treatment in our sample. Clinical and demographic characteristics of the sample are summarized in **Table 2**. First, we applied linear regression models with individual predictors as independent variables on the self-reported treatment outcomes as dependent variables. The aim of this analysis was to test how much variance could be explained by individual predictors for the different treatments. **Figure 1** shows the average amount of variance explained by each predictor on all 25 different treatments. A summary of all statistical models can be found in the **Supplementary Materials**. The amount of variance explained by single predictors over all treatments was 2% on average.

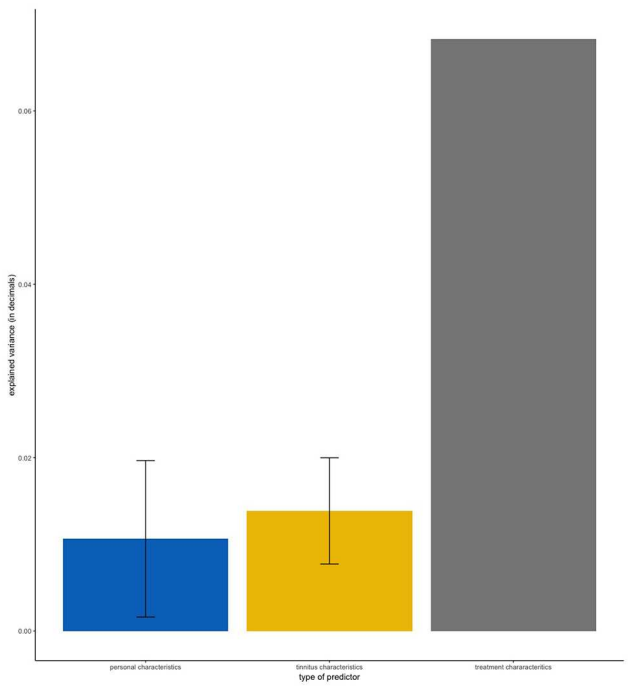


FIGURE 2 | Mean amount of variance explained by type of predictor. Error bars represent standard deviation. "Personal Characteristics" contains the predictors Age, Gender, and Tinnitus Onset. "Tinnitus Characteristics" contains the predictors Tinnitus Frequency, Laterality of Hearing Loss, Perception of Tinnitus, Reactiveness to Noise, Hearing Loss, Laterality of Tinnitus, Hyperacusis, and Jaw/Neck Problems. "Treatment Characteristics" contains the predictor Treatment Duration.

Next, we investigated what type of predictor could explain most of the variance of treatment outcomes. For this analysis, we grouped predictors in three groups: personal, tinnitus and treatment characteristics (**Figure 2**). Personal and tinnitus characteristics could explain, on average, the same amount of variance.

As shown in **Figures 1**, **2**, the predictor "Duration of Treatment" explained on average more variance than the remaining predictors (p <0.05). To further explore the relationship between treatment duration and treatment outcome, we clustered the average treatment outcomes based on their duration. The results can be found in **Figure 3**, where our analysis of variance showed no trend of time over treatment outcome ($p = 0.99$).

Next, we fitted all predictors as independent variables and self-reported treatment outcomes as the dependent variable in our elastic net regression model. This analysis aimed to measure how much of the deviance on treatment outcomes can be explained by combining all analyzed items. **Figure 3** shows the amount of deviance explained by all predictors for each of the 25 treatments. On average, 16% of the deviance could be explained by all predictors combined. **Table 3** summarizes which predictors were considered statistically significant by the elastic net and linear regressions respectively.

Lastly, we conducted one exploratory analysis based on the coefficients obtained by both models to identify clinical markers of treatment success. From coefficients estimated by linear regression, we observed that participants who reported responding positively to sounds (i.e., rating a 4 or 5 in the Likert scale) reported more frequently benefiting positively to treatments with an acoustic component. Thus, we subset only patients who reacted positively to sounds and divided treatments with and without an acoustic component (**Figure 4**). Our group mean comparison analysis corroborated our data-driven hypothesis, as patients who reported reacting positively to sounds also reported higher outcomes with treatments with an acoustic component ($p = 0.02$, Cohen's d = 1.07).

4. DISCUSSION

In this study we investigated whether personal, tinnitus, and treatment characteristics collected from an internet self-help platform population can be used to explain which patients are responding to different treatments. Similar attempts to predict treatment outcomes with patients' characteristics have been tried in a spectrum of mental conditions, including lower back pain (28), depression (29), post traumatic stress disorder (30), obsessive-compulsive disorder (31), substance abuse (32), and tinnitus itself (33). To the best of our knowledge, this is the first study attempting to capture tinnitus' heterogeneity in terms of a wide range of treatment responses using crowdsensing technology. Moreover, whereas most studies tried to predict the outcome of a single treatment, our study aimed to predict the outcome of 25 different treatments.

Our results showed that 2% of the variance of treatment outcomes could be explained, on average, by individual

TABLE 3 | Predictors identified as significant by the elastic net model (X) and by linear regressions (O).

	Duration of treatment	Laterality of hearing loss	Fluctuation of sound perception	Noise reactiveness	Jaw/Neck problems	Onset	Age	Hyperacusis	Gender	Tinnitus frequency	Hearing loss	Somatic	Laterality of tinnitus
Acoustic Neuromodulation	X	X	X	X	X	X	X	X	X	X	X		
Hearing aid	X/O	X/O	X	X/O	X/O	X/O	X	X	X	X	X/O		O
Self Admin. Sound Therapy	X/O	X	X/O	X/O	X	X	X	X	X/O	X			
TRT	X/O	X	X	X	X	X	X/O	X		X	X/O		
Antidepressants	X/O	X	X/O	X	X	X	X/O	X/O		X			
Soundcure	X/O	X	X	X		X	X	X	X	X/O			
Psychiatrist	X	X	X	X		X	X	X	X		X		
Psychologist	X	X	X/O	X		X	X	X	X		X		
Supplements/Herbal admin.	X	X	X	X	X		X			X	X		
Homeopathic admin.	X	X	X	X	X	X	X		X				
GABA admin.	X/O			O	X/O		X/O	X		X	X		
In ear masker	X/O	X	X	O				X		X	X		
Acunpuncture	X	X	X		X/O	X	X		X/O				
Hyperbaxic Oxygen Therapy	X	X			X	X		X		X			
Notched music	X/O	X/O	X	X					X				
Off Label Medication admin.	X		X	X	X	X							
Self learning	X	X		X	X	X							
CBT	X		X/O					X	O				
Chiropractor	X				X								
Neurofeedback	X/O						X						
Steroids admin.					X	X							
LowLevelLaser Therapy	X												
Neuromonics	X												
Retigabine admin.		X											
Transcranianl Magnetic Stim.		X											

Coefficients associated with significant predictors can be found in ***Supplementary Materials****.*

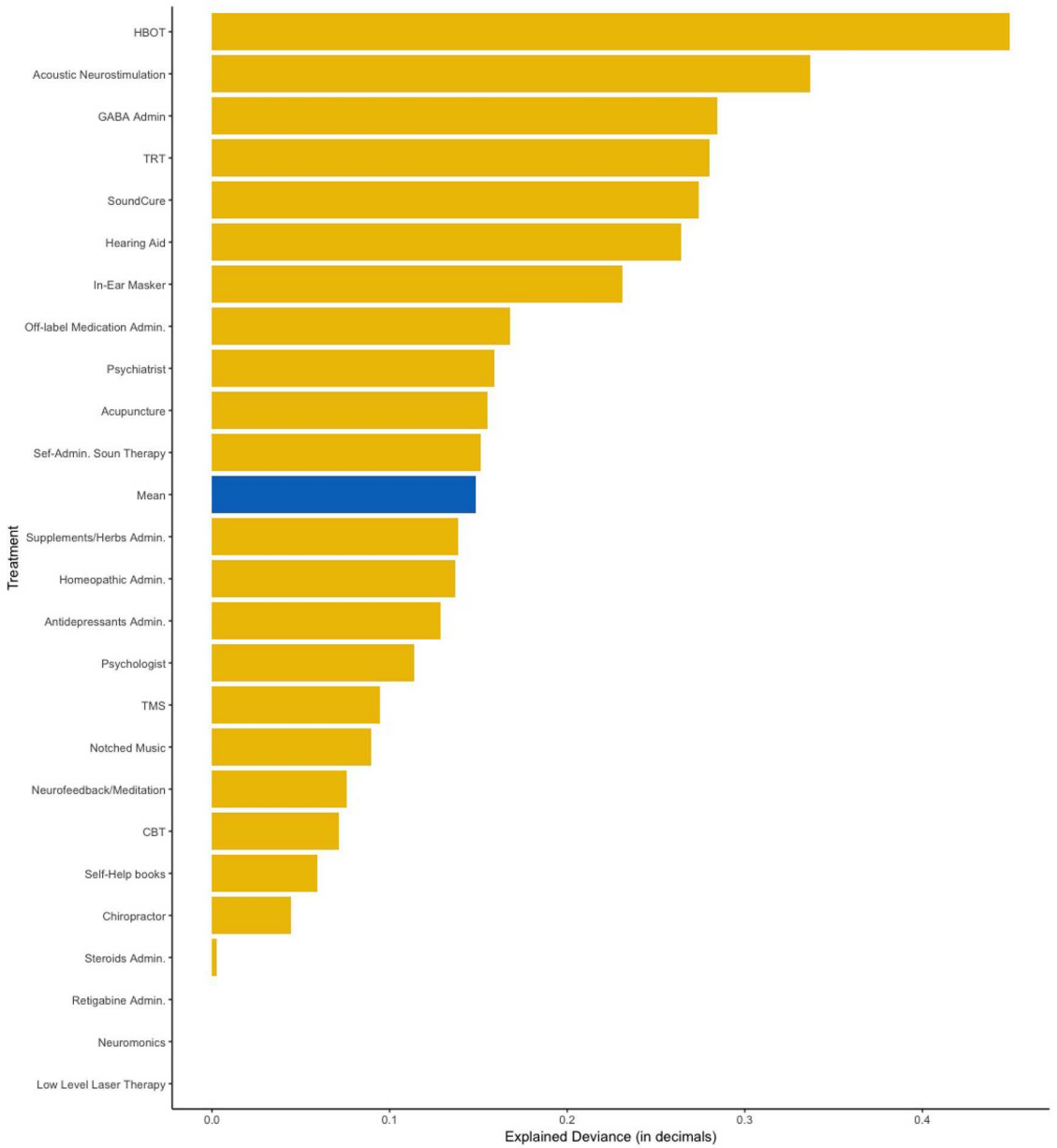


FIGURE 3 | Amount of variance explained by the Elastic Net model with all the 13 predictors added simultaneously. HBOT, Hyperbaric Oxygen Therapy; TRT, Tinnitus Retraining Therapy; TMS, Transcranial Magnetic Stimulation; CBT, Cognitive Behavior Therapy.

predictors (**Figure 2**). Additionally, our analysis showed that both personal characteristics and tinnitus characteristics, despite being significant predictors for multiple treatments (**Table 3**), could explain little variance on average. At first glance, it seems that the analyzed parameters have only a small impact on treatment outcome, but the average amount of deviance explained by the elastic net combining all 13 predictors into a single model was 16%, after accounting for covariance. We identified multiple statistically significant predictors in both regression setups (**Table 3**), but the individual amount of variance they could explain was limited. These results suggest that although no single predictor is paramount to predict the treatment outcomes, personal, tinnitus, and treatment characteristics may have a predictive role when combined. Altogether, those characteristics could be used in the future to predict treatment responsiveness in tinnitus, especially after better markers of treatment success are identified. For instance, our analysis did not include information about patients' personality, depression or tinnitus-related distress, nor did it collect information of the sequence in which treatments were tried or whether treatments were tried simultaneously.

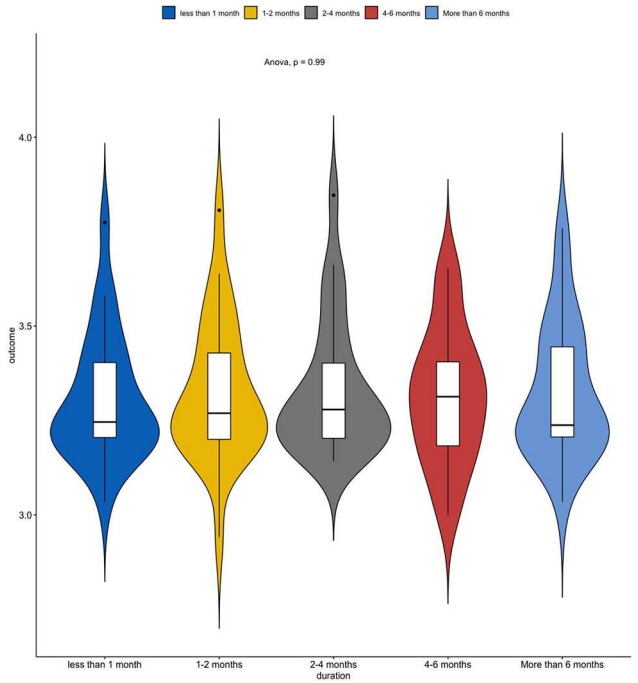


FIGURE 4 | Mean treatment outcomes on a 1–5 Likert scale clustered by treatment duration.

Capturing tinnitus heterogeneity has been proposed as an important clinical and scientific goal, but previous attempts obtained limited results (12, 13). Importantly, tinnitus heterogeneity may explain why only a subset of patients are responding to specific treatments (10). A broader comprehension of tinnitus, encompassing not only demographics and tinnitus characteristics, but also treatment response, could, for example, explain the limited treatment efficacy seen in clinical practice (2). For instance, it is yet unclear whether previous successful or unsuccessful treatments have any predictive power on the outcomes of future treatments. Ultimately, the subtyping of tinnitus could lead to personalized care, a long-standing request by both clinicians and patients (6). Our results, though modest, suggest that personalized treatment for tinnitus patients based on patients' personal, tinnitus, and treatment characteristics should be feasible.

One example of future implications that this type of analysis could lead to, is the effect of noise reactiveness in the outcomes of treatments with and without an acoustic component (**Figure 5**). Our results suggest that participants whose tinnitus respond positively to sounds tend to benefit more from treatments with an acoustic component than from treatment without such component. Although future studies should try to replicate these results, we believe that the insights from large data sets such as these could have meaningful effects in tinnitus care and research. For instance, such insights could help researchers define new, fine-grained inclusion criteria for future clinical trials in acoustic-based treatments.

Regarding treatment duration, the predictor that could, on average, explain most of the variance, did not show any statistically significant difference between time periods. These results should be interpreted with caution as it is well-known that certain treatments, such as cochlear implants, require some time for adaptation whereas other treatments, such as antidepressants, require longer periods to be effective. Nonetheless, our results support the notion that the duration of treatment is not inherently beneficial or detrimental to the treatment's efficacy.

Our study comes with some inherent limitations. First, we did not have access to information about treatments which were performed in an overlapping span of time, thus we were unable to account for possible interaction between treatments. Second, our outcome measure was retrospective and subjective, which could have biased the results. We consider a subjective metric, although coarser than an objective one such and the Tinnitus Handicap Inventory, adequate for this type of analysis given the multiple treatments that a single patient tried and the sometimes-long period of time between the administration of a treatment and the survey. Nevertheless, further prospective studies analyzing outcome predictors would be desirable. Third, although we examined 25 different treatments, this number was insufficient to capture the whole complexity of available interventions for tinnitus treatments. Cognitive Behavior Therapy (CBT), for example, can be performed in a span of days or months, sessions can be individual or in group, a wide range of techniques can be applied in each session, etc. Such variety of treatment details and subtypes were not exclusive to CBT, but rather a commonality across treatments. Fourth, we chose a limited number of potential predictors for the survey, but we might have missed other important items. Particularly we would expect that there may exist further items that may be relevant for response to some of the investigated treatments. Finally we are aware that the investigated sample, albeit large and international, might not be representative of all patients with tinnitus (14).

5. CONCLUSION

Our results suggest that tinnitus heterogeneity could be expressed in terms of treatment response. The variance explained by individual predictors on treatment outcomes suggests that specific traits could explain why certain people are responding positively to a given treatment. In the future, especially with the availability of "big" multi-faceted data, a better understanding of the factors involved in treatment responsiveness could lead to individualized, optimal tinnitus management.

ETHICS STATEMENT

The data set was collected in 2016 through a survey in the tinnitus hub online forum (https://www.tinnitushub.com), and was shared to the authors.

AUTHOR CONTRIBUTIONS

JS wrote the manuscript. JS, WS, and JB defined the study design and interpreted the results. PN, MS, SS, and BL interpreted the results, and provided critical feedback during the review. SH and MV were responsible for data collection.

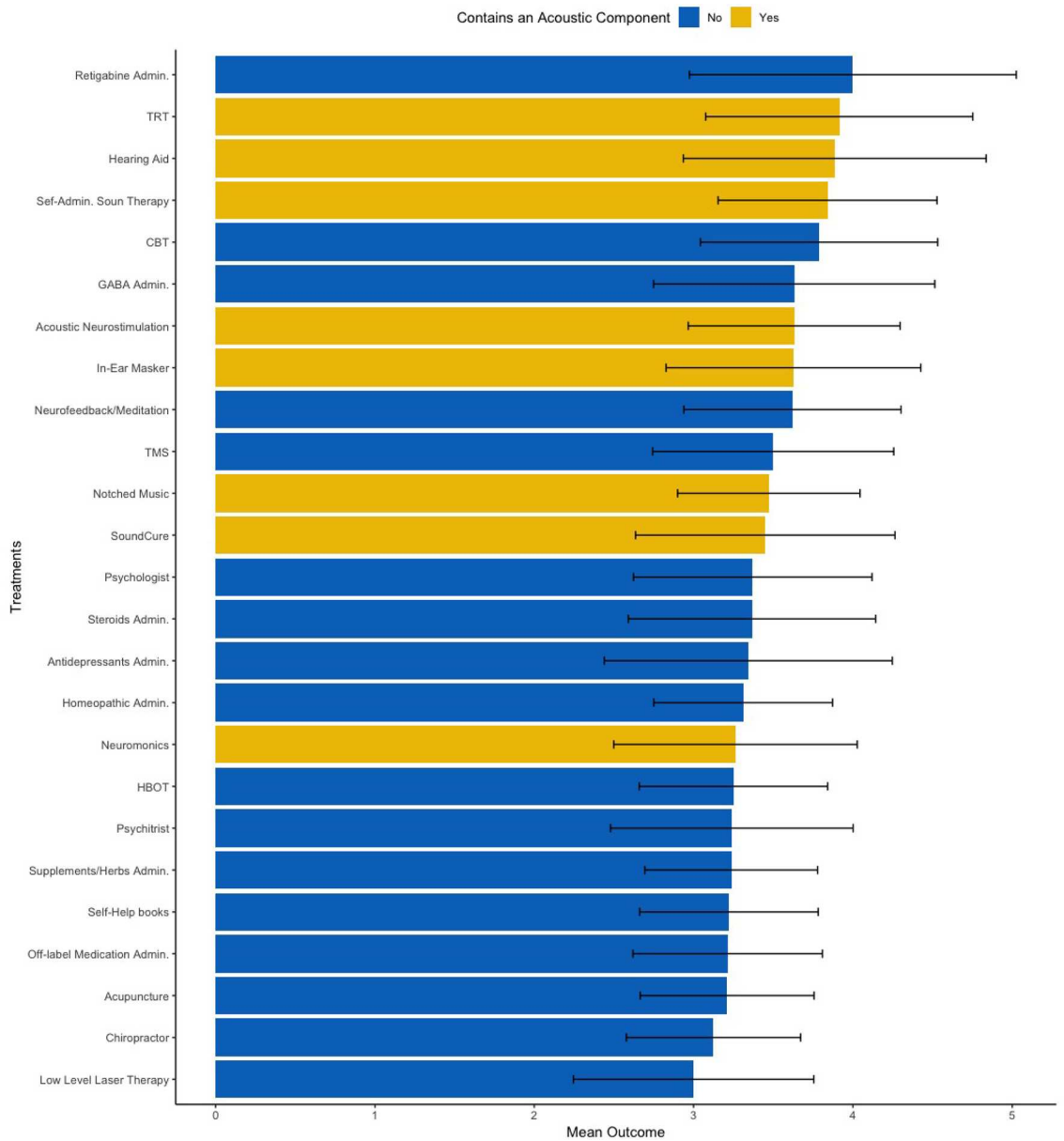


FIGURE 5 | Mean treatment outcomes on a 1–5 Likert scale clustered by treatments with an acoustic component (yellow) and without an acoustic component (blue). Error bar accounts for the standard deviation across all 25 treatments. TRT, Tinnitus Retraining Therapy; CBT, Cognitive Behavior Therapy; TMS, Transcranial Magnetic Stimulation; HBOT, Hyperbaric Oxygen Therapy.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 722046.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2019.00157/full#supplementary-material

REFERENCES

1. Baguley D, McFerran D, Hall D. Tinnitus. *Lancet*. (2013) 382:1600–7. doi: 10.1016/S0140-6736(13)60142-7
2. Tunkel DE, Bauer CA, Sun GH, Rosenfeld RM, Chandrasekhar SS, Cunningham ER, et al. Clinical Practice Guideline: Tinnitus. *Otolaryngol Head Neck Surg*. (2014) 151:S1–40. doi: 10.1177/0194599814545325
3. McCormack A, Edmondson-Jones M, Somerset S, Hall D. A systematic review of the reporting of tinnitus prevalence and severity. *Hear Res*. (2016) 337:70–9. doi: 10.1016/j.heares.2016.05.009
4. Henry JA, Zaugg TL, Myers PJ, Schechter MA. The role of audiologic evaluation in progressive audiologic tinnitus management. *Trends Hear*. (2008) 12:170–87. doi: 10.1177/1084713808319941
5. Maes IHL, Cima RFF, Vlaeyen JW, Anteunis LJC, Joore MA. Tinnitus: a Cost Study. *Ear Hear*. (2013) 34:508–14. doi: 10.1097/AUD.0b013e31827d113a
6. Langguth B, Kreuzer PM, Kleinjung T, De Ridder D. Tinnitus: causes and clinical management. *Lancet Neurol*. (2013) 12:920–30. doi: 10.1016/S1474-4422(13)70160-1
7. Zenner HP, Delb W, Kröner-Herwig B, Jäger B, Peroz I, Hesse G, et al. A multidisciplinary systematic review of the treatment for chronic idiopathic tinnitus. *Eur Arch Otorhinolaryngol*. (2017) 274:2079–91. doi: 10.1007/s00405-016-4401-y
8. Langguth B, Elgoyhen AB, Cederroth CR. Therapeutic approaches to the treatment of tinnitus. *Annu Rev Pharmacol Toxicol*. (2019) 59:291–313. doi: 10.1146/annurev-pharmtox-010818-021556
9. Schlee W, Hall DA, Canlon B, Cima RF, de Kleine E, Hauck F, et al. Innovations in doctoral training and research on tinnitus: the European School on Interdisciplinary Tinnitus Research (ESIT) Perspective. *Front Aging Neurosci*. (2018) 9:447. doi: 10.3389/fnagi.2017.00447
10. Cederroth CR, Gallus S, Hall DA, Kleinjung T, Langguth B, Maruotti A, et al. Editorial: towards an understanding of tinnitus heterogeneity. *Front Aging Neurosci*. (2019) 11:53. doi: 10.3389/fnagi.2019.00053
11. Langguth B, Landgrebe M, Schlee W, Schecklmann M, Vielsmeier V, Steffens T, et al. Different patterns of hearing loss among tinnitus patients: a latent class analysis of a large sample. *Front Neurol*. (2017) 8:46. doi: 10.3389/fneur.2017.00046
12. Tyler R, Coelho C, Tao P, Ji H, Noble W, Gehringer A, et al. Identifying tinnitus subgroups with cluster analysis. *Am J Audiol*. (2008) 17:S176–84. doi: 10.1044/1059-0889(2008/07-0044)
13. van den Berge MJC, Free RH, Arnold R, de Kleine E, Hofman R, van Dijk JMC, et al. Cluster analysis to identify possible subgroups in tinnitus patients. *Front Neurol*. (2017) 8:115. doi: 10.3389/fneur.2017.00115
14. Probst T, Pryss RC, Langguth B, Spiliopoulou M, Landgrebe M, Vesala M, et al. Outpatient tinnitus clinic, self-help web platform, or mobile application to recruit tinnitus study samples? *Front Aging Neurosci*. (2017) 9:113. doi: 10.3389/fnagi.2017.00113
15. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res*. (2012) 14:e46. doi: 10.2196/jmir.1988
16. Wazny K. "Crowdsourcing" ten years in: a review. *J Global Health*. (2017) 7:020602. doi: 10.7189/jogh.07.020602
17. Pryss R, Reichert M, Herrmann J, Langguth B, Schlee W. Mobile crowd sensing in clinical and psychological trials – A case study. In: *2015 IEEE 28th International Symposium on Computer-Based Medical Systems*. São Carlos; Ribeirão Preto: IEEE (2015). p. 23–4.
18. Ma H, Zhao D, Yuan P. Opportunities in mobile crowd sensing. *IEEE Commun Magaz*. (2014) 52:29–35. doi: 10.1109/MCOM.2014.6871666
19. Pryss R, Probst T, Schlee W, Schobel J, Langguth B, Neff P, et al. Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus-stress associations based on the TrackYourTinnitus mobile platform. *Int J Data Sci Anal*. (2018) 12:1–12. doi: 10.1007/s41060-018-0111-4
20. Muniandi LP, Schlee W, Pryss R, Reichert M, Schobel J, Kraft R, et al. Finding tinnitus patients with similar evolution of their ecological momentary assessments. In: *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. Karlstand (2018). p. 112–7.
21. Hommel G. A comparison of two modified Bonferroni procedures. *Biometrika*. (1989) 76:624–5. doi: 10.1093/biomet/76.3.624
22. Shaffer JP. Multiple hypothesis testing. *Annu Rev Psychol*. (1995) 46:561–84. doi: 10.1146/annurev.ps.46.020195.003021
23. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Vol. 112. Springer (2013).
24. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. (2010) 33:1–22. doi: 10.18637/jss.v033.i01
25. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (2018). Available online at: https://www.R-project.org (accessed 02 May, 2019).
26. Wickham H. *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. (2017). Available online at: https://CRAN.R-project.org/package=tidyverse (accessed 02 May, 2019).
27. Torchiano M, Torchiano MM. Package "effsize" version 0.7.4. (2017). Available online at: https://cran.r-project.org/web/packages/effsize/effsize.pdf (accessed 02 May, 2019).
28. Kleinke CL, Spangler AS Jr. Predicting treatment outcome of chronic back pain patients in a multidisciplinary pain clinic: methodological issues and treatment implications. *Pain*. (1988) 33:41–8. doi: 10.1016/0304-3959(88)90201-1
29. Jarrett RB, Eaves GG, Grannemann BD, Rush AJ. Clinical, cognitive, and demographic predictors of response to cognitive therapy for depression: a preliminary report. *Psychiatry Res*. (1991) 37:245–60. doi: 10.1016/0165-1781(91)90061-S
30. Karatzias A, Power K, McGoldrick T, Brown K, Buchanan R, Sharp D, et al. Predicting treatment outcome on three measures for post-traumatic stress disorder. *Eur Arch Psychiatry Clin Neurosci*. (2007) 257:40–6. doi: 10.1007/s00406-006-0682-2
31. Cavedini P, Riboldi G, D'Annucci A, Belotti P, Cisima M, Bellodi L. Decision-making heterogeneity in obsessive-compulsive disorder: ventromedial prefrontal cortex function predicts different treatment outcomes. *Neuropsychologia*. (2002) 40:205–11. doi: 10.1016/S0028-3932(01)00077-X
32. McLellan AT, Luborsky L, Woody GE, O'Brien CP, Druley KA. Predicting response to alcohol and drug abuse treatments: role of psychiatric severity. *Arch Gen Psychiatry*. (1983) 40:620–5. doi: 10.1001/archpsyc.1983.04390010030004
33. Kröner-Herwig B, Zachriat C, Weigand D. Do patient characteristics predict outcome in the outpatient treatment of chronic tinnitus? *Psychosoc Med*. (2006) 3:Doc07.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

SYSTEMATIC REVIEW
published: 24 January 2020
doi: 10.3389/fnins.2020.00003

Efficacy of Self-Management Smartphone-Based Apps for Post-traumatic Stress Disorder Symptoms: A Systematic Review and Meta-Analysis

Andreas Goreis[1,2]*, *Anna Felnhofer*[3], *Johanna Xenia Kafka*[4], *Thomas Probst*[5] and *Oswald D. Kothgassner*[4]

[1] Department for Clinical and Health Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria, [2] Outpatient Unit for Research, Teaching and Practice, Faculty of Psychology, University of Vienna, Vienna, Austria, [3] Department of Pediatrics and Adolescent Medicine, Medical University of Vienna, Vienna, Austria, [4] Department of Child and Adolescent Psychiatry, Medical University of Vienna, Vienna, Austria, [5] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, Austria

OPEN ACCESS

Edited by:
Waldemar Karwowski, University of Central Florida, United States

Reviewed by:
James Douglas Bremner, Emory University, United States
Olga Santesteban-Echarri, University of Calgary, Canada
Susanne Schuett, University of Vienna, Austria

***Correspondence:**
Andreas Goreis
andreas.goreis@univie.ac.at

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *21 March 2019*
Accepted: *06 January 2020*
Published: *24 January 2020*

Citation:
Goreis A, Felnhofer A, Kafka JX, Probst T and Kothgassner OD (2020) Efficacy of Self-Management Smartphone-Based Apps for Post-traumatic Stress Disorder Symptoms: A Systematic Review and Meta-Analysis. Front. Neurosci. 14:3. doi: 10.3389/fnins.2020.00003

Post-traumatic stress disorder (PTSD) symptoms are prevalent in both civilian and military service members. As the number of smartphone-based applications (apps) grows rapidly in health care, apps are also increasingly used to help individuals with subthreshold PTSD or full PTSD. Yet, if the apps are self-managed, the feasibility and efficacy of such interventions are still rather unclear in these two populations with PTSD symptoms. Hence, the present meta-analysis set out to evaluate the effect of self-management smartphone-based apps on PTSD and depressive symptoms in populations with subthreshold PTSD or full PTSD. Studies were included if they conducted randomized controlled trials or pre-post comparisons. Six studies ($n = 2$ randomized controlled trials) were identified for meta-analysis. In pre-post comparisons, $N = 209$ participants were included in the analyses. In randomized controlled trials, $N = 87$ participants received smartphone-based self-management interventions and $N = 82$ participants were in waitlist control conditions. Meta-analysis for pre-post comparisons concluded an effect of $g = 0.55$ ($p < 0.001$) regarding the overall reduction in PTSD symptoms ($n = 6$) and $g = 0.45$ ($p < 0.001$) for reduction in depressive symptoms ($n = 5$). Yet, in randomized controlled trials, no significant difference was found between app-based treatment and waitlist control groups ($g = 0.09$, $p = 0.574$). The duration of the interventions did not significantly influence the results. Overall, despite positive pre-post effects, current results indicate that smartphone-apps for PTSD patients are not significantly more effective than waitlist control conditions. Nevertheless, a combined smartphone and standard therapy approach may be a fruitful field for future research.

Keywords: smartphone app, PTSD, post-traumatic stress disorder, mHealth, trauma intervention, depression, meta-analysis, mobile phone intervention

INTRODUCTION

Post-traumatic stress disorder (PTSD) is a cause of substantial disability in both civilian and military populations, leading to long-term problems for individuals, families, and society in terms of compromised emotional well-being, productivity loss, and high cost of treatment (Kessler, 2000; Breslau et al., 2004; Buckley et al., 2004; Cohen et al., 2009; Kok et al., 2012; Marmar et al., 2015). PTSD is characterized by a multitude of symptoms resulting from exposure to one or more traumatic events (World Health Organization, 1993; American Psychiatric Association, 2013). Individuals with PTSD are typically affected by anhedonia, emotional numbness, social detachment, unresponsiveness to external stimuli, insomnia, and suffer from hyperarousal (Elhai and Palmieri, 2011; e.g., Armour et al., 2016). The experience of traumatic events is also associated with elevated symptoms of depression and anxiety (Etkin and Wager, 2007; Mandelli et al., 2015).

PTSD has an estimated lifetime prevalence ranging from 2% in Europe (Darves-Bornoz et al., 2008; Maercker et al., 2008) to 7% in the United States (Kessler et al., 2005), and 4% in a cross-national study of 24 countries (Koenen et al., 2017). Furthermore, a significant number of individuals experience symptoms of subthreshold or subclinical PTSD in response to traumatic events that do not meet diagnostic criteria (Brancu et al., 2016). Subthreshold PTSD has previously been identified to affect around 20% of U.S. veterans returning from Afghanistan (Hoge et al., 2006) and prevalence estimates for civilian populations are mirroring—at least—the prevalence rates of those with full PTSD (Stein et al., 1997; Marshall et al., 2001; Breslau et al., 2004; Bergman et al., 2015). Research has shown that levels of distress and functional impairment are significantly heightened for individuals with subthreshold PTSD (e.g., Mylle and Maes, 2004), underscoring the fact that both subthreshold PTSD and the full PTSD cause impairment and represent considerable public health concerns (Bergman et al., 2015).

Evidence-based treatments are available for PTSD (Foa et al., 2009), and guidelines generally recommend exposure therapy and cognitive therapies, and pharmacological treatment as an adjunct treatment (for an overview of psychological treatments see Cusack et al., 2016). Literature on treatment options for subthreshold PTSD is limited (Dickstein et al., 2013), but treatment with lower intensity may be favorable (Shiner et al., 2012; Korte et al., 2016). Many affected individuals, however, remain without treatment due to negative beliefs about efficacy, stigma, logistic reasons, or shortage of qualified treatment centers in the adjacent geographic region (Hoge et al., 2004; Shalev et al., 2011; Kazdin and Rabbitt, 2013). Early and accessible interventions are equally important in subthreshold PTSD, as 25% of those affected develop the full PTSD (Marshall et al., 2001; Breslau et al., 2004; Cukor et al., 2010).

Innovative technology, such as applications (apps) for smartphones, can address the need for accessible and effective interventions after traumatic experiences, especially on a population level (Cernvall et al., 2018). Smartphones are carried by a majority of adults with ownership rates ranging between 77% in the U.S. (Pew Research Center, 2018) and 79% in the European Union (Eurostat, 2016). Promisingly, no ethnic disparities exist in smartphone ownership in U.S. adults (Pew Research Center, 2018) and applications for smartphones could be a feasible means of reaching minority populations with a possibly limited access to health care (López et al., 2012). Applications allow individuals to approach specific treatments at their own pace, individually, and confidentially, which may result in greater acceptance and compliance (Juarascio et al., 2014). Emerging evidence suggests that smartphone applications improve depression and anxiety symptoms (Donker et al., 2013; Firth et al., 2017), health behaviors such as physical activity, diet (Schoeppe et al., 2016), smoking cessation (Whittaker et al., 2016), and reduces alcohol consumption (Gustafson et al., 2014). Preliminary results also exist for potential benefits in patients with schizophrenia (Firth and Torous, 2015) and eating disorders (Juarascio et al., 2014).

Based on this, a multitude of applications, which specifically target subthreshold PTSD have been developed. In a literature review of mobile health apps for PTSD, Rodriguez-Paras et al. (2017) found 45 publicly available PTSD-specific apps in their recent review and they stated that minimal effort and transparency has been made regarding development, usability, and validation of this plethora of apps. The PTSD Coach app, for example, was jointly developed by the U.S. Department of Veterans Affairs' and the Department of Defense, providing users with self-management, psychoeducative elements concerning PTSD symptoms and treatment, symptom monitoring, and coping skills (U. S. Department of Veterans Affairs, 2011a,b; Possemato et al., 2016). PTSD Coach is available for iOS and Android devices, and preliminary studies reported a high satisfaction and acceptance among veteran (Kuhn et al., 2014) and community samples (Miner et al., 2016). Another app, PE Coach (U. S. Department of Veterans Affairs, 2017a,b), was also developed by the U.S. Department of Veteran Affairs and provides psychoeducation, symptom tracking, and—optionally—support features to improve patient compliance (e.g., appointment reminders, audio recordings, imaginal exposure homework). It was previously utilized to support users who were in primary care settings or receiving therapy (Reger et al., 2013, 2015). Some studies have been conducted to test the efficacy of these applications for individuals with (subthreshold) PTSD (Miner et al., 2016; Possemato et al., 2016; e.g., Kuhn et al., 2017). Results were promising, with moderate to large effects ($d = 0.78$) regarding the reduction of PTSD-symptoms post-intervention in the PTSD Coach group when compared to a waitlist-condition (Miner et al., 2016). In another study, 57% of PTSD Coach users reported a reduction of PTSD symptoms compared to 26% in a waitlist condition (Kuhn et al., 2017). In both studies, however, the two groups did not differ significantly in PTSD or depressive symptoms post treatment. Yet, sample sizes for the PTSD Coach condition were small in both studies ($n = 25$ in Miner et al., 2016; $n = 62$ in Kuhn et al., 2017), possibly impeding significant differences to be detected. Similar patterns emerged in Cernvall et al. (2018) with 11 participants, pre-post effect sizes for the reduction of symptoms were moderate for PTSD and depressive symptoms ($d = 0.51$ and $d = 0.58$, respectively), but both failed to reach nominal significance. As symptoms of depression and anxiety often have profound effects on affected individuals

that overlap and co-occur with PTSD symptoms (Norris et al., 1997; e.g., Luxton et al., 2010) it is of additional interest to investigate the efficacy of smartphone-based apps on depressive and anxiety symptoms.

It is discernable that this field of research is underpowered and conclusions about the benefits of smartphone-based applications cannot be drawn on single trials alone. A recent study (Wickersham et al., 2019) reviewed the efficacy of mobile interventions, both self-managed and with clinician support, for the treatment of PTSD symptoms in randomized controlled trials (RCTs) and found inconclusive yet promising results, with a decrease of symptoms in app-based treatments, but not compared to control groups. To evaluate the efficacy of self-managed apps alone, further granulation and meta-analysis of individual studies is needed. We therefore present a meta-analysis on all available studies assessing the effects of self-management smartphone-based applications for PTSD treatment. The aim of the present meta-analysis is two-fold: (1) to conduct a meta-analysis of studies reporting the effect of self-managed mobile application on PTSD symptoms, and (2) to conduct a meta-analysis of studies reporting the effect of mobile applications on depression and anxiety symptoms as secondary outcome variables.

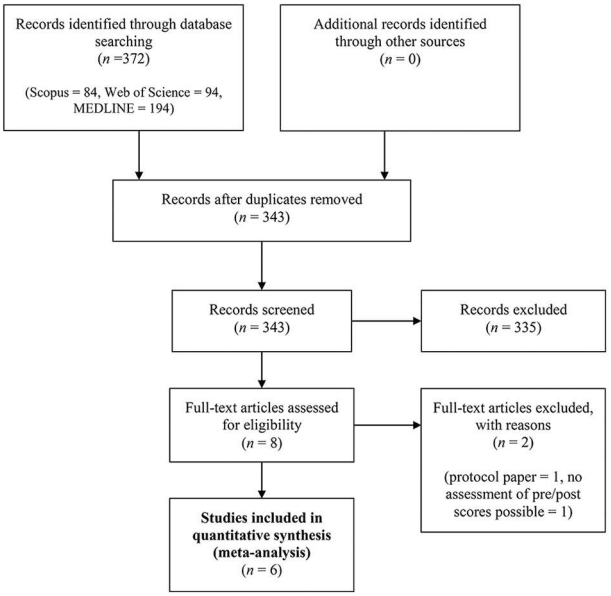


FIGURE 1 | PRISMA flowchart of screening, exclusion, and inclusion criteria.

METHOD

Search Strategy and Inclusion Criteria

A search of MEDLINE, Scopus, and Web of Science was conducted using the keywords "PTSD OR trauma OR posttraumatic-stress disorder AND Smartphone OR App OR Application OR mobile phone" from the beginning of database records until January 2019. Studies were eligible to be included in the meta-analysis if they (i) conducted randomized controlled trials with waitlist controls or (ii) pre-post studies assessing the effect of self-management smartphone-based apps on PTSD symptoms. No other inclusion or exclusion criteria were applied. No limitations on language or publication status were invoked. We additionally coded and analyzed symptoms of depression and anxiety if they were reported. Furthermore, Google Scholar alerts were enabled to ensure inclusion of accepted articles and articles in preprint, and authors were contacted to ensure inclusion of unpublished studies. Two Authors (ODK and JXK) independently examined the title, abstract, and main text of each study and full text papers were obtained where necessary to evaluate inclusion. Any discrepancies were discussed by the two authors. Final inclusion was based on the following criteria:

(1) Participants: Individuals with varying severity of PTSD symptoms as indicated by self-report questionnaires or via clinical interview conducted by a psychologist or physician.
(2) Intervention: Self-managed smartphone-based apps.
(3) Comparison: Studies with and without control groups were included.
(4) Outcomes: Reported at least a PTSD symptom severity score before and after the intervention.
(5) Study design: Pre-post studies or randomized controlled trials.

Exclusion of documents occurred at each stage (see **Figure 1** for PRISMA flow diagram and **Supplementary Table 3** for PRISMA checklist). The initial search generated 343 results. After the article selection process, six studies were identified and included in our meta-analysis.

Data Extraction and Analysis

To analyze the effect of app-based interventions from pre to post, we computed the standardized mean difference (Hedges' g) of PTSD-symptoms, depressive symptoms, and anxiety symptoms based on means and standard deviations (Dunlap et al., 1996) before and after the app-based intervention. We used the formula $d = (M_{pre} - M_{post})/SD_{pooled}$, where M_{pre} is the mean of the measure before the intervention and M_{post} after the intervention, with SD_{pooled} as the standard deviation for both measurements, defined as $SD_{pooled} = SQRT(SD^2_{pre} + SD^2_{post})/2$ (Lakens, 2013). For the standardized mean difference between intervention and control groups as indicator of the efficacy of the intervention in randomized-controlled trials, we calculated Cohen's d for the post-intervention scores, based on means and standard deviations, with the formula $d = (M_{Intervention} - M_{Control})/SD_{pooled}$, with the respective means of measurements for the intervention and control groups. To investigate changes from baseline separately in the intervention and control groups of the RCTs, we also computed the above-mentioned effect sizes for pre-post changes. Means, standard deviations and sample sizes were retrieved and entered into a spreadsheet. The calculations of the effect sizes and the subsequent meta-analysis were then conducted using the package metafor for R (Viechtbauer, 2010), which automatically corrects Cohen's d for a potential positive bias in small samples, yielded the effect size Hedges' g (Hedges, 1981). Following general convention

(Cohen, 1988), an effect size of 0.20 was considered a small effect, 0.50 a moderate effect, and 0.80 a large effect. Random effects models were applied to estimate aggregated effect sizes (Borenstein et al., 2011). Heterogeneity across study outcomes was reported with I^2 values, where 0 to 40% might not be important, 30 to 60% may represent moderate heterogeneity, and 50 to 90% may represent substantial heterogeneity (Higgins and Green, 2011).

Egger's regressions were conducted to analyze indications for publication bias (Sterne and Egger, 2005). Trim-and-fill analyses were calculated to provide estimates for adjusted effect sizes and, based on funnel plot asymmetry, numbers of imputed missing studies (Duval and Tweedie, 2000). Publication bias can be tested by entering data in a funnel graph (a plot of dispersion between study effect and a measure of study size). A symmetrical inverted distribution of the studies around the mean effect size represented in the funnel would indicate an absence of publication bias. Moderator analysis (meta-regression) was calculated to test whether the durations of interventions (in weeks) moderate the effect of the self-management app-based interventions on PTSD and depressive symptoms. The alpha level was set at 5% for all analyses. All data and codes are stored on a repository of the Open Science Framework (doi: 10.17605/OSF.IO/DZJT7).

Risk of Bias Assessment

We assessed risk of bias for each study using predefined criteria based on the AHRQ Method Guide for Comparative Effectiveness Reviews (Viswanathan et al., 2018). Therefore, categories regarding randomization, selection and attrition bias, confounding bias, measurement bias and statistical problems were included for coding. We rated all studies according to low, moderate or high risk of bias. Results assessed as having low risk of bias are considered to be valid, moderate risk of bias indicate some risk of bias, but probably this does not invalidate its results, a high risk indicates significant issues with design, measurement, conduct or analysis, all of which probably invalidates the results. We predefined that inappropriate methods of randomization, no control for confounding factors high attrition ≥40% or differential loss ≥30%, problems in participant selection and adequate statistical power are reasons for high risk of bias ratings. However, we rated grades of overall strength of evidence (SOE) according to Owens et al. (2010) for all studies as displayed in **Table 1**. The supplemental materials (**Supplementary Tables 1**, **2**) deliver an overview concerning the coding categories and risk of bias assessments. The assessments were independently determined by two investigators (AG and ODK); disagreements between the two investigators were discussed.

RESULTS

Study Characteristics

The six studies included in our meta-analysis covered data from 209 participants in self-management app-based intervention groups and 82 in control groups. All study samples included persons with both PTSD and subthreshold PTSD. Three studies (Possemato et al., 2016; Roy et al., 2017; Tiet et al., 2019) included samples of military service members, the remainder evaluated participants from the general population. All studies were conducted in the U.S., with the exception of Cernvall et al. (2018), which was conducted in Sweden. Additionally, all studies used the same application (PTSD Coach, U. S. Department of Veterans Affairs, 2011a,b), with the exception of Roy et al. (2017), who provided their sample with a multitude of applications with varying content (e.g., LifeArmor and PE Coach for psychoeducation concerning prolonged exposure, Tactical Breather for breathing exercises, Eventful to facilitate positive social engagement). See **Table 1** for detailed study characteristics and SOE assessments for each study.

Four of these six studies were included as pre-post comparisons (Possemato et al., 2016; Roy et al., 2017; Cernvall et al., 2018; Tiet et al., 2019) and two were included as randomized controlled trials with waitlist control conditions (Miner et al., 2016; Kuhn et al., 2017). Two studies (Possemato et al., 2016; Roy et al., 2017) had a randomized controlled design, but only pre-post comparisons were included to be in line with the aim of the present meta-analysis, i.e., to examine the effect of self-management apps. One study (Possemato et al., 2016) randomly assigned participants to either self-managed or clinician-managed PTSD Coach conditions ($n = 10$ per condition). The clinician managed condition received four 20-min sessions (via phone) which focused on providing instructions for app use, setting goals for symptom reduction, and assigning activities between sessions (Possemato et al., 2016). In order to assure cross-study comparability, we only included the self-managed PTSD Coach condition in which no support by a clinician was provided in our meta-analysis as a pre-post comparison. Roy et al. (2017) compared the efficacy of an app-based intervention supported by daily brief text messages with elements of resilience enhancement and cognitive-behavioral therapy to a self-management control group without such support. As the aim of the present meta-analysis was to evaluate the effect of self-management app-based interventions, we included only the self-management group of the study by Roy et al. (2017) as a pre-post comparison in our meta-analysis.

All included studies used the DSM-IV based PTSD checklist (PCL) in either the civilian or specific versions (Weathers et al., 1994, 2001; Weathers and Ford, 1996) to assess PTSD symptoms. Four studies assessed depressive symptoms with the Patient Health Questionnaire Depression Scale (PHQ-9; Kroenke et al., 2001), one study used the PHQ-8 (Kroenke et al., 2009). Except for Roy et al. (2017), none of the studies assessed symptoms of anxiety. Therefore, we were not able to meta-analytically evaluate the effects of smartphone apps on anxiety symptoms.

The study by Owen et al. (2015) was excluded although PTSD symptoms were measured using the PCL-C via the app; the authors analyzed data from users who had downloaded and used the app between 2012 and 2014 ($N = 3{,}462$) and, thus, had aggregated over 12,449 sessions. Yet, sample characteristics during the time points of assessment were not

TABLE 1 | Characteristics of the six studies included in the meta-analysis.

Study	Country	Sample	Treatment Group			Control Group*			App	Duration	Design	SOE
			n	Age *M* (*SD*)	% male	*n*	Age *M* (*SD*)	% male				
Cernvall et al. (2018)	Sweden	General population with full or partial PTSD (according to CAPS-5)	11	38.6 (Range 32–55)	27	–	–	–	PTSD Coach	4 weeks	Pre-test post-test design	low
Kuhn et al. (2017)	USA	General population with PCL-C score > 34 (subthreshold)	62	39.43 (15.16)	26	58	39.12 (14.08)	36	PTSD Coach	3 months	RCT with waitlist control group	high
Miner et al. (2016)	USA	General population with PCL-C score > 24 (subthreshold)	25	whole sample: 45.7 (13.9)	16	24	–	21	PTSD Coach	1 month	RCT with waitlist control group	moderate
Possemato et al. (2016)	USA	Veterans with PCL-S score > 40 (subthreshold)	10	42 (12)	95	–	–	–	PTSD Coach	2 months	RCT with clinician-support control group	low
Roy et al. (2017)	USA	Military service members and relatives with PCL score > 27 (subthreshold)	72	33.97 (10.8)	50	–	–	–	LifeArmor, PE Coach, Eventful, Positive Activity Jackpot, Tactical Breather, Daily Yoga, Simple Yoga	6 weeks	RCT with clinician-support control group	moderate
Tiet et al. (2019)	USA	Military service members with PC-PTSD score > 2 (probable PTSD)	29	Median: 61	97	–	–	–	PTSD Coach	4 months	Pre-test post-test design	moderate

*The description of Tiet et al. (2019) is the treatment arm without clinician support. SOE, Strength of Evidence; CAPS-5, Clinician-Administered PTSD for DSM-5 (Weathers et al., 2017). PCL, PTSD Checklist; PC-PTSD, Primary Care-PTSD Screen (Prins et al., 2004). *Only RCTs with waitlist control groups are reported.*

readily available, making it unfeasible to calculate effect sizes for meta-analysis. Mean scores for the PCL-C changed in the study by Owen et al. (2015) from $M = 57.2$ ($SD = 15.7$) at the first session to $M = 55.1$ ($SD = 16.6$) at individual return sessions. Reger et al. (2015) subjected two active-duty military service members with a current diagnose of PTSD to 8 weeks of prolonged exposure treatment, half of the duration with the support of PE Coach and the other half without the app. Since the participants in this study were both receiving prolonged exposure treatment and Reger et al. (2015) used a crossover design, it was not possible to isolate the effects of self-administer app. Participants, however, indicated higher levels of satisfaction concerning the weeks in which they were supported by the app.

Effects of Self-Management App-Based Interventions on PTSD Symptoms (Pre-post Comparisons)

Six effect sizes covering 209 participants were extracted to calculate the overall effect, operationalized in changes in PCL scores before and after the intervention. Meta-analysis concluded an effect of $g = 0.55$ (CI 0.29–0.80, $p < 0.001$) regarding the reduction in PTSD symptoms post intervention. Low heterogeneity between studies was found ($I^2 = 31.47$, $Q(5) = 6.38$, $p = 0.271$). Meta-regression did not reveal a significant coefficient for the duration of the intervention on PTSD symptoms ($b = -0.02$, $SE = 0.03$, $p = 0.622$). See **Figure 2** for forest plot.

Effects of Self-Management App-Based Interventions on Depressive Symptoms (Pre-post Comparisons)

Five effect sizes covered the changes in PHQ scores of 184 participants before and after the intervention. Meta-analysis revealed an effect of $g = 0.45$ (CI 0.24–0.65, $p < 0.001$). Low heterogeneity between studies was found for depressive symptoms ($I^2 = 0.58$, $Q(4) = 2.52$, $p = 0.642$). Furthermore, meta-regression did not reveal a significant coefficient for the duration of the intervention on depressive symptoms ($b = 0.01$, $SE = 0.03$, $p = 0.629$). See **Figure 3** for forest plot.

Efficacy of Self-Management App-Based Interventions in Randomized Controlled Trials

Two studies (Miner et al., 2016; Kuhn et al., 2017; overall $N = 169$) compared app-based interventions to waitlist control-groups in randomized controlled trials. Meta-analysis of post-treatment scores in PTSD symptoms of these two studies resulted in no significant difference between app-based treatment and waitlist groups ($g = 0.09$ [CI −0.22–0.39], $p = 0.574$). No

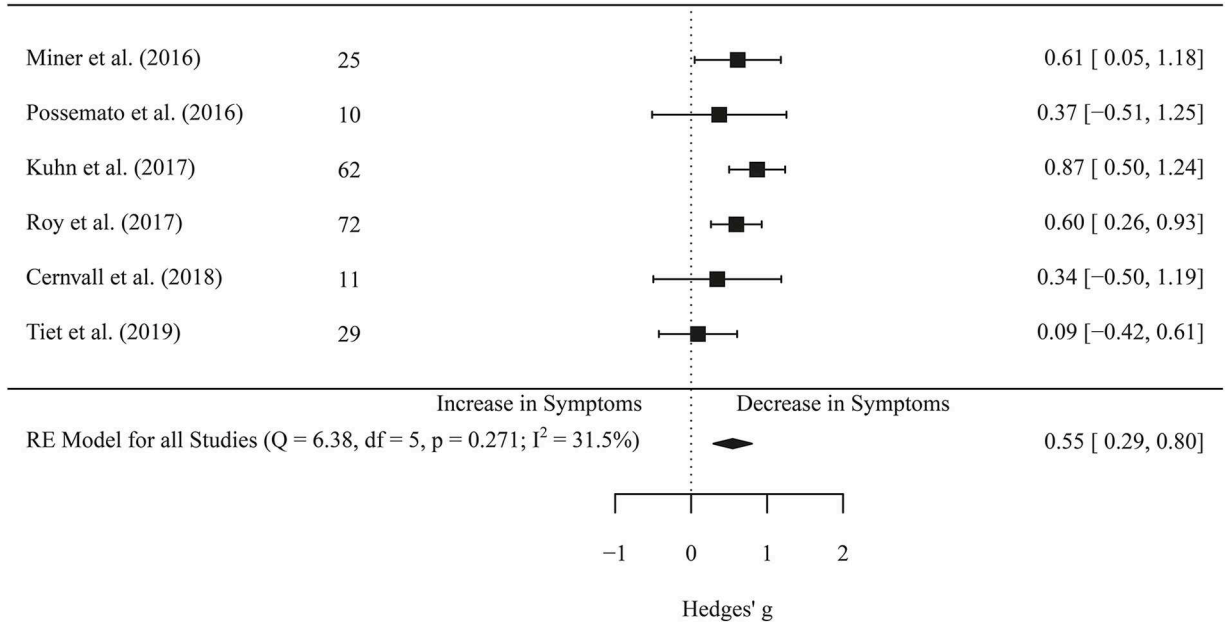


FIGURE 2 | Forest plot of the standardized mean difference (Hedges' *g*) of the effect of self-management smartphone-based apps on PTSD symptoms (pre-post changes). A positive effect size indicates that the PTSD symptoms decreased at the post measurement.

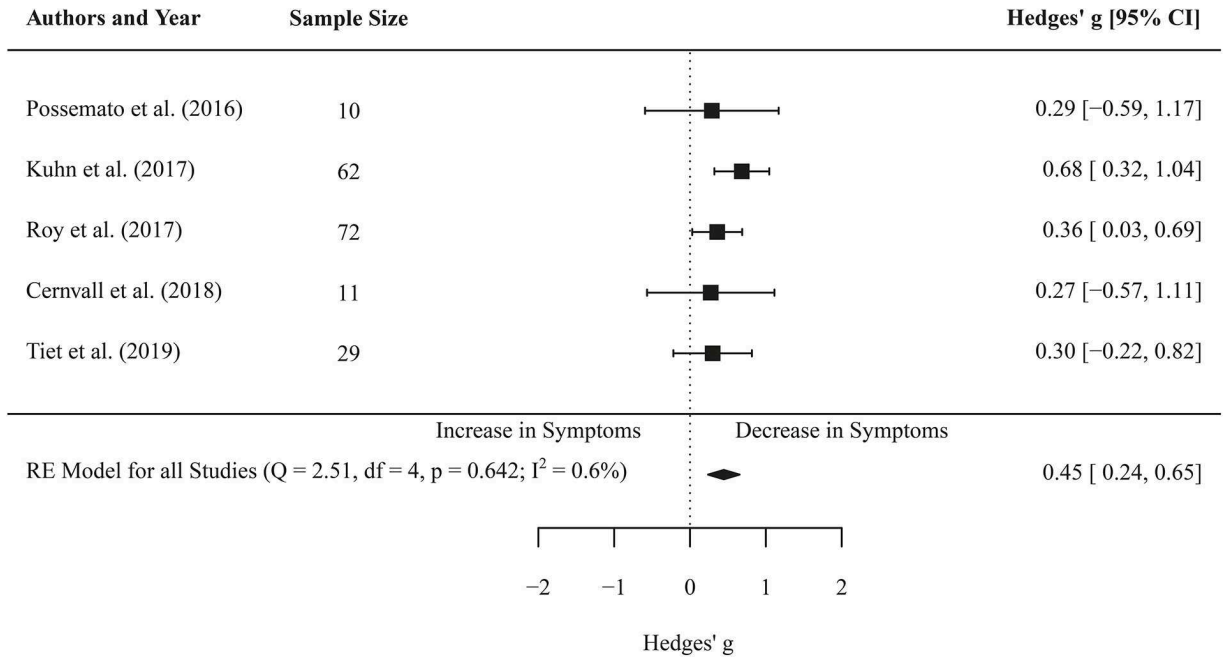


FIGURE 3 | Forest plot of the standardized mean difference (Hedges' *g*) of the effect of self-management smartphone-based apps on depressive symptoms (pre-post changes). A positive effect size indicates that the depressive symptoms decreased at the post measurement.

heterogeneity was found between the two studies ($I^2 = 0.00$, $Q(1) = 0.30$, $p = 0.584$; results not shown). Interestingly, meta-analysis concluded an effect post treatment of $g = 0.47$ for PTSD symptom reduction in waiting list controls compared to an effect post treatment of $g = 0.79$ in the treatment groups. Studies were rated with moderate-high SOE.

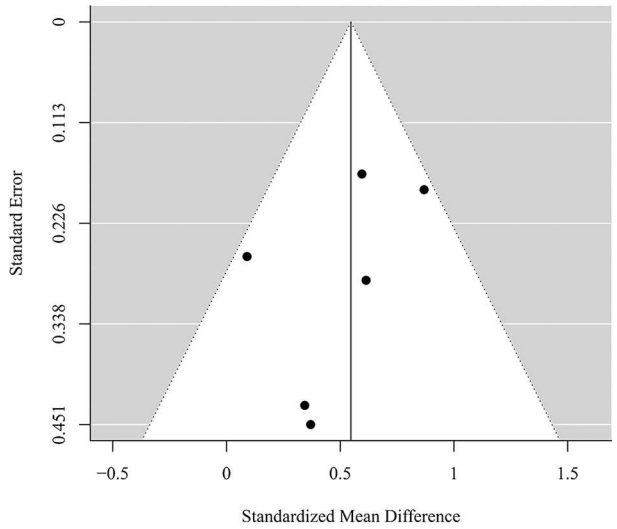


FIGURE 4 | Funnel plot for the meta-analysis of the effect of self-management smartphone-based apps on PTSD symptoms.

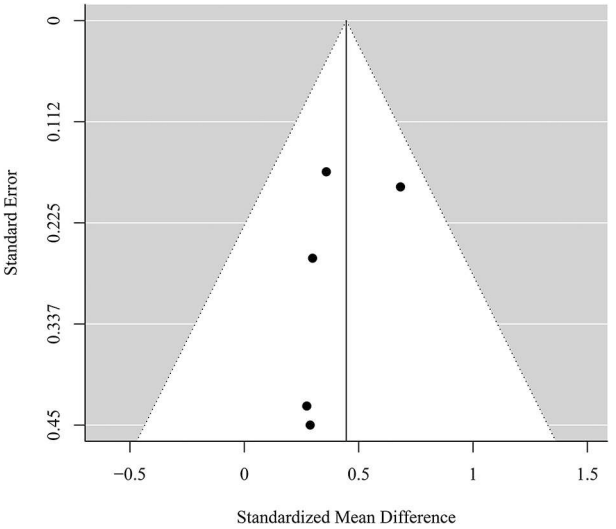


FIGURE 5 | Funnel plot for the meta-analysis of the effect of self-management smartphone-based apps on depressive symptoms.

Publication Bias and Risk of Bias Assessment

Visual inspection of the funnel plots (see **Figures 4**, **5**) did not suggest a publication bias in the present meta-analysis. Results for Egger's regression for funnel plot asymmetry were not significant both for the analysis of PTSD symptoms ($z = -1.09$, $p = 0.277$) and the analysis of depressive symptoms ($z = -0.67$, $p = 0.503$). No adjustments were needed according to the trim-and-fill analysis (no studies added left of the summary effect) in both analyses. This suggests no indication for publication bias in the present meta-analysis. Studies are heterogeneous regarding strengths of evidence in overall quality of evidence assessment. Our review revealed that majority of studies showed high or moderate risk of bias as presented in **Figure 6**.

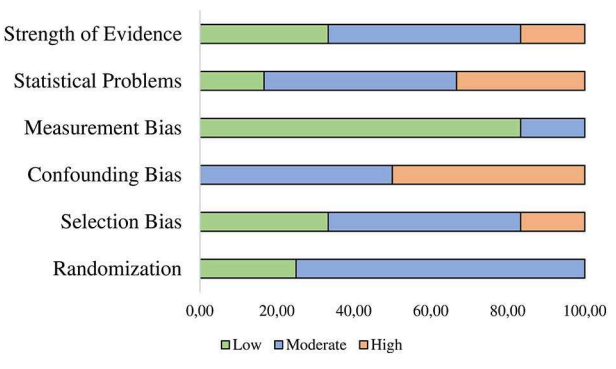


FIGURE 6 | Graphical Representation of the Risk of Bias Assessment.

DISCUSSION

In light of the ever increasing, promising use of innovative technologies in the context of treatment, the current meta-analysis set out to systematically analyze the effect of self-management smartphone-based applications as a means of intervention in populations with PTSD. Six studies with an overall sample of 209 participants with both subthreshold and full PTSD who used one or more self-management applications as an intervention were included in the meta-analysis. PTSD as well as anxiety and depressive symptoms were used as outcomes.

In the overall sample, self-management smartphone-based applications showed a moderate effect size ($g = 0.55$) for the reduction of PTSD symptoms (assessed with the PCL) post treatment. In the two included RCTs with waitlist controls, however, no significant decrease in PTSD symptoms was found after the intervention ($g = 0.09$). Regarding depressive symptoms (assessed with the PHQ), the overall effect was $g = 0.47$, bordering on a moderate effect size. A separate analysis for depressive symptoms in RCTs was not possible, as they were not assessed in these trials. In addition, the effect of self-management apps on anxiety symptoms could not be analyzed as only one study (Roy et al., 2017) reported according scores. As anxiety symptoms are regarded a frequent comorbidity of PTSD (e.g., Ginzburg et al., 2010), it is crucial to systematically assess them in future controlled trials which evaluate the efficacy of PTSD interventions. This would allow for a more differentiated picture regarding the differential effect of according treatments on the reduction of anxiety.

Overall, the current results suggest that PTSD symptom severity is reduced while using self-management smartphone-based apps, yet, the factors to which these changes may be attributed remain unclear. The app-specific effect evaluated in the RCTs was not significant.

Unexpectedly, the results of our meta-analysis indicate that there is no difference between an app-based intervention and

waitlist control conditions regarding PTSD symptom severity post treatment. This might be due to the small number of RCT studies ($n = 2$) included in our analysis as well as to the high pre-post effect size of $g = 0.47$ for PTSD symptom reduction in the waitlist control group. A possible explanation for symptom reduction in the absence of treatment may be that the inclusion of a patient in a study often entails a beneficial shift in attentional focus. Even though no treatment is provided, the patient is still subject to repeated clinical assessments and receives support and information regarding his/her symptoms. Accordingly, Smith et al. (2007) found that patients improved significantly simply by monitoring their PTSD symptoms. This questions the efficacy of self-management applications and encourages further RCT research regarding smartphone-based apps, and furthermore, a deeper discussion of its usefulness as a stand-alone intervention. Moreover, the content of self-management apps used by most included studies was similar or even the same, limiting a possible generalization of the effects for other or future smartphone-based applications. Further research of content-based factors for treatment outcomes (e.g., level of interactivity, type of tasks such as relaxation tasks, self-monitoring tasks) would be beneficial for the field of smartphone-based therapy apps. Nevertheless, such research would be advantageous for all mobile applications in the context of psychological therapy. Hence, based on the current results, the conceptual integration of smartphone-based apps for self-management intervention in consisting therapies seems to be essential, as well as a further development to reach an exponentially higher efficacy with a combined treatment.

Moreover, findings suggest that depressive symptoms decrease during the use of smartphone-based apps. However, it was not possible to conduct additional meta-analysis to assess the app-specific effect on depression change in RCTs, as only one study (Kuhn et al., 2017) assessed depressive symptoms in a randomized control design. Kuhn et al. (2017) reported a reduction in depressive symptoms, yet—similarly to our meta-analysis of PTSD symptoms in RCTs—scores between the intervention and waitlist group did not differ at post treatment. Both the utilization and the prospect of being able to utilize apps appear to have a supportive, stress-buffering effect. This means, that the individual is protected against the detrimental consequences of stress over time through continuing support. Accordingly, a recent experimental study (Kothgassner et al., 2019a) succeeded in demonstrating a considerable stress buffering effect of virtually provided support compared with face-to-face support. Results indicate that acute stress regulation, negative emotions of shame and rumination—as essential markers for PTSD and depressive symptoms—improved when people received digitally mediated social support, yet this support was only effective in terms of stress buffering if participants thought it was provided by another person (via an avatar) and not by a computer (via an agent). Following this, it can be argued that the patients' assumption that they are being supported—either virtually or physically—by another human could be a crucial factor influencing the efficacy of innovative, interactive intervention apps and may limit the efficacy of apps providing only self-administered content without a supporting person.

Avatar-based technology facilitates several therapy approaches, as it can substitute face-to-face contact with a clinician. According to Rehm et al. (2016), two concepts exist of how to include avatars into therapy: On the one hand, the patient interacts with an avatar, this was used as an effective tool in Virtual Reality Exposure Therapy (e.g. Cárdenas and De La Rosa, 2012 for PTSD), and as the embodiment of a real clinician or a supporting tool for self-management technology (e.g., Pinto et al., 2016 for depressive symptoms). On the other hand, patients may represent themselves as a virtual avatar, either as representation of the self for assessment or to be involved in a therapy setting. As an avatar can be seen as digital representation of the self that may become part of a person's overall identity after a certain time (Bessière et al., 2007), it reflects a link to a person's personality, strengths, and impairments. Further, the matter of how individuals behave and interact via avatars can be used for assessment or therapeutic information. It has been shown that avatar preference of persons with traumatic events differs from persons without traumatic events and that there are differences between men and women with emotional or physical abuse regarding their choices of avatar characteristics. Women choose avatar characteristics to help others, while men tend to use aggressive features for their avatars (Kothgassner et al., 2020). Other studies already showed evidence for the effect of avatars as representations of the patient to assess PTSD symptoms through a computer-based avoidance task (Myers et al., 2016; e.g., Allen et al., 2017).

In sum, it is—at this point—difficult to deduce specific recommendation for future apps from the current results since all but one study (Roy et al., 2017) have used the same self-management app. The PTSD Coach entails four modules including psychoeducative elements (about the disorder itself as well as about treatment options and family relations), the option to track symptoms (i.e., in the form of repeated assessments of related thoughts and emotions), symptom management tasks (e.g., stress relief) as well as a feature for receiving support (e.g., in the event of crisis). In line with the idea of self-management, this app offers only limited interactivity with another person (e.g., psychologist, friends, peers etc.). Based on the consideration that social resources (e.g., involvement of significant others in the treatment process) and virtual social support (see above, Kothgassner et al., 2019a) may show particularly beneficial effects on treatment outcome (see Heaney and Israel, 2008), we may, with caution, suggest the inclusion of more social interactive elements in future apps, be it in the form of actual interactions (via chat, voice recordings, video etc.) or via a pre-programmed virtual human which implies the presence of another person. Being accompanied by an avatar throughout the online treatment process has proven beneficial in past studies (see Rehm et al., 2016 for a review). Further, this lack of knowledge regarding the design of smartphone-based therapy applications strengths the need for including therapy naïve and experienced patients in the development for future therapy applications.

In general, the effects of self-management smartphone-based intervention apps are smaller compared with the effects found by another meta-analysis on more established and evidence-based interventions for PTSD like prolonged exposure therapy (PE) ($g = 1.08$ for PTSD symptoms; see Powers et al., 2010). Similarly, studies in the field of child and adolescent trauma-focused cognitive behavioral therapy (tf-CBT) showed a higher effect ($d = 0.88$ for PTSD symptoms; see Goldbeck et al., 2016) compared to a waiting list control group. However, compared with another technology-mediated therapy approach—the Virtual Reality exposure therapy—larger effect sizes are found for PTSD and depression symptom reduction compared to waiting list controls in a recent quantitative review ($g = 0.62$; $g = 0.50$; see Kothgassner et al., 2019b). For the current results, the inclusion of subthreshold and full PTSD is a clear limitation, because self-management smartphone-apps may be helpful and supportive for people with experienced trauma and mild symptoms, but not for full PTSD. According to this, it is necessary to state that the inclusion of patients with subthreshold PTSD strongly limits comparability to other studies including only full PTSD patients for treatment. However, this is a major point for future original studies investigating smartphone-based interventions. In light of the present results, self-management smartphone-apps might be a supportive intervention, but not a stand-alone solution. Another shortcoming is the small number of studies included which made it impossible to evaluate the efficacy of self-management apps only via RCTs. Some studies were pilot trials and did not have randomized control groups, others did not have control conditions that would make comparisons feasible (e.g., treatment as usual or apps with clinician support). This was explicated by SOE ratings, showing only one study with high, yet two studies with low SOE. However, by including non-randomized studies and reporting an overall pre-post effect, we were able to analyze the efficacy of self-managed apps as a stand-alone intervention—-in a granulated manner—-with more confidence. Additionally, as the studies included in the meta-analysis predominantly used one specific smartphone-app for treatment it was not possible to compare different solutions and designs. This hinders generalization for all smartphone-app approaches treating PTSD symptoms.

Furthermore, third variables, which are not possible to control for, might have influenced the extracted effect sizes concerning PTSD and depressive symptoms. For instance, both the duration and the daily use of the applications seem to be vital for the method's success (Henson et al., 2019). Although we did not find a moderating effect of duration, it was not possible to test for the actual use of applications in the daily life due to a lack of consistent reporting. Only few authors assessed use of applications in self-report; here, individuals indicated that they used the mobile app between 2.27 (Kuhn et al., 2017) and 2.65 times a week (Miner et al., 2016). The interventions' duration ranged between fo0ur and 12 weeks in the included studies, and it did not explain heterogeneity neither in PTSD nor in depressive symptoms. Future studies should investigate the relation between frequency of usage and improvement of symptoms.

Standardizing treatment duration, frequency of usage, and comparing key outcomes to treatment-as-usual control groups in a randomized controlled design would certainly add to a better understanding of processes underlying the efficacy of smartphone-based intervention applications for PTSD for example by mediation analyses. Another open question pertains to the fact that, to date, it is unclear how patients with PTSD perceive health-related mobile apps in terms of usability and acceptability (Rodriguez-Paras et al., 2017). This, however, may be a crucial issue when it comes to patient compliance and adherence in the context of mobile health applications, in particular with regards to self-management but also regarding data protection and security concerns. Understanding these technologies and perceiving them as useful may be an essential prerequisite for an adequate usage by patients. Furthermore, the investigation of guided and unguided support via smartphone apps could be a future interest for research in PTSD treatment. Research synthesis already showed guidance as a beneficial feature in Internet- and mobile-based interventions and reveals that clinical qualification of the person providing guidance is surprisingly of minor importance (Baumeister et al., 2014). Furthermore, first results concerning Internet- and mobile-based interventions used as supportive, adjunct tools in face-to-face therapy (blended care) seem promising (e.g., in the context of depression, Berger et al., 2018).

CONCLUSION

The current meta-analysis found small-to-moderate pre-post effect sizes for the reduction of PTSD and depressive symptoms in an overall sample of 209 participants. Even though effects are smaller than those of typical evidence-based interventions and therapies for PTSD (Powers et al., 2010), smartphone-based apps—due to their reach and availability—have a considerable potential to become vital parts of treatment strategies and interventions for communities and military populations suffering from subthreshold PTSD. In particular, the option of assessing health data on a day-to-day basis and in an ecologically valid fashion would not only allow for pinpointed assessments of key symptoms in future. It would also add to more customized technology-based interventions with an improved interaction between patient needs and clinician resources.

The results of our study imply that a self-managed smartphone-based app is not superior to waitlist control. It might therefore not be recommended to use these tools as stand-alone interventions. Following recent research, the social component seems to be important in basic computer mediated as well as in more complex virtual social interactions. According to this, it is safe to assume that a professional social entity is needed for a significant impact on symptomatology (e.g., Kothgassner et al., 2019a), but further smartphone-based apps have the potential to enrich traditional therapy protocols. Currently, there is a definitive lack of research on combined treatments (traditional face-to-face therapy including mobile app interventions) in the field of PTSD treatment. Evaluating the benefits of such blended care approaches during PTSD therapy as well as in the context of ambulatory recovery seems to be a particularly fruitful field for future research.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

AG and OK wrote the first draft of the manuscript. JK and OK conducted the literature search and coded the studies. AG prepared the statistical procedures and analyzed the data. AF, JK, and TP contributed extensively to the first draft. All authors have approved the final manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2020.00003/full#supplementary-material

REFERENCES

Allen, M. T., Jameson, M. M., and Myers, C. E. (2017). Beyond behavioral inhibition: a computer avatar task designed to assess behavioral inhibition extends to harm avoidance. *Front. Psychol.* 8:1560. doi: 10.3389/fpsyg.2017.01560

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders 5th Edn.* Washington, DC: American Psychiatric Association. doi: 10.1176/appi.books.9780890425596

Armour, C., Contractor, A., Shea, T., Elhai, J. D., and Pietrzak, R. H. (2016). Factor structure of the PTSD checklist for DSM-5: relationships among symptom clusters, anger, and impulsivity. *J. Nerv. Ment. Dis.* 204, 108–115. doi: 10.1097/NMD.0000000000000430

Baumeister, H., Reichler, L., Munzinger, M., and Lin, J. (2014). The impact of guidance on internet-based mental health interventions—a systematic review. *Internet Intervent.* 1, 205–215. doi: 10.1016/j.invent.2014.08.003

Berger, T., Krieger, T., Sude, K., Meyer, B., and Maercker, A. (2018). Evaluating an e-mental health program ("deprexis") as adjunctive treatment tool in psychotherapy for depression: results of a pragmatic randomized controlled trial. *J. Affect. Disord.* 227, 455–462. doi: 10.1016/j.jad.2017.11.021

Bergman, H. E., Kline, A. C., Feeny, N. C., and Zoellner, L. A. (2015). Examining PTSD treatment choice among individuals with subthreshold PTSD. *Behav. Res. Ther.* 73, 33–41. doi: 10.1016/j.brat.2015.07.010

Bessière, K., Seay, A. F., and Kiesler, S. (2007). The ideal elf: Identity exploration in World of Warcraft. *Cyberpsychol. Behav.* 10, 530–535. doi: 10.1089/cpb.2007.9994

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. Hoboken, NJ: John Wiley & Sons.

Brancu, M., Mann-Wrobel, M., Beckham, J. C., Wagner, H. R., Elliott, A., Robbins, A. T., et al. (2016). Subthreshold posttraumatic stress disorder: a meta-analytic review of DSM–IV prevalence and a proposed DSM—5 approach to measurement. *Psychol. Trauma* 8, 222–232. doi: 10.1037/tra0000078

Breslau, N., Lucia, V. C., and Davis, G. C. (2004). Partial PTSD versus full PTSD: an empirical examination of associated impairment. *Psychol. Med.* 34, 1205–1214. doi: 10.1017/S0033291704002594

Buckley, T. C., Mozley, S. L., Bedard, M. A., Dewulf, A. C., and Greif, J. (2004). Preventive health behaviors, health-risk behaviors, physical morbidity, and health-related role functioning impairment in veterans with post-traumatic stress disorder. *Mil. Med.* 169, 536–540. doi: 10.7205/MILMED.169.7.536

Cárdenas, G., and De La Rosa, A. (2012). Using virtual reality for PTSD treatment in criminal violence victims. *J. Cyber Ther Rehab.* 5, 65–67. Available online at: https://link.gale.com/apps/doc/A321462391/AONE?u=43wien&sid=AONE&xid=11ac7dbe

*Cernvall, M., Sveen, J., Bergh Johannesson, K., and Arnberg, F. (2018). A pilot study of user satisfaction and perceived helpfulness of the Swedish version of the mobile app PTSD Coach. *Eur. J. Psychotraumatol.* 9 (Suppl. 1):1472990. doi: 10.1080/20008198.2018.1472990

Cohen, B. E., Marmar, C. R., Neylan, T. C., Schiller, N. B., Ali, S., and Whooley, M. A. (2009). Posttraumatic stress disorder and health-related quality of life in patients with coronary heart disease. *Arch. Gen. Psychiatr.* 66, 1214–1220. doi: 10.1001/archgenpsychiatry.2009.149

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Cukor, J., Wyka, K., Jayasinghe, N., and Difede, J. (2010). The nature and course of subthreshold PTSD. *J. Anxiety Disord.* 24, 918–923. doi: 10.1016/j.janxdis.2010.06.017

Cusack, K., Jonas, D. E., Forneris, C. A., Wines, C., Sonis, J., Middleton, J. C., et al. (2016). Psychological treatments for adults with posttraumatic stress disorder: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 43, 128–141. doi: 10.1016/j.cpr.2015.10.003

Darves-Bornoz, J.-M., Alonso, J., de Girolamo, G., de Graaf, R., Haro, J.-M., Kovess-Masfety, V., et al. (2008). Main traumatic events in Europe: PTSD in the European study of the epidemiology of mental disorders survey. *J. Trauma. Stress* 21, 455–462. doi: 10.1002/jts.20357

Dickstein, B. D., Walter, K. H., Schumm, J. A., and Chard, K. M. (2013). Comparing response to cognitive processing therapy in military veterans with subthreshold and threshold posttraumatic stress disorder. *J. Trauma. Stress* 26, 703–709. doi: 10.1002/jts.21869

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M. R., and Christensen, H. (2013). Smartphones for smarter delivery of mental health programs: a systematic review. *J. Med. Internet Res.* 15:e247. doi: 10.2196/jmir.2791

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., and Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychol. Methods* 1, 170–177. doi: 10.1037/1082-989X.1.2.170

Duval, S., and Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.* 95, 89–98. doi: 10.1080/01621459.2000.10473905

Elhai, J. D., and Palmieri, P. A. (2011). The factor structure of posttraumatic stress disorder: a literature update, critique of methodology, and agenda for future research. *J. Anxiety Disord.* 25, 849–854. doi: 10.1016/j.janxdis.2011.04.007

Etkin, A., and Wager, T. D. (2007). Functional neuroimaging of anxiety: a meta-analysis of emotional processing in PTSD, social anxiety disorder, and specific phobia. *Am. J. Psychiatr.* 164, 1476–1488. doi: 10.1176/appi.ajp.2007.07030504

Eurostat (2016). *Internet Use by Individuals*. Retrieved from: https://ec.europa.eu/eurostat/documents/2995521/7771139/9-20122016-BP-EN.pdf/f023d81a-dce2-4959-93e3-8cc7082b6edd. (accessed March 5, 2019).

Firth, J., and Torous, J. (2015). Smartphone apps for schizophrenia: a systematic review. *JMIR mHealth uHealth* 3:e102. doi: 10.2196/mhealth.4930

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., et al. (2017). The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatr.* 16, 287–298. doi: 10.1002/wps.20472

Foa, E. B., Keane, T. M., Friedman, M. J., and Cohen, J. A. (2009). *Effective Treatments for PTSD: Practice Guidelines from the International Society for Traumatic Stress Studies, 2nd Edn*. New York, NY: Guilford Press.

Ginzburg, K., Ein-Dor, T., and Solomon, Z. (2010). Comorbidity of posttraumatic stress disorder, anxiety and depression: a 20-year longitudinal study of war veterans. *J. Affect. Disord.* 123, 249–257. doi: 10.1016/j.jad.2009.08.006

Goldbeck, L., Muche, R., Sachser, C., Tutus, D., and Rosner, R. (2016). Effectiveness of trauma-focused cognitive behavioral therapy for children and adolescents: a randomized controlled trial in eight German mental health clinics. *Psychother. Psychosom.* 85, 159–170. doi: 10.1159/000442824

Gustafson, D. H., McTavish, F. M., Chih, M. Y., Atwood, A. K., Johnson, R. A., Boyle, M. G., et al. (2014). A smartphone application to support recovery from alcoholism: a randomized clinical trial. *JAMA Psychiatr.* 71, 566–572. doi: 10.1001/jamapsychiatry.2013.4642

Heaney, C. A., and Israel, B. A. (2008). "Social networks and social support," in: *Health Behavior and Health Education: Theory, Research, and Practice* eds K. Glanz, B. K. Rimer, and K. Viswanath (San Francisco, CA: Jossey-Bass), 189–2104.

Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *J. Educ. Stat.* 6, 107–128. doi: 10.3102/10769986006002107

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., and Torous, J. (2019). Digital mental health apps and the therapeutic alliance: initial review. *BJPsych Open* 5:e15. doi: 10.1192/bjo.2018.86

Higgins, J. P. T., and Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011].* Chichester: The Cochrane Collaboration. Retrieved from: http://handbook.cochrane.org.

Hoge, C. W., Auchterlonie, J. L., and Milliken, C. S. (2006). Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. *JAMA* 295, 1023–1032. doi: 10.1001/jama.295.9.1023

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., and Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *N. Engl. J. Med.* 351, 13–22. doi: 10.1056/NEJMoa040603

Juarascio, A. S., Manasse, S. M., Goldstein, S. P., Forman, E. M., and Butryn, M. L. (2014). Review of smartphone applications for the treatment of eating disorders. *Eur. Eat. Disord. Rev.* 23, 1–11. doi: 10.1002/erv.2327

Kazdin, A. E., and Rabbitt, S. M. (2013). Novel models for delivering mental health services and reducing the burdens of mental illness. *Clin. Psychol. Sci.* 1, 170–191. doi: 10.1177/2167702612463566

Kessler, R. C. (2000). Posttraumatic stress disorder: the burden to the individual and to society. *J. Clin. Psychiatr.* 61, 4–12. doi: 10.4088/JCP.v61n0713e

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., and Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Arch. Gen. Psychiatr.* 62, 593–602. doi: 10.1001/archpsyc.62.6.593

Koenen, K. C., Ratanatharathorn, A., Ng, L., McLaughlin, K. A., Bromet, E. J., Stein, D. J., et al. (2017). Posttraumatic stress disorder in the world mental health surveys. *Psychol. Med.* 47, 2260–2274. doi: 10.1017/S0033291717000708

Kok, B. C., Herrell, R. K., Thomas, J. L., and Hoge, C. W. (2012). Posttraumatic stress disorder associated with combat service in Iraq or Afghanistan: reconciling prevalence differences between studies. *J. Nerv. Ment. Dis.* 200, 444–450. doi: 10.1097/NMD.0b013e3182532312

Korte, K. J., Allan, N. P., Gros, D. F., and Acierno, R. (2016). Differential treatment response trajectories in individuals with subclinical and clinical PTSD. *J. Anxiety Disord.* 38, 95–101. doi: 10.1016/j.janxdis.2016.01.006

Kothgassner, O. D., Goreis, A., Kafka, J. X., Kaufmann, M., Atteneder, K., Beutl, L., et al. (2019a). Virtual social support buffers stress response: an experimental comparison of real-life and virtual support prior to a social stressor. *J. Behav. Ther. Exp. Psychiatr.* 63, 57–65. doi: 10.1016/j.jbtep.2018.11.003

Kothgassner, O. D., Goreis, A., Kafka, J. X., Van Eickels, R. L., Plener, P. L., and Felnhofer, A. (2019b). Virtual reality exposure therapy for post-traumatic stress disorder (PTSD): a meta-analysis. *Eur. J. Psychotraumatol.* 10:1654782. doi: 10.1080/20008198.2019.1654782

Kothgassner, O. D., Plener, P. L., Sachser, C., Witt, A., Brähler, E., and Fegert, J. M. (2020). Adverse childhood experiences and avatar preferences in online games. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie.* 48, 15–23. doi: 10.1024/1422-4917/a000647

Kroenke, K., Spitzer, R. L., and Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* 16, 606–613. doi: 10.1046/j.1525-1497.2001.016009606.x

Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., and Mokdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114, 163–173. doi: 10.1016/j.jad.2008.06.026

Kuhn, E., Greene, C., Hoffman, J., Nguyen, T., Wald, L., Schmidt, J., et al. (2014). Preliminary evaluation of PTSD coach, a smartphone app for post-traumatic stress symptoms. *Mil. Med.* 179, 12–18. doi: 10.7205/MILMED-D-13-00271

*Kuhn, E., Kanuri, N., Hoffman, J. E., Garvert, D. W., Ruzek, J. I., and Taylor, C. B. (2017). A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J. Consult. Clin. Psychol.* 85, 267–273. doi: 10.1037/ccp0000163

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863

López, S. R., Barrio, C., Kopelowicz, A., and Vega, W. A. (2012). From documenting to eliminating disparities in mental health care for Latinos. *Am. Psychol.* 67, 511–523. doi: 10.1037/a0029737

Luxton, D. D., Skopp, N. A., and Maguen, S. (2010). Gender differences in depression and PTSD symptoms following combat exposure. *Depress. Anxiety* 27, 1027–1033. doi: 10.1002/da.20730

Maercker, A., Forstmeier, S., Wagner, B., Glaesmer, H., and Brähler, E. (2008). Posttraumatische belastungsstörungen in deutschland. Ergebnisse einer gesamtdeutschen epidemiologischen Untersuchung. *Nervenarzt* 79, 577–586. doi: 10.1007/s00115-008-2467-5

Mandelli, L., Petrelli, C., and Serretti, A. (2015). The role of specific early trauma in adult depression: a meta-analysis of published literature. childhood trauma and adult depression. *Eur. Psychiatr.* 30, 665–680. doi: 10.1016/j.eurpsy.2015.04.007

Marmar, C. R., Schlenger, W., Henn-Haase, C., Qian, M., Purchia, E., Li, M., et al. (2015). Course of posttraumatic stress disorder 40 years after the vietnam war: findings from the national vietnam veterans longitudinal study. *JAMA Psychiatr.* 72, 875–881. doi: 10.1001/jamapsychiatry.2015.0803

Marshall, R. D., Olfson, M., Hellman, F., Blanco, C., Guardino, M., and Struening, E. L. (2001). Comorbidity, impairment, and suicidality in subthreshold PTSD. *Am. J. Psychiatr.* 158, 1467–1473. doi: 10.1176/appi.ajp.158.9.1467

*Miner, A., Kuhn, E., Hoffman, J. E., Owen, J. E., Ruzek, J. I., and Taylor, C. B. (2016). Feasibility, acceptability, and potential efficacy of the PTSD coach app: a pilot randomized controlled trial with community trauma survivors. *Psychol. Trauma.* 8, 384–392. doi: 10.1037/tra0000092

Myers, C. E., Radell, M. L., Shind, C., Ebanks-Williams, Y., Beck, K. D., and Gilbertson, M. W. (2016). Beyond symptom self-report: use of a computer "avatar" to assess post-traumatic stress disorder (PTSD) symptoms. *Stress* 19, 593–598. doi: 10.1080/10253890.2016.1232385

Mylle, J., and Maes, M. (2004). Partial posttraumatic stress disorder revisited. *J. Affect. Disord.* 78, 37–48. doi: 10.1016/S0165-0327(02)00218-5

Norris, F. H., Kaniasty, K., and Thompson, M. P. (1997). "The psychological consequences of crime: findings from a longitudinal population-based study," in *Victims of Crime, 2nd Edn.* eds R. C. Davis, A. J. Lurigio, and W. G. Skogan (Thousand Oaks, CA: Sage Publications), 146–166.

Owen, J. E., Jaworski, B. K., Kuhn, E., Makin-Byrd, K. N., Ramsey, K. M., and Hoffman, J. E. (2015). mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. *JMIR Mental Health* 2:e7. doi: 10.2196/mental.3935

Owens, D. K., Lohr, K. N., Atkins, D., Treadwell, J. R., Reston, J. T., Bass, E. B., et al. (2010). AHRQ series paper 5: grading the strength of a body of evidence when comparing medical interventions—agency for healthcare research and quality and the effective health-care program. *J. Clin. Epidemiol.* 63, 513–523. doi: 10.1016/j.jclinepi.2009.03.009

Pew Research Center (2018). *Mobile Fact Sheet*. Retrieved from: http://www.pewinternet.org/fact-sheet/mobile/. (accessed March 5, 2019).

Pinto, M. D., Greenblatt, A. M., Hickman, R. L., Rice, H. M., Thomas, T. L., and Clochesy, J. M. (2016). Assessing the critical parameters of eSMART-MH: a promising avatar-based digital therapeutic intervention to reduce depressive symptoms. *Perspect. Psychiatr. Care* 52, 157–168. doi: 10.1111/ppc.12112

*Possemato, K., Kuhn, E., Johnson, E., Hoffman, J. E., Owen, J. E., Kanuri, N., et al. (2016). Using PTSD coach in primary care with and without clinician support: a pilot randomized controlled trial. *Gen. Hosp. Psychiatr.* 38, 94–98. doi: 10.1016/j.genhosppsych.2015.09.005

Powers, M. B., Halpern, J. M., Ferenschak, M. P., Gillihan, S. J., and Foa, E. B. (2010). A meta-analytic review of prolonged exposure for posttraumatic stress disorder. *Clin. Psychol. Rev.* 30, 635–641. doi: 10.1016/j.cpr.2010.04.007

Prins, A., Ouimette, P., Kimerling, R., Camerond, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., et al. (2004). The primary care PTSD screen (PC–PTSD): development and operating characteristics. *Primary Care Psychiatr.* 9, 9–14. doi: 10.1185/135525703125002360

Reger, G. M., Hoffman, J., Riggs, D., Rothbaum, B. O., Ruzek, J., Holloway, K. M., et al. (2013). The "PE coach" smartphone application: an innovative approach to improving implementation, fidelity, and homework adherence during prolonged exposure. *Psychol. Serv.* 10, 342–349. doi: 10.1037/a0032774

Reger, G. M., Skopp, N. A., Edwards-Stewart, A., and Lemus, E. L. (2015). Comparison of prolonged exposure (PE) coach to treatment as usual: a case series with two active duty soldiers. *Mil. Psychol.* 27, 287–296. doi: 10.1037/mil0000083

Rehm, I. C., Foenander, E., Wallace, K., Abbott, J. A. M., Kyrios, M., and Thomas, N. (2016). What role can avatars play in e-mental health interventions? Exploring new models of client–therapist interaction. *Front. Psychiatr.* 7:186. doi: 10.3389/fpsyt.2016.00186

Rodriguez-Paras, C., Tippey, K., Brown, E., Sasangohar, F., Creech, S., Kum, H. C., et al. (2017). Posttraumatic stress disorder and mobile health: app investigation and scoping literature review. *JMIR mHealth uHealth* 5:e156. doi: 10.2196/mhealth.7318

*Roy, M. J., Costanzo, M. E., Highland, K. B., Olsen, C., Clayborne, D., and Law, W. (2017). An app a day keeps the doctor away: guided education and training via smartphones in subthreshold post traumatic stress disorder. *Cyberpsychol. Behav. Soc. Netw.* 20, 470–478. doi: 10.1089/cyber.2017.0221

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., et al. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *Int. J. Behav. Nut. Phys. Activ.* 13:127. doi: 10.1186/s12966-016-0454-y

Shalev, A. Y., Ankri, Y. L. E., Peleg, T., Israeli-Shalev, Y., and Freedman, S. (2011). Barriers to receiving early care for PTSD: results from the Jerusalem trauma outreach and prevention study. *Psychiatric Serv.* 62, 765–773. doi: 10.1176/ps.62.7.pss6207_0765

Shiner, B., Bateman, D., Young-Xu, Y., Zayed, M., Harmon, A. L., Pomerantz, A., et al. (2012). Comparing the stability of diagnosis in full vs. partial posttraumatic stress disorder. *J. Nerv. Mental Dis.* 200, 520–525. doi: 10.1097/NMD.0b013e318257c6da

Smith, P., Yule, W., Perrin, S., Tranah, T., Dalgleish, T., and Clark, D. M. (2007). Cognitive-behavioral therapy for PTSD in children and adolescents: a preliminary randomized controlled trial. *J. Am. Acad. Child Adolesc. Psychiatr.* 46, 1051–1061. doi: 10.1097/CHI.0b013e318067e288

Stein, M. B., Walker, J. R., Hazen, A. L., and Forde, D. R. (1997). Full and partial posttraumatic stress disorder: findings from a community survey. *Am. J. Psychiatr.* 154, 1114–1119. doi: 10.1176/ajp.154.8.1114

Sterne, J. A. C., and Egger, M. (2005). "Regression methods to detect publication and other bias in meta-analysis," in *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments,* eds H. R. Rothstein, A. J. Sutton, and M. Borenstein (New York, NY: Wiley), 99–110.

*Tiet, Q. Q., Duong, H., Davis, L., French, R., Smith, C. L., Leyva, Y. E., et al. (2019). PTSD coach mobile application with brief telephone support: a pilot study. *Psychol. Serv.* 16, 227–232. doi: 10.1037/ser0000245

U. S. Department of Veterans Affairs (2011a). *PTSD Coach [Mobile Application Software, Android Version]*. Retrieved from: https://play.google.com/store/apps/details?id=is.vertical.ptsdcoach&hl=en. (accessed March 5, 2019).

U. S. Department of Veterans Affairs (2011b). *PTSD Coach* (*Mobile Application Software, iOS Version) [Mobile Application Software]*. Retrieved from: https://itunes.apple.com/us/app/ptsd-coach/id430646302?mt=8. (accessed March 5, 2019).

U. S. Department of Veterans Affairs (2017a). *PE Coach [Mobile Application Software, Android Version]*. Retrieved from: https://play.google.com/store/apps/details?id=gov.va.mobilehealth.ncptsd.pecoach&hl=en_U.S. (accessed March 5, 2019).

U. S. Department of Veterans Affairs (2017b). *PE Coach [Mobile Application Software, iOS Version]*. Retrieved from: https://itunes.apple.com/us/app/pe-coach-2/id1281266434?mt=8. (accessed March 5, 2019).

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 36, 1–48. doi: 10.18637/jss.v036.i03

Viswanathan, M., Patnode, C. D., Berkman, N. D., Bass, E. B., Chang, S., Hartling, L., et al. (2018). Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *J. Clin. Epidemiol.* 97, 26–34. doi: 10.1016/j.jclinepi.2017.12.004

Weathers, F. W., Bovin, M. J., Lee, D. J., Sloan, D. M., Schnurr, P. P., Kaloupek, D. G., et al. (2017). The clinician-administered PTSD scale for DSM-5 (CAPS-5): development and initial psychometric evaluation in military veterans. *Psychol. Assess.* 30, 383–395. doi: 10.1037/pas0000486

Weathers, F. W., and Ford, J. (1996). "Psychometric properties of the PTSD Checklist (PCL-C, PCL-S, PCL-M, PCL-PR)," in *Measurement of Stress, Trauma, and Adaptation*. ed B. H. Stamm. Lutherville, MD: Sidran, 250–251.

Weathers, F. W., Keane, T. M., and Davidson, J. R. T. (2001). Clinician-administered PTSD scale: a review of the first ten years of research. *Depress. Anxiety* 13, 132–156. doi: 10.1002/da.1029

Weathers, F. W., Litz, B. T., Herman, D., Huska, J., and Keane, T. (1994). *The PTSD Checklist—Civilian Version (PCL–C)*. Boston, MA: National Center for PTSD. doi: 10.1037/t02622-000

Whittaker, R., McRobbie, H., Bullen, C., Borland, R., Rodgers, A., and Gu, Y. (2016). Mobile phone-based interventions for smoking cessation. *Cochrane Database Syst. Rev.* 4:CD006611. doi: 10.1002/14651858.CD006611.pub4

Wickersham, A., Petrides, P. M., Williamson, V., and Leightley, D. (2019). Efficacy of mobile application interventions for the treatment of post-traumatic stress disorder: a systematic review. *Digital Health* 5:2055207619842986. doi: 10.1177/2055207619842986

World Health Organization (1993). *ICD-10, the ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. Geneva: World Health Organization.

*Studies included in the meta-analysis are marked with an asterisk.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SS declared a shared affiliation with no collaborations with one of the Authors AG to the handling editor at the time of review.

frontiers in Neuroscience

HYPOTHESIS AND THEORY
published: 28 February 2020
doi: 10.3389/fnins.2020.00164

Combining Mobile Crowdsensing and Ecological Momentary Assessments in the Healthcare Domain

Robin Kraft[1,2], Winfried Schlee[3], Michael Stach[1], Manfred Reichert[1], Berthold Langguth[3], Harald Baumeister[2], Thomas Probst[4], Ronny Hannemann[5] and Rüdiger Pryss[6]*

[1] Institute of Databases and Information Systems, Ulm University, Ulm, Germany, [2] Department of Clinical Psychology and Psychotherapy, Ulm University, Ulm, Germany, [3] Clinic and Policlinic for Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany, [4] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems an der Donau, Austria, [5] Sivantos GmbH, Erlangen, Germany, [6] Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

The increasing prevalence of smart mobile devices (e.g., smartphones) enables the combined use of mobile crowdsensing (MCS) and ecological momentary assessments (EMA) in the healthcare domain. By correlating qualitative longitudinal and ecologically valid EMA assessment data sets with sensor measurements in mobile apps, new valuable insights about patients (e.g., humans who suffer from chronic diseases) can be gained. However, there are numerous conceptual, architectural and technical, as well as legal challenges when implementing a respective software solution. Therefore, the work at hand (1) identifies these challenges, (2) derives respective recommendations, and (3) proposes a reference architecture for a MCS-EMA-platform addressing the defined recommendations. The required insights to propose the reference architecture were gained in several large-scale mHealth crowdsensing studies running for many years and different healthcare questions. To mention only two examples, we are running crowdsensing studies on questions for the tinnitus chronic disorder or psychological stress. We consider the proposed reference architecture and the identified challenges and recommendations as a contribution in two respects. First, they enable other researchers to align our practical studies with a baseline setting that can satisfy the variously revealed insights. Second, they are a proper basis to better compare data that was gathered using MCS and EMA. In addition, the combined use of MCS and EMA increasingly requires suitable architectures and associated digital solutions for the healthcare domain.

Keywords: mobile crowdsensing (MCS), crowdsourcing, ecological momentary assessments (EMA), mobile healthcare application, chronic disorders, reference architecture

OPEN ACCESS

Edited by:
Alejandro Rodríguez González, Polytechnic University of Madrid, Spain

Reviewed by:
Kazuhiro Yoshiuchi, The University of Tokyo, Japan
Peter A. Tass, Stanford University School of Medicine, United States

****Correspondence:***
Robin Kraft
robin.kraft@uni-ulm.de

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *05 September 2019*
Accepted: *13 February 2020*
Published: *28 February 2020*

Citation:
Kraft R, Schlee W, Stach M, Reichert M, Langguth B, Baumeister H, Probst T, Hannemann R and Pryss R (2020) Combining Mobile Crowdsensing and Ecological Momentary Assessments in the Healthcare Domain. Front. Neurosci. 14:164. doi: 10.3389/fnins.2020.00164

1. INTRODUCTION

For many use cases in the healthcare domain, e.g., in the assessment of chronic diseases and disorders, there is a need for the collection of large, qualitative, longitudinal, and ecologically valid data sets. Additionally, contextual information like environmental factors can give even more valuable insights to researchers, healthcare providers (e.g., physicians or therapists), and last but not least, the patients themselves. At the same time, smart mobile devices (e.g., smartphones and smartwatches) and low-powered sensors are becoming increasingly ubiquitous. Two concepts that

highly benefit from these advancements are *mobile crowdsensing* (MCS) and *ecological momentary assessments* (EMA). They can be used in combination in the form of mobile apps to correlate EMA assessment data with sensor measurement data in order to gain even more valuable insights about patients. However, there are numerous challenges when implementing a software solution in order to provide the desired functionality, to cope with technical aspects, as well as to comply with high standards and regulations in the healthcare domain. In this work, we discuss these challenges, derive several recommendations and propose a reference architecture for a respective software platform. These insights were mainly gained through several studies that combined MCS and EMA based on mHealth apps that we have developed in the last years. The mentioned studies, in turn, address different healthcare questions and are mostly running for many years. This provides us with a proper basis for the proposed reference architecture as well as the introduced set of recommendations. To conclude, the work at hand provides the following contributions:

- Various challenges are pointed out and discussed on the basis of the ongoing research project TrackYourTinnitus (TYT), which has been running since 2014.
- A number of recommendations are derived from the findings during this and other related projects.
- A reference architecture for a platform enabling the combination of MCS and EMA is proposed that aims to address the defined recommendations. Additionally, technical considerations for the implementation of the architecture are discussed.

The remainder of this paper is organized as follows. In section 2, related work in the fields of mobile crowdsensing and ecological momentary assessments is presented, and the combination of both concepts is discussed. Lessons learned during the operation of the TrackYourTinnitus (TYT) project are presented in section 3. In section 4, we derive recommendations for a MCS-EMA platform, propose a reference architecture to address these recommendations, and discuss selected technical considerations. Furthermore, the findings and their implications for MCS and EMA research are discussed in section 5. Finally, section 6 concludes the paper with a summary and an outlook.

2. MOBILE CROWDSENSING IN HEALTHCARE

In this section, we discuss mobile crowdsensing (MCS) in the healthcare domain. We cover related work in the fields of MCS and EMA and explain how we relate ecological momentary assessments (EMA) apps to MCS.

2.1. Mobile Crowdsensing (MCS)

Mobile crowdsensing is a paradigm in which a community is leveraging devices with sensing and computing capabilities to collectively share data and extract information in order to measure and map phenomena of common interest. Therefore, it is also referred to as *community sensing*. As opposed to *personal sensing*, where the phenomena that are monitored belong to an individual user, community sensing applications focus on monitoring large-scale phenomena that cannot easily be measured by a single user or device (Ganti et al., 2011). This set of applications can then further be classified into *participatory sensing* (Burke et al., 2006) and *opportunistic sensing* (Lane et al., 2010) applications. Participatory sensing requires an active and conscious involvement of the user in order to contribute sensor data, while in opportunistic sensing, user involvement is minimal and sensor measurements as well as data transmission are done passively. In reality, mobile crowdsensing applications will often be located somewhere between these two extremes and use both paradigms to some extent. Furthermore, there exist recent works that reflect the categories of participatory and opportunistic sensing in the healthcare context (e.g., Pryss, 2019).

Furthermore, we consider the concept of mobile crowdsensing in the healthcare domain. Therefore, we are focusing on correlating personal sensing data with assessment data in order to gain insights on specific health conditions, (chronic) diseases and the patients' behavior. We consider the potential knowledge generated from this data as the phenomenon of common interest in terms of mobile crowdsensing. There are a number of applications in the field of healthcare (Guo et al., 2015). Its use cases include data collection in clinical and health/psychological trials (Pryss et al., 2015; Schobel et al., 2015), environmental monitoring and pollution measurement like noise pollution (Schweizer et al., 2011; Zappatore et al., 2017) or air pollution (Mun et al., 2009), public health (Wesolowski et al., 2012), and personal well-being (Consolvo et al., 2006). Although various mobile applications and solutions have been proposed, less works exist that cover reference settings to build generic solutions (Tokosi and Scholtz, 2019). In addition, few works are based on comprehensive experiences that are gained through various long-running projects (Tokosi and Scholtz, 2019).

2.2. Ecological Momentary Assessments (EMA)

Ecological Momentary Assessment (EMA) (Stone and Shiffman, 1994) denotes a range of research methods aiming to assess phenomena with ecological validity by allowing subjects and patients to repeatedly report in real time, in real-world settings, over time, and across contexts and therefore avoiding the bias of retrospective reports (Pryss et al., 2018a). Among numerous other aspects, EMA is characterized by several key features (Shiffman et al., 2008):

- Ecological: Data is collected *in situ*, i.e., in real-world settings and environments, which constitutes the ecological validity.
- Momentary: Assessments focus on current or very recent states in real time, which aims to avoid a bias associated with retrospective assessments.
- Strategic sampling: Assessment timings are strategically selected by specific sampling schemes, e.g., based on particular events of interest or by random, representative samplings across contexts.

- Longitudinal data: Subjects complete multiple assessments over time, which provides longitudinal data with insights on how the state varies over time and across situations.

A related methodology in the field of momentary research is the *Experience Sampling Method* (ESM) (Larson and Csikszentmihalyi, 2014; Van Haren, 2018), which aims at measuring momentary behavior, thoughts, symptoms, and feelings of participants, collected through self-reports that are typically filled out several times a day over several consecutive days (Myin-Germeys et al., 2009; Van Berkel et al., 2018). Generally, ESM has a focus on random time sampling and private, subjective experiences, while EMA is defined more broadly, as it also includes other sampling approaches and behavioral as well as physiological measures (Stone and Shiffman, 2002). Since we are striving to make our architecture as generic as possible and to additionally address physiological sampling via mobile sensors, we focus on EMA within the scope of this work.

2.2.1. Implementation of EMA With Mobile Devices

EMA studies can be carried out with the help of portable electronic devices, which support the following EMA key functions (Shiffman, 2007; Shiffman et al., 2008):

1. Present assessment content to the subject (i.e., display questions and response options).
2. Manage assessment logic (e.g., handle branching and validate inputs).
3. Provide time-stamp data to document when assessments are completed.
4. Store assessment data.
5. Manage prompting schedules (e.g., determine when assessments should be made).
6. Prompt the subject to complete assessments.

Modern smartphones offer all of these functions, as they provide high-resolution displays, advanced processing power and storage, as well as push notifications (Raento et al., 2009). They have already been used in different EMA studies (Ebner-Priemer and Kubiak, 2007; Schlee et al., 2016). We summarize smartphone applications that offer EMA functionality using the term *EMA apps*. Smartphones offer additional capabilities that go beyond the initially defined EMA key functions, most importantly advanced processing capabilities, an (almost) always available network connection and built-in as well locally connected sensors (Van Berkel et al., 2018). Furthermore, data can be stored locally on the device and synchronized with the server, enabling an offline availability. Therefore, we explore different extensions of EMA apps and their combinations and study their effects. These extensions can be broadly categorized in (1) *guidance*, (2) *feedback*, (3) *adjustable prompts*, and (4) *dynamic questionnaires*. Generally, we distinguish between EMA apps and features that are used for data collection only (mainly research) and others that offer a benefit to the user (research and health care). The four categories of extensions we consider are described in the following:

- Guidance: We refer to *guidance* as the option for the user of the EMA app to link to a contact person. This contact person might be some kind of healthcare provider (HCP) that has some professional qualifications, e.g., a physician or therapist. The HCP might influence the process of EMA prompts, provide feedback to submitted data, and offer general advice to the user, or just act as an observer.
- Feedback: The EMA app could offer feedback to the user when he/she submits questionnaires. This feedback can be in the form of text messages by the HCP or automated feedback by the app, like tips and warnings when certain thresholds are exceeded, as well as graphical feedback in the form of graphs about the history of different measurements. We assume that feedback of this kind might act as an incentive to users and therefore increase adherence, but we also want to study the effects of this feedback on the EMA data.
- Adjustable prompts: Assessment prompts (i.e., notifications) can either be fixed and determined by the system, defined by the HCP, event-triggered (e.g., when a patient perceives his tinnitus, or when a context change is detected through sensor data), or can be adjusted by the user in a flexible manner.
- Dynamic questionnaires: The content of EMA questionnaires could be dynamic and adjusted depending on answered questionnaires in the past, occurring events, or other external parameters (e.g., the current weather retrieved through a web service).

2.2.2. Potential Challenges

There are a number of potential challenges when employing EMA studies, which are outlined in the following (Van Berkel et al., 2018):

- Participant burden: Answering questionnaires multiple times a day can be burdensome for participants. To counteract this issue, the number of questions, alerts, and question types should be kept as small as possible.
- Participant retention: Related to the frequent answering of questionnaires, study dropout rates are generally high. There has to be some sort of incentive for participants in order to keep them entering their data in a constant manner.
- Programming: There is no generic software solution that allows to employ EMA studies on mobile devices without requiring at least basic programming skills.
- Platform heterogeneity: Flexible software is required in order to support a large number of different hardware devices and operating systems.
- Data quality: Since data is not collected in a controlled environment, participants' data might be of low quality or noisy. Mechanisms should be in place to avoid or compensate missing, wrong or careless answers, as well as response shifts (i.e., changes in the participant's internal standards) or changes in the participants reactivity (i.e., behavioral adjustments because the participants know that they are being observed). In the context of participant retention, participants might answer the questionnaires as often as possible, even in a dishonest way, if they expect a reward or think they are supporting the platform in this way.

2.3. Combining Mobile Crowdsensing and Ecological Momentary Assessments

Crucially, smartphones enable us to not only collect explicit answers to EMA questionnaires, but additionally capture the *context* in which they are collected (Van Berkel et al., 2018). We consider EMA apps similar to mobile crowdsensing, in which the assessed phenomenon in terms of mobile crowdsensing is the ecological data collected in EMA questionnaires. Consequently, we combine the concepts and features of EMA apps with the paradigms of mobile crowdsensing by correlating questionnaire responses and sensor data in order to gain new insights on certain phenomena. Furthermore, we derive different classes of mobile crowdsensing EMA apps depending on the EMA features they provide and the crowdsensing paradigms that they make use of. **Table 1** shows examples for apps that are incorporating both EMA and mobile crowdsensing features that were developed by the authors. The *TrackYourTinnitus (TYT)* project tracks one's individual tinnitus and is described in detail in section 3. Similar to TYT, *TrackYourHearing* (TYH), *TrackYourDiabetes (TYD)*, and *TrackYourStress (TYS)* (Pryss et al., 2019) help the user to assess and track the progress of their hearing loss, diabetes, or stress level, respectively, and allow them to be more sensitive to symptom changes in specific contexts. The *TinnitusTipps* app was designed to enable the communication between healthcare providers (HCP) and tinnitus patients, including the assessment of the user's tinnitus and various automatic as well as manual feedback options. The *KINDEX mum screen* enables the assessment of psychosocial stress factors during pregnancy (Ruf-Leuschner et al., 2016). Finally, the *Intersession* app focuses on the assessment and guidance of users during the time between therapy sessions. Even though the last two apps are not incorporating any sensor measurements and are therefore by definition not utilizing MCS, we consider their assessed ecological data as phenomenon of common interest in terms of mobile crowdsensing and their contributions regarding guidance and feedback as a valuable basis for MCS-EMA platforms. The number of users, submitted answer sheets and released versions as well as the incorporated sensor measurements for the developed apps are shown in **Table 2**. The sensor measurements are performed while the patients answer the questionnaires and stored together with the answer data in order to allow to investigate correlations. To put these apps into perspective, **Figure 1** shows how they are incorporating guidance and feedback (as defined in section 2.2.1) on a relative two-dimensional scale based on a subjective rating (however, guided by the extensive experiences) by the authors.

3. LESSONS LEARNED FROM THE TRACKYOURTINNITUS PROJECT

The TrackYourTinnitus (TYT) platform is available and has been maintained since April 2014. It consists of a website for registration[1], two native mobile applications (iOS and Android), and a central backend that stores the collected data in a relational database. The mobile apps track the individual tinnitus perception by asking the patients to complete tinnitus assessment EMA questionnaires at different times during the day and on a random basis. The daily questionnaire is assessing tinnitus by measuring eight dimensions, e.g., tinnitus loudness and distress, utilizing the questions shown in **Table 3**. Furthermore, the apps measure the environmental sound level while patients fill out the questionnaires (Pryss et al., 2015). Medically, tinnitus is the perception of a sound when no corresponding external sound is present. The symptoms, in turn, are subjective and vary over time. Hence, TYT was realized to monitor and evaluate the variability of symptoms over time based on EMA and mobile crowdsensing (Schlee et al., 2016).

One potential risk worth considering is whether continuous tracking of tinnitus with the app could aggravate the patient's symptoms by drawing additional attention to them. However, it has been shown that the regular use of the TYT app has no significant negative effect on the perceived tinnitus loudness and the tinnitus distress. Therefore, the app can be considered as a safe method for the longitudinal assessment of tinnitus symptoms in the everyday life of patients (Schlee et al., 2016). Another health risk is that patients (or their HCP) use TYT as a treatment tool and unnecessarily change their treatment plan due to self-reported symptoms in the app. In order to make patients aware of these risks, they are outlined on the TYT website[2].

Figures 2, **3** show the general process a user is going through when using the TYT iOS or Android application. Note that these figures are process-oriented graphs in terms of the Business Process Modeling Notation (BPMN). This notation is an industry standard and also well-known for the documentation of healthcare-related procedures (Reichert and Pryss, 2017). With respect to these figures, first of all, a user authenticates himself/herself with his/her login data. Then, all available questionnaires are loaded from a central backend. If the loading is unsuccessful (e.g., no connection to the server can be established), locally stored data is used until the next synchronization attempt. In case there are no locally stored questionnaires, the synchronization attempt is retried until it succeeds. The app then checks if there are *first usage* (i.e., questionnaires that are only answered once after the first login) or *one-time* (i.e., questionnaires that are only answered once but might be answered at a later time) questionnaires available. If this is the case, these questionnaires are displayed and can be filled in by the user one after the other. Data is then synchronized with the backend by uploading all newly answered questionnaire data and loading all studies the user is subscribed to. If the synchronization is unsuccessful, the local storage is checked once again and the process is retried after some time if no data can be retrieved both remotely and locally. In the next step, an overview of all available studies is presented to the user. He/She may then select a study from that overview. Depending on the study and the user's subscription status, the following process differs. If the user is currently not subscribed to the study, he/she will be able to (a) directly subscribe to that study if it is *public*, or (b) be prompted to enter a password if it is a *private* study. For private studies,

[1]https://www.trackyourtinnitus.org/

[2]https://www.trackyourtinnitus.org/about

TABLE 1 | Examples of apps developed by the authors combining mobile crowdsensing (MCS) and ecological momentary assessments (EMA), compared according to their respective features.

App name	Guidance	Feedback	Adjustable prompts	Dynamic questionnaires	Participatory sensing	Opportunistic sensing
TrackYourTinnitus (TYT)			✓	✓	✓	
TrackYourHearing (TYH)[a]			✓	✓	✓	
TrackYourDiabetes (TYD)	✓	✓	✓		✓	
TrackYourStress (TYS)[b]	✓	✓	✓		✓	
TinnitusTipps	✓	✓	✓	✓	✓	
KINDEX	✓	✓		✓	(✓)	
Intersession	✓	✓	✓	✓	(✓)	

[a] *https://www.trackyourhearing.org/*
[b] *https://www.trackyourstress.org/*

TABLE 2 | Descriptive statistics on mobile crowdsensing EMA apps developed by the authors.

App	Number of total users	Number of users with at least one answer sheet[†]	Submitted answer sheets	Sensor measurements
TrackYourTinnitus (TYT)	4,480	2,905	76,105	Environmental sound level
TrackYourHearing (TYH)	437	167	6,102	Environmental sound level, EEG*
TrackYourDiabetes (TYD)	58	36	3,097	Position (GPS), environmental sound level, blood sugar*
TrackYourStress (TYS)	204	138	2,989	Position (GPS), environmental sound level, heart rate sensor
TinnitusTipps	95	66	8,209	Position (GPS)
KINDEX	1,779	1,779	1,943	–
Intersession	6	4	220	–
Total	7,059	5,095	98,665	

Numbers extracted on 05 Dec 2019.
**External sensor measurements.*
[†]*Compared to the second column, this column does not include users that quit using the app after registration and are therefore considered as early dropouts.*

the password is then checked with the backend. If the password is correct, the user is subscribed to the study. Otherwise, an error hint is displayed and the user is redirected back to the study overview. If the user is currently subscribed to the study and that study is already finished, its details are loaded from the backend and the user is forwarded to the main menu. If the user is currently subscribed to the study and that study is still running, the user is also forwarded to the main menu. From the main menu, the user can choose to go back to the study overview, display his/her results, fill in questionnaires and perform sensor measurements, and finally, change the settings. From the results, questionnaire and settings views, he/she can always return to the main menu. If the user selects the study overview, or if the study period is expired (respectively, if the study is finished), the study overview is displayed once again.

During the development and advancement of the platform, we faced several challenges and peculiarities. Additionally, we gained some valuable insights when implementing such a combination of an EMA and MCS approach. First, we required a basic functionality to identify different users. One could argue that, since data has to be stored anonymized, a device ID would be sufficient, but this would prevent the users from changing devices without data loss. Therefore, we implemented basic authentication and authorization mechanisms, including registration via email, login with username and password, as well as password reset features.

The core of the application is the presentation and fill-in process of (EMA) questionnaires. In order to facilitate adding new questionnaires and adjusting existing questionnaires at a later time, the platform should offer a generic approach to handle questionnaires. We achieved this by defining the questionnaires as *JavaScript Object Notation (JSON)* objects containing an array of *questionnaire elements* (e.g., headline, text, multiple-choice-question), stored on the backend. The apps provide components with functionalities to render, configure, and handle the input for each of these elements. The components are then put together in a *list view*, and additional checks like input validation or ensuring that required questions are filled in are performed. Another requirement was to make the questionnaires easy to use, while not introducing bias. In order to improve usability of the apps, we tried to make the questionnaires look similar to their paper-pencil counterpart while using as many system-provided and default UI elements as possible when implementing the element components. However, some of the default UI elements are not suitable for the use in psychological questionnaires and had to be adjusted. For instance, default iOS and Android sliders have a pre-selected value, which fosters undesirable anchoring affects (Tversky and Kahneman, 1974).

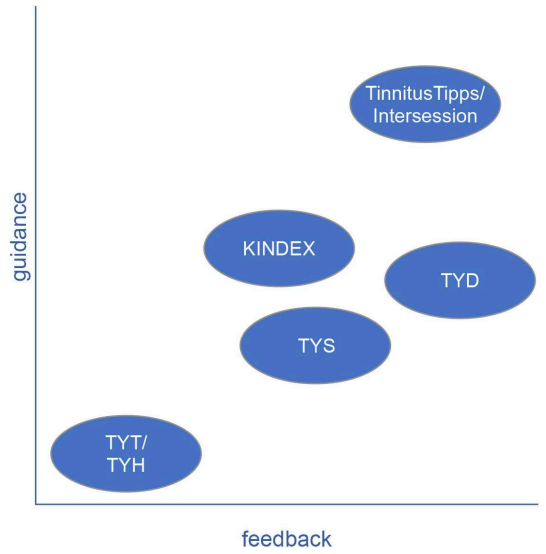


FIGURE 1 | Subjective and relative rating of guidance and feedback of selected EMA apps developed by the authors.

TABLE 3 | Questions of the daily questionnaire in the TrackYourTinnitus (TYT) smartphone application, along with their scale and the dimension they measure (Schlee et al., 2016; Pryss et al., 2017).

#	Question	Scale	Dimension
1	Did you perceive the tinnitus right now?	BS	Perception
2	How loud is the tinnitus right now?	VAS	Loudness
3	How stressful is the tinnitus right now?	VAS	Distress
4	How is your mood right now?	VAS	Mood
5	How is your arousal right now?	VAS	Arousal
6	Do you feel stressed right now?	VAS	Stress
7	How much did you concentrate on the things you are doing right now?	VAS	Concentration
8	Do you feel irritable right now?	BS	Irritability

BS, binary scale; VAS, visual analog scale.

Furthermore, a sophisticated algorithm has to be deployed in order to implement the notification (i.e., prompting) schedules. The algorithm has to account for the users' sleep and work schedules, ensuring notifications are not too close to each other[3] and allowing different adjustments by the user (e.g., the time frame and number of notifications per day). Since we managed the notification schedules exclusively inside the apps, there was no way to retrieve any information on the scheduled and received notifications. Therefore, we were unable to extract valuable information on how users change their notification schedules and, most importantly, we could not evaluate the notification adherence. We offered both *random* (in a given time frame, adjustable by the user) and *fixed* (at an exact point in time, chosen by the user) notifications. However, users reported problems with random notifications not being delivered as configured or not delivered at all. While fixed notifications have proven to be more reliable, more flexible, and less disruptive to the user, their value in terms of EMA is to be questioned. Users might integrate answering the questionnaire into their daily routine, which can lead to a possible bias.

In our first version of the app, we incorporated an environmental sound measurement. If enabled by the user, the app tracks the average loudness recorded by the smartphone microphone while the user answers the questionnaires. This value is then stored together with the questionnaire data and can be correlated to gain new valuable insights on tinnitus and its interrelations with environmental sound. However, due to manufacturer and device model differences, measurements are not comparable across users. Calibrations with different device models or other measures to ensure comparability should be performed before integrating similar measurements into mobile applications. Additionally, these sensor measurements are hard-coded into the apps. A dynamic framework to integrate internal and external sensors would facilitate studies aiming to correlate different sensor data with questionnaire data. In this way, one could integrate additional sensors, e.g., positioning with GPS in order to investigate the interrelations to motion patterns or the influence of weather-related factors.

Another aspect worth considering is *incentives*. There needs to be some sort of motivation for users to continuously submit data. Zhang et al. (2015) divide incentives in mobile crowd sensing applications into entertainment, service, and monetary incentives. Since we do not consider monetary incentives sustainable in the long term (especially in the research context), we focus on the former two categories in order to increase the users' extrinsic and intrinsic motivation. While in TYT, we provided some minimalistic feedback in the form of a chart of the perceived tinnitus loudness and an option to review the history of submitted questionnaires for each individual user, we believe the main incentive for users is the contribution to research on a chronic disorder from which they are suffering. However, more than 78% of users drop out after 10 days of participation. More incentive mechanisms, like advanced feedback, gamification, or social features should be implemented (Agrawal et al., 2018).

In order to perform different studies with the app (and to exclude test users from the actual data set), the need to separate users into study groups inside the app emerged. We updated the app to incorporate a basic study allocation. Users are able to join studies by manually selecting them from a list inside the app. However, users can currently only be member of a single study at a time and there is no functionality in place for the study manager to control or verify which user joins which study without checking the database.

Since mobile devices are not guaranteed to always be connected to the internet (i.e., *be online*), the app should also be functional without internet connection whenever possible. TYT offers a basic *offline functionality* by initially downloading all questionnaires and storing them on the device. Additionally, the users' given answers for questionnaires are cached on the device if there is no internet connectivity until the connection is restored. This way, the feedback features also remain functional. However, other features, e.g., the study management, are only available if the device is online.

[3]We chose 15 min as minimal distance between two consecutive notifications.

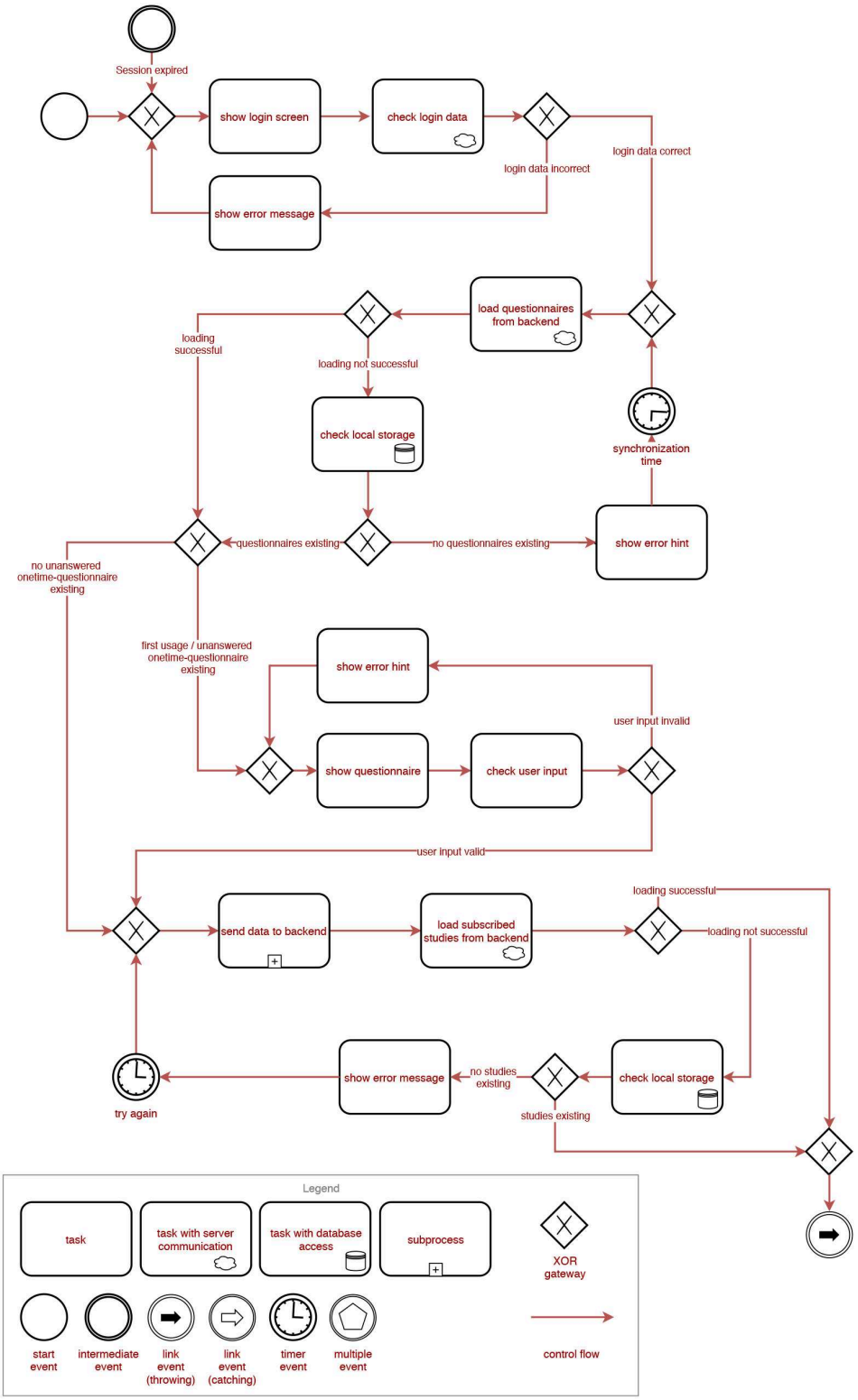


FIGURE 2 | BPMN representation of the general process of the TrackYourTinnitus (TYT) smartphone application (Part 1).

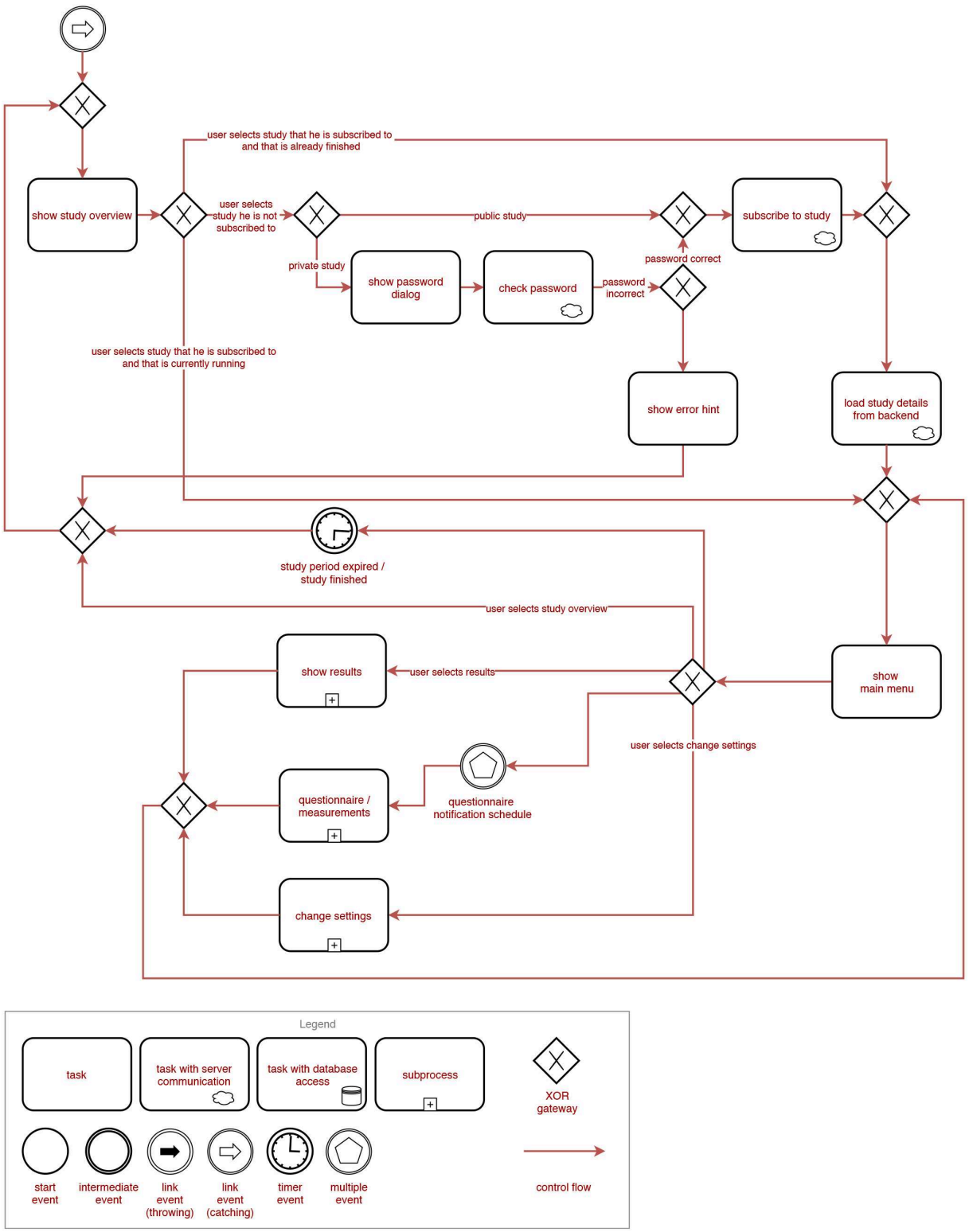


FIGURE 3 | BPMN representation of the general process of the TrackYourTinnitus (TYT) smartphone application (Part 2).

Furthermore, *safety, security, and privacy* are aspects of high importance in the healthcare domain. Region-specific regulations, e.g., the *General Data Protection Regulation (GDPR)* and the *Medical Device Regulation (MDR)* in the EU, as well as high expectations of patients need to be considered when designing a software system in this field. TYT applies state-of-the-art security measures with an email verification as part of the registration process, credential-based authentication and token-based authorization (see above) as well as encrypted data transmission via SSL/TLS (Rescorla, 2018). Health risks are outlined on the website. However, since safety, security, and privacy requirements are constantly evolving, a more transparent informed consent, additional security measures and a privacy-preserving design would be desirable for the future (e.g., Beierle et al., 2019).

Data quality of the submitted data is another critical issue in MCS-EMA apps (see section 2.2.2). As already discussed above, reliable sensor and comparable measurements on mobile devices are difficult to achieve due to the variety of device models. But, also for the questionnaire data, no real statement can be made regarding its quality. Since we only require the user to answer two of the eight questions in the daily questionnaire, users can skip most of the questions if they would like to do so, which leads to missing values. Also, if users feel forced to answer a question or have malicious intentions, they might provide untruthful data. In addition, the use of a smartphone to gather large amounts of personal data in real life that is stored to a large database for scientific research could boost competition thoughts. Consequently, participants might provide data only for the purpose of providing more data than others. Such factors should be taken into account and mechanisms should be in place to cope with data quality.

Moreover, scientists providing the platform and HCPs want to analyze the collected data. In TYT, data analysis is only possible in a static way by querying the raw database. More flexible, on-demand analysis functionalities for scientists evaluating the platform data are desirable. Furthermore, HCPs and their patients could benefit from a dynamic analysis of the patients' data, providing detailed insights and building the baseline for tailored feedback.

Finally, the experiences gained with TrackYourTinnitus and the projects shown in **Table 2** are discussed in the light of their general contribution and their generalizability. A recent review of mobile health crowdsensing research (Tokosi and Scholtz, 2019) shows that the projects shown in **Table 2** and the related papers are heavily recognized by their selected key terms of existing works. Tokosi and Scholtz (2019) also shows that although more and more research is pursued in this context, less experiences are reported that were gained over multiple large-scale and long-running projects. Therefore, we consider our experiences as a proper starting point to conceive a reference architecture that incorporates aspects that are relevant on one hand. On the other, these aspects have shown their importance at multiple times. Furthermore, the authors have already worked on better generic solutions for parts of the reference architecture. For example, for the REST interface (see **Figure 4**) in Pryss et al. (2018b), a more generic solution was proposed. This solution, in turn, is utilized by all projects shown in **Table 2** that have been started after TrackYourTinnitus. However, as for other purposes, like mobile data collection, better generic solutions have been proposed (e.g., Schobel et al., 2019). A configurable crowdsensing platform based on (1) the archetype shown in (Schobel et al., 2019) and (2) the results of this work is currently conceived. Moreover, developments, such as *PACO*[4] show that easily customizable MCS-EMA apps are highly welcome by users. In addition, commercial tools, such as *ilumivu*[5] emphasize the need of generic solutions in the given context of EMA and mobile crowdsensing. Thereby, the ilumivu technical solution provides already sophisticated features for EMA apps on a generic level. Importantly, these features deal with many aspects raised in this work. On the other, ilumivu still does not consider all of the discussed aspects. For example, ilumivu does not convey how they cope with a management of incentives. Following this, the work at hand can be utilized to reflect existing solutions or new developments with the shown experiences and derived recommendations, especially as they are gained over time and across projects. We do not claim that these recommendations are complete or cover every aspect, but we consider them as a proper starting point for various projects and questions in the context of healthcare and the combination of mobile crowdsensing and EMA.

4. TOWARD A REFERENCE ARCHITECTURE

Based on the findings in section 3, we derive a number of recommendations for a mature and contemporary MCS-EMA platform. We then propose a reference architecture to address these recommendations and discuss technical considerations with respect to the implementation.

4.1. Recommendations

We derived twelve recommendations from the lessons learned during the TYT project (see section 3), various discussions with colleagues and domain experts, as well as general considerations when building a modern software system. Namely, these recommendations are (R1) *User Identity*, (R2) *Generic Questionnaires*, (R4) *Sensors and Context-Awareness*, (R5) *Incentive Mechanisms*, (R6) *Groups, Studies and HCPs*, (R7) *High Availability and Performance*, (R8) *Offline Availability*, (R9) *Safety, Security, and Privacy*, (R10) *Data Quality*, (R11) *Data Analysis*, and (R12) *Interoperability*. The recommendations are described in detail in **Tables 4**, **5**.

4.2. Architecture

Based on the recommendations defined in section 4.1, we propose a reference architecture for a platform supporting the combination of mobile crowdsensing and ecological momentary assessments in the healthcare domain. **Figure 4** shows the general architecture. It comprises a central backend with different services, a database and a file server, as well as mobile apps

[4]https://pacoapp.com/
[5]https://ilumivu.com/

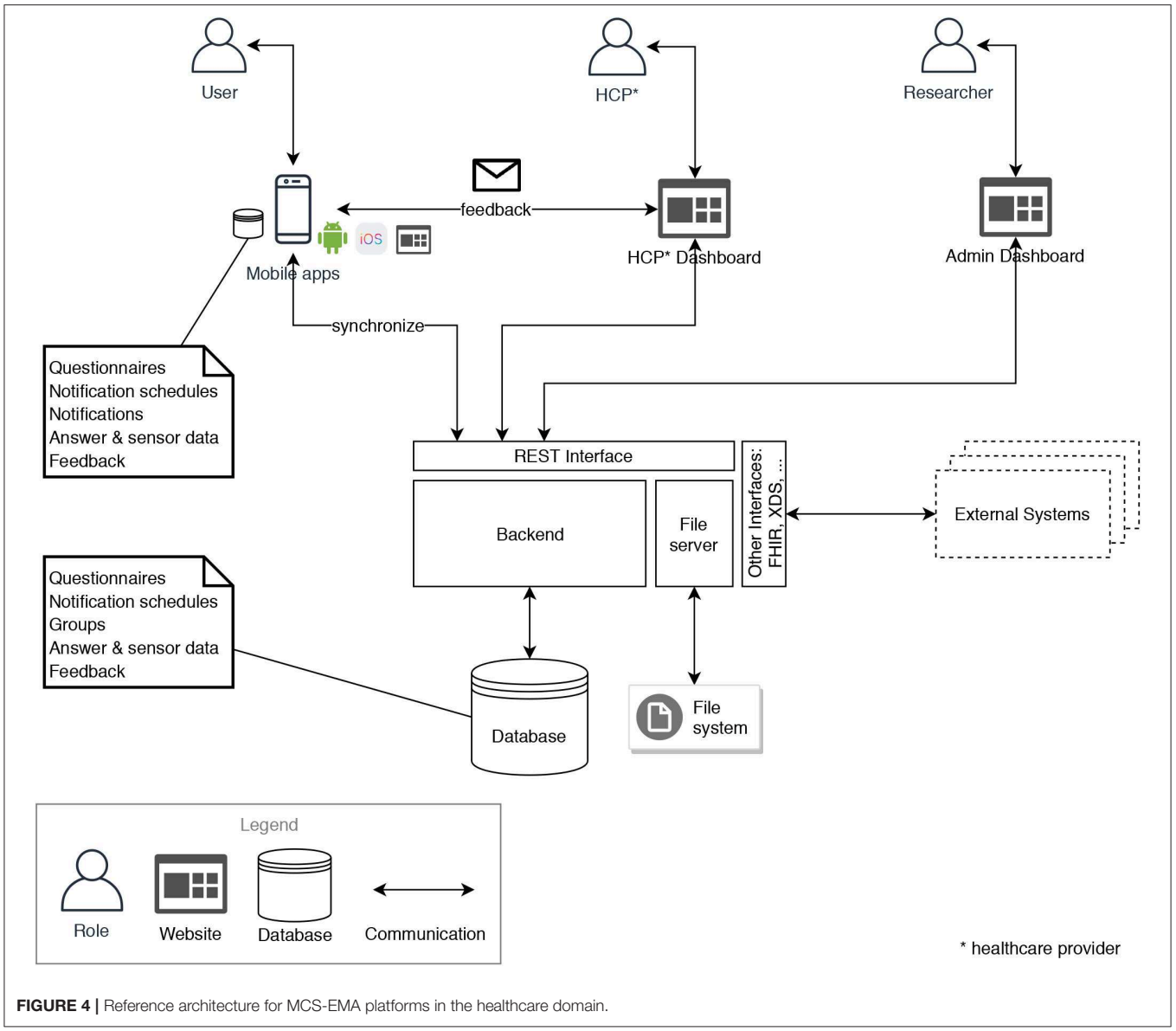


FIGURE 4 | Reference architecture for MCS-EMA platforms in the healthcare domain.

for both Android and iOS, a web dashboard for HCPs and another web dashboard for system administrators (*admins*). The clients (mobile apps, HCP dashboard and admin dashboard) communicate with the backend via a RESTful interface. Files, like multimedia and documents, are stored on a file server. Relevant files are downloaded and additionally stored on the mobile devices. All relevant data, like questionnaires, notification schedules as well as answer and sensor data are synchronized between the central database in the backend and the mobile apps' local databases. The backend additionally provides other interfaces for external systems, implementing common standards in the healthcare domain.

4.3. Selected Technical Considerations

Furthermore, we discuss technical considerations in order to address some of the architectural aspects of the defined recommendations in respect to our reference architecture. First, in order to achieve high availability, the system has to be scalable, and in the best case, elastic. According to definitions provided by Herbst et al., *scalability* is "the ability of a system to handle increasing workloads with adequate performance," while *elasticity* is "the degree to which a system is able to adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible" (Herbst et al., 2013). We suggest to use a *cloud-native* approach to address these recommendations. A *cloud-native application (CNA)* is explicitly designed to be operated in the cloud. Therefore, such application is—by design—distributed, elastic, and horizontally scalable. Furthermore, it is composed of microservices with a minimum of isolated states (Kratzke and Quint, 2017). The internal architecture for a cloud-native

TABLE 4 | Recommendations for a platform combining mobile crowdsensing (MCS) and ecological momentary assessments (EMA) in the healthcare domain (Part 1).

ID	Name	Description
R1	User identity	The platform should allow authentication and authorization in order to uniquely identify users. The user should be able to log into the platform with multiple devices, change and recover his/her password if it is lost, and deactivate as well as delete his/her account.
R2	Generic questionnaires	The platform should be able to handle generically defined questionnaires. Both one-time (e.g., demographic) and repeating (e.g., EMA) questionnaires should be supported. The mobile application should be able to display multiple questionnaires, which are available at different intervals, concurrently. Supported question types should be at least *single choice*, *multiple choice*, *text input*, and *date input*. There should be an option to define dynamic questionnaires, which adapt to the previous input of the user (i.e., *conditional content*). Optionally, the user can also adapt his/her own questionnaire according to his/her needs (e.g., add additional questions).
R3	Notifications	The platform should be able to prompt the user to fill in questionnaires. For each questionnaire, one or multiple *notification schedules* can be defined, which determines how and how often the user is notified. A default configuration for each questionnaire can be provided, which is optionally adjustable by the user. Notifications can be set for fixed times (i.e., *fixed*), or randomly within a given time frame for each day (i.e., *random*). An algorithm should ensure that notifications from different schedules are not conflicting with each other. Additionally, notifications that are event-triggered (e.g., by a context change) can be defined. Information on the notification adherence (i.e., when the notification has been displayed; if/when did the user trigger the notification) should be stored and made available for analysis.
R4	Sensors and context-awareness	For each questionnaire, a set of sensor measurements (e.g., GPS coordinates, sound level, brightness, or wearable sensors) that are performed on the mobile devices should be definable. These measurements can be configured to be performed (a) once or (b) continuously during the fill-in process of the respective questionnaire; (c) continuously during the app usage; or (d) continuously in the background. Additionally, different sensors can be combined (i.e., *sensor fusion*) to retrieve various context information.
R5	Incentive mechanisms	Different incentive mechanisms should be deployed in order to support the patients' adherence. We define three types of incentives: *feedback*, *gamification*, or *social features*.
R5.1	Feedback	The platform should provide different types of feedback to the user. Graphical feedback (e.g., charts or graphs), daily tips, automatic feedback based on the given answers, as well as manual feedback in the form of messages by the HCP can be incorporated. Manual feedback could be supported or partly be replaced by incorporating a chatbot with automated analysis of the user's input (both answer data and text messages).
R5.2	Gamification	The platform should offer gamification features like achievements (e.g., submission streaks), badges, points, and leaderboards.
R5.3	Social features	The platform should offer social features like public user profiles, group chats, discussion boards on certain topics and following as well as sharing functionalities.

implementation of the backend in our reference architecture is shown in **Figure 5**. The backend can be decomposed to multiple microservices, and these microservices can then be replicated in order to enable horizontal scalability. Optimally, the database, file server and file system should be distributed and/or replicated as well. In order to provide elasticity, an *orchestration system* is used to monitor metrics describing the load of the system and automatically orchestrate resources based on these metrics in order to scale in and scale out. A common approach would be to use Docker[6] as container technology to implement microservices and Kubernetes[7] (Burns et al., 2016) as container-orchestration system.

In order to provide high levels of security and privacy, all communication between different components of the architecture should be encrypted. All personal and private user data should be stored separately from the application data to reduce the risk of it being exposed in case of a data breach. Optionally, in a *privacy-preserving design*, this data should be encrypted in a way that it can only be decrypted by each respective user himself. In the best case, a dedicated privacy model is incorporated or developed (e.g., Beierle et al., 2019).

Furthermore, for the development of the mobile apps, it has to be decided whether to develop a native app for each target platform (e.g., Android, iOS, web browser) or use cross-platform frameworks that enable the developer to use a single code-base and deploy this code to different platforms. We recommend to use cross-platform frameworks (e.g., *Xamarin*[8], *Flutter*[9], or *Ionic*[10]) for small developer teams and teams which are prone to changes (e.g., research projects), since the single code base requires less efforts for development and maintenance, as well as causes lesser heterogeneity-based challenges in programming languages and tools, which makes it easier for new developers to enter the team. However, for bigger and more consistent developer teams, native app development might be better suited. Native apps might provide a better interface to the operating system and therefore more control over sensors and the user interface, as well as potentially better performance. This has special value to MCS apps incorporating advanced sensor usage.

[6]https://www.docker.com/
[7]https://kubernetes.io/
[8]https://dotnet.microsoft.com/apps/xamarin
[9]https://flutter.dev/
[10]https://ionicframework.com/

TABLE 5 | Recommendations for a platform combining mobile crowdsensing (MCS) and ecological momentary assessments (EMA) in the healthcare domain (Part 2).

ID	Name	Description
R6	Groups, studies, and HCPs	Users should be able to join one or multiple groups. These groups can represent studies, HCPs or other groupings (e.g., test users). Users can be invited to groups by their respective group owner (e.g., the HCP) or join them via different join mechanisms (e.g., join requests, password-restricted or freely).
R7	High availability and Performance	The platform should be available to its users in the best possible way. There should not be any noticeable performance drops under higher loads.
R8	Offline availability	The mobile app should still be functional when there is no internet connection (or more generally, no connection to the server) whenever possible. All data should be stored on the device where appropriate and synchronized with the server.
R9	Safety, security, and privacy	The platform should meet high safety, security and privacy standards. Region-specific regulations like the *EU General Data Protection Regulation (GDPR)* and the *Medical Device Regulation (MDR)* should be considered. All confidential data should be stored securely and transmitted in encrypted form. User data and credentials should be stored separately from the answer data. Health risks should be identified and addressed at an early stage and outlined to users and HCPs in a transparent way. A security model for the mobile apps and the entire platform should exist.
R10	Data quality	Data quality should be kept as high as possible. Different data quality aspects like believability, relevancy, accuracy (i.e., error-free, reliable, precise), interpretability, understandability, accessibility, objectivity, timeliness, completeness and (representational) consistency (Wang and Strong, 1996) should be addressed depending on the specific requirements of the use case. The platform should perform input validation and prevent invalid inputs, perform plausibility checks, as well as other measures to improve quality of answer and sensor data. This also includes measures for detecting and handling misstatements by users, which might be both intentional and malicious (e.g., *faking*), as well as unintentional (e.g., *self-deception*), summarized with the terms *faking* and *socially desirable responding (SDR)* (Paulhus, 2001; Van de Mortel, 2008).
R11	Data analysis	The platform should offer easy-to-use data analysis functionalities on live data for researchers, HCPs, and also the users themselves. Both static and dynamic data analysis (e.g., aggregation with the help of filters and time windows or clustering) should be enabled. All relevant data should be exportable to common formats (e.g., CSV, SPSS, R, PDF). The HCP and the user should be able to review and analyze the individual answers to questionnaires as well as sensor measurements and compare them to the data of other users.
R12	Interoperability	The platform should offer a good interoperability with other (external) systems. This includes implementing common data exchange format standards and communication protocols, as well as providing uniform, understandable, and well-documented interfaces.

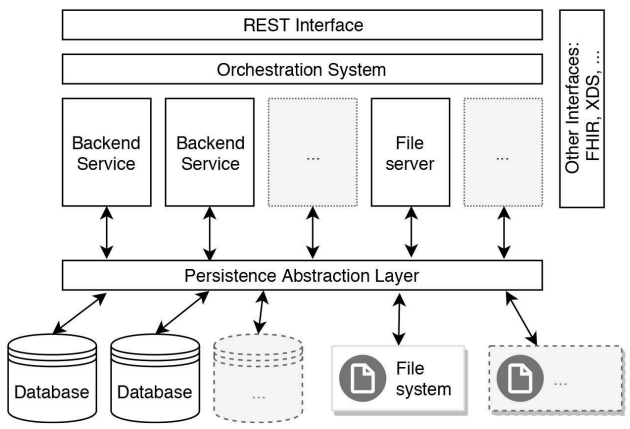


FIGURE 5 | Scalable design of a backend in the reference architecture for MCS-EMA platforms in the healthcare domain.

Finally, in order to provide good interoperability with other internal as well as external systems, common interfaces should be provided. This includes state-of-the-art architectural styles in web technology like *REST* (Fielding and Taylor, 2000; Pryss et al., 2018b), but also standards in the healthcare domain [e.g., *FHIR*[11] or *XDS* (Trotter and Uhlman, 2011)]. Standards that one wants to support should be considered at an early stage when designing the data models.

5. DISCUSSION

We argue that, when considering mobile crowdsensing in the healthcare domain, differentiating only between participatory and opportunistic sensing is not sufficient. Other aspects like context-awareness, incentive mechanisms, groups, security, and privacy, data quality, as well as technical aspects like availability, performance, offline availability and interoperability should be also thoroughly taken into account. Additionally, although personal sensing data on its own only belongs to an individual user, it can be used in order to be beneficial for the community as a whole by processing, clustering, and correlating this type of data. Therefore, we further argue that in the context of mobile crowdsensing in healthcare, there is no distinct separation between community sensing and personal sensing, and that both concepts should be considered depending on the scenario that is addressed.

Furthermore, in the literature, MCS and EMA are considered as separate, mostly unrelated concepts. While they have different origins, we argue that both concepts make use of similar approaches, namely leveraging the crowd and their (already

[11] https://www.hl7.org/fhir/

existing) mobile devices in order to assess phenomena of common interest. Therefore, they should be considered closely related to each other, and their combination should get more awareness. Beyond that, the architectural model is often not provided in publications on MCS and EMA studies, although we argue that it has meaningful implications on the comparability of their results. We believe that a reference architecture, such as that introduced in this work, can raise awareness and counteract this issue to a certain degree. In this context, we have particularly shown which aspects the reference architecture incorporates to develop more generic technical solutions based on it.

6. CONCLUSION

In this work, we discussed the combination of mobile crowdsensing (MCS) and ecological momentary assessment (EMA) in the healthcare domain. We introduced both terms and described how we considered their underlying concepts that are similar to each other, which fosters combining MCS and EMA in a single approach. Furthermore, we discussed the lessons we learned from the TrackYourTinnitus project, which is running for over 5 years. Based on these findings, we derived recommendations for a platform supporting the combination of MCS and EMA in the healthcare domain. We then proposed a reference architecture for such a platform, described its components and how they interact. Additionally, we outlined how the reference architecture could be implemented in order to address the defined recommendations from the technical side. Furthermore, we discussed how MCS and EMA research should be considering both concepts in combination and propose that publications in this field should refer to the used architectural model.

In conclusion, one can see that there are numerous conceptual, architectural and technical, as well as legal challenges when designing a MCS-EMA platform for the healthcare domain. We believe that the defined recommendations can—adjusted to the individual factors, needs and requirements of a (research) project or product—act as foundation for future MCS-EMA systems. All the different aspects should be considered at an early stage of the project. Additionally, the reference architecture can serve as a generic template for a platform implementation. Technical considerations should be kept in mind in order to be able to scale and cope with future requirements. However, we believe that the combination of MCS and EMA is a promising approach for many different use cases in the healthcare domain. For this endeavor, our reference architecture and recommendations shall be a basis for more generic and comparable technical solutions.

AUTHOR CONTRIBUTIONS

RK and RP substantially contributed to the TrackYourTinnitus platform, drafted, and revised the manuscript. WS, MS, MR, BL, and TP substantially contributed to the TrackYourTinnitus platform and revised the manuscript. HB read and revised the manuscript. RH substantially contributed to the TinnitusTipps platform and revised the manuscript.

REFERENCES

Agrawal, K., Mehdi, M., Reichert, M., Hauck, F., Schlee, W., Probst, T., et al. (2018). Towards incentive management mechanisms in the context of crowdsensing technologies based on trackyourtinnitus insights. *Proc. Comput. Sci.* 134, 145–152. doi: 10.1016/j.procs.2018.07.155

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., et al. (2019). What data are smartphone users willing to share with researchers? *J. Ambient Intell. Human. Comput.* 2019, 1–13. doi: 10.1007/s12652-019-01355-6

Burke, J. A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., et al. (2006). *Participatory Sensing.* Los Angeles, CA: UCLA; Center for Embedded Network Sensing.

Burns, B., Grant, B., Oppenheimer, D., Brewer, E., and Wilkes, J. (2016). *Borg, Omega, and Kubernetes.* New York, NY: ACM Queue.

Consolvo, S., Everitt, K., Smith, I., and Landay, J. A. (2006). "Design requirements for technologies that encourage physical activity," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Montreal, QC: ACM), 457–466.

Ebner-Priemer, U. W., and Kubiak, T. (2007). Psychological and psychophysiological ambulatory monitoring. *Eur. J. Psychol. Assess.* 23, 214–226. doi: 10.1027/1015-5759.23.4.214

Fielding, R. T., and Taylor, R. N. (2000). *Architectural styles and the design of network-based software architectures* (Doctoral dissertation), University of California, Irvine, Irvine, CA, United States.

Ganti, R. K., Ye, F., and Lei, H. (2011). Mobile crowdsensing: current state and future challenges. *IEEE Commun. Mag.* 49, 32–39. doi: 10.1109/MCOM.2011.6069707

Guo, B., Wang, Z., Yu, Z., Wang, Y., Yen, N. Y., Huang, R., et al. (2015). Mobile crowd sensing and computing: the review of an emerging human-powered sensing paradigm. *ACM Comput. Surv.* 48:7. doi: 10.1145/2794400

Herbst, N. R., Kounev, S., and Reussner, R. (2013). "Elasticity in cloud computing: what it is, and what it is not," in *Proceedings of the 10th International Conference on Autonomic Computing (ICAC 13)* (San Jose, CA), 23–27.

Kratzke, N., and Quint, P.-C. (2017). Understanding cloud-native applications after 10 years of cloud computing–a systematic mapping study. *J. Syst. Softw.* 126, 1–16. doi: 10.1016/j.jss.2017.01.001

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., and Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Commun. Mag.* 48, 140–150. doi: 10.1109/MCOM.2010.5560598

Larson, R., and Csikszentmihalyi, M. (2014). "The experience sampling method," in *Flow and the Foundations of Positive Psychology*, ed M. Csikszentmihalyi (Dordrecht: Springer), 21–34.

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., et al. (2009). "Peir, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services* (Kraków: ACM), 55–68.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., and Van Os, J. (2009). Experience sampling research in psychopathology: opening the black box of daily life. *Psychol. Med.* 39, 1533–1547. doi: 10.1017/S0033291708004947

Paulhus, D. L. (2001). "Socially desirable responding: the evolution of a construct," in *The Role of Constructs in Psychological and Educational Measurement*, eds H. I. Braun, D. N. Jackson, and D. E. Wiley (Abingdon: Routledge), 61–84.

Pryss, R. (2019). "Mobile crowdsensing in healthcare scenarios: taxonomy, conceptual pillars, smart mobile crowdsensing services," in *Digital Phenotyping and Mobile Sensing*, eds H. Baumeister and C. Montag (Cham: Springer), 221–234.

Pryss, R., John, D., Schlee, W., Schlotz, W., Schobel, J., Kraft, R., et al. (2019). Exploring the time trend of stress-levels while using the crowdsensing mhealth platform trackyourstress and the influence of stress reactivity. *JMIR mHealth uHealth* 7:e13978. doi: 10.2196/13978

Pryss, R., Probst, T., Schlee, W., Schobel, J., Langguth, B., Neff, P., et al. (2018a). Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the trackyourtinnitus mobile platform. *Int. J. Data Sci. Anal.* 8, 327–338. doi: 10.1007/s41060-018-0111-4

Pryss, R., Reichert, M., Herrmann, J., Langguth, B., and Schlee, W. (2015). "Mobile crowd sensing in clinical and psychological trials–a case study," in *2015 IEEE 28th International Symposium on Computer-Based Medical Systems* (São Carlos: IEEE), 23–24.

Pryss, R., Schlee, W., Langguth, B., and Reichert, M. (2017). "Mobile crowdsensing services for tinnitus assessment and patient feedback," in *2017 IEEE International Conference on AI & Mobile Services (AIMS)* (Honolulu, HI: IEEE), 22–29.

Pryss, R., Schobel, J., and Reichert, M. (2018b). "Requirements for a flexible and generic api enabling mobile crowdsensing mhealth applications," in *2018 4th International Workshop on Requirements Engineering for Self-Adaptive, Collaborative, and Cyber Physical Systems (RESACS)* (Banff, AB: IEEE), 24–31.

Raento, M., Oulasvirta, A., and Eagle, N. (2009). Smartphones: an emerging tool for social scientists. *Sociol. Methods Res.* 37, 426–454. doi: 10.1177/0049124108330005

Reichert, M., and Pryss, R. (2017). Flexible support of healthcare processes. *Process Model. Manag. Healthc.* 1, 35–66. doi: 10.1201/b21884-2

Rescorla, E. (2018). *The Transport Layer Security (TLS) Protocol Version 1.3*. Technical report, RFC Editor.

Ruf-Leuschner, M., Brunnemann, N., Schauer, M., Pryss, R., Barnewitz, E., Liebrecht, M., et al. (2016). The kindex-app–an instrument for assessment and immediate analysis of psychosocial risk factors in pregnant women in daily practice by gynecologists, midwives and in gynecological hospitals. *Verhaltenstherapie* 26, 171–181. doi: 10.1159/000448455

Schlee, W., Pryss, R. C., Probst, T., Schobel, J., Bachmeier, A., Reichert, M., et al. (2016). Measuring the moment-to-moment variability of tinnitus: the trackyourtinnitus smart phone app. *Front. Aging Neurosci.* 8:294. doi: 10.3389/fnagi.2016.00294

Schobel, J., Probst, T., Reichert, M., Schickler, M., and Pryss, R. (2019). Enabling sophisticated lifecycle support for mobile healthcare data collection applications. *IEEE Access* 7, 61204–61217. doi: 10.1109/ACCESS.2019.2916142

Schobel, J., Pryss, R., and Reichert, M. (2015). "Using smart mobile devices for collecting structured data in clinical trials: results from a large-scale case study," in *2015 IEEE 28th International Symposium on Computer-Based Medical Systems* (São Carlos: IEEE), 13–18.

Schweizer, I., Bärtl, R., Schulz, A., Probst, F., and Mühläuser, M. (2011). "NoiseMap–real-time participatory noise maps," in *Second International Workshop on Sensing Applications on Mobile Phones* (Seattle, WA: Citeseer), 1–5.

Shiffman, S. (2007). "Designing protocols for ecological momentary assessment," in *The Science of Real-Time Data Capture: Self-Reports in Health Research*, eds A. Stone, S. Shiffman, A. Atienza, and L. Nebeling (Oxford: Oxford University Press), 27–53.

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

Stone, A. A., and Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. *Ann. Behav. Med.* 16, 199–202. doi: 10.1093/abm/16.3.199

Stone, A. A., and Shiffman, S. (2002). Capturing momentary, self-report data: a proposal for reporting guidelines. *Ann. Behav. Med.* 24, 236–243. doi: 10.1207/S15324796ABM2403_09

Tokosi, T. O., and Scholtz, B. M. (2019). "A classification framework of mobile health crowdsensing research: a scoping review," in *Proceedings of the South African Institute of Computer Scientists and Information Technologists 2019* (Skukuza: ACM), 4.

Trotter, F., and Uhlman, D. (2011). *Hacking Healthcare: A Guide to Standards, Workflows, and Meaningful Use*. Sebastopol, CA: O'Reilly Media, Inc.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Van Berkel, N., Ferreira, D., and Kostakos, V. (2018). The experience sampling method on mobile devices. *ACM Comput. Surv.* 50:93. doi: 10.1145/3123988

Van de Mortel, T. F. (2008). Faking it: social desirability response bias in self-report research. *Aust. J. Adv. Nurs.* 25:40.

Van Haren, S. (2018). *An exploratory study into the use of mobile experience sampling as a method for assessing infant language development* (Master's thesis), Tilburg University, Tilburg, Netherlands.

Wang, R. Y., and Strong, D. M. (1996). Beyond accuracy: what data quality means to data consumers. *J. Manag. Inform. Syst.* 12, 5–33. doi: 10.1080/07421222.1996.11518099

Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., et al. (2012). Quantifying the impact of human mobility on malaria. *Science* 338, 267–270. doi: 10.1126/science.1223467

Zappatore, M., Longo, A., and Bochicchio, M. A. (2017). Crowd-sensing our smart cities: a platform for noise monitoring and acoustic urban planning. *J. Commun. Softw. Syst.* 13:53. doi: 10.24138/jcomss.v13i2.373

Zhang, X., Yang, Z., Sun, W., Liu, Y., Tang, S., Xing, K., et al. (2015). Incentives for mobile crowd sensing: a survey. *IEEE Commun. Surv. Tutor.* 18, 54–67. doi: 10.1109/COMST.2015.2415528

Conflict of Interest: RH was employed by the company Sivantos GmbH, Germany.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 06 March 2020
doi: 10.3389/fnins.2020.00174

Motorized Shoes Induce Robust Sensorimotor Adaptation in Walking

*Yashar Aucie[1], Xunjie Zhang[2], Randy Sargent[3] and Gelsy Torres-Oviedo[1]**

[1] Department of Bioengineering, Swanson School of Engineering, University of Pittsburgh, Pittsburgh, PA, United States, [2] Nimbus Robotics, Pittsburgh, PA, United States, [3] The Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

The motor system has the flexibility to update motor plans according to systematic changes in the environment or the body. This capacity is studied in the laboratory through sensorimotor adaptation paradigms imposing sustained and predictable motor demands specific to the task at hand. However, these studies are tied to the laboratory setting. Thus, we asked if a portable device could be used to elicit locomotor adaptation outside the laboratory. To this end, we tested the extent to which a pair of motorized shoes could induce similar locomotor adaptation to split-belt walking, which is a well-established sensorimotor adaptation paradigm in locomotion. We specifically compared the adaptation effects (i.e. after-effects) between two groups of young, healthy participants walking with the legs moving at different speeds by either a split-belt treadmill or a pair of motorized shoes. The speeds at which the legs moved in the split-belt group was set by the belt speed under each foot, whereas in the motorized shoes group were set by the combined effect of the actuated shoes and the belts' moving at the same speed. We found that the adaptation of joint motions and measures of spatial and temporal asymmetry, which are commonly used to quantify sensorimotor adaptation in locomotion, were indistinguishable between groups. We only found small differences in the joint angle kinematics during baseline walking between the groups – potentially due to the weight and height of the motorized shoes. Our results indicate that robust sensorimotor adaptation in walking can be induced with a paired of motorized shoes, opening the exciting possibility to study sensorimotor adaptation during more realistic situations outside the laboratory.

Keywords: locomotion, motor learning, rehabilitation robotics, real-world, portable device

OPEN ACCESS

Edited by:
Thomas Probst,
Danube University Krems, Austria

Reviewed by:
Emilia Biffi,
IRCCS Eugenio Medea, Italy
Natalia Sanchez,
University of Southern California, United States
Ryan Thomas Roemmich,
Kennedy Krieger Institute, United States

****Correspondence:***
Gelsy Torres-Oviedo
gelsyto@pitt.edu

Specialty section:
This article was submitted to Neural Technology, a section of the journal Frontiers in Neuroscience

Received: *04 October 2019*
Accepted: *17 February 2020*
Published: *06 March 2020*

Citation:
Aucie Y, Zhang X, Sargent R and Torres-Oviedo G (2020) Motorized Shoes Induce Robust Sensorimotor Adaptation in Walking. Front. Neurosci. 14:174. doi: 10.3389/fnins.2020.00174

INTRODUCTION

The motor system has the flexibility to update motor plans according to systematic changes in the environment or the body. This human ability is studied in the laboratory through sensorimotor adaptation paradigms imposing sustained and predictable motor demands specific to the task at hand, such as unusual visuomotor rotations (e.g. Krakauer et al., 2000) or constant forces during walking (Savin et al., 2010) or reaching (Shadmehr and Mussa-ivaldi, 1994). For example, split-belt walking is a well-established paradigm in which participants update spatiotemporal gait features in response to a persistent speed difference between their legs (Dietz et al., 1994; Reisman et al., 2005; Malone et al., 2012). Important motor adaptation principles have been learned from these sensorimotor adaptation paradigms, such as the computations underlying motor adaptation (Thoroughman and Shadmehr, 2000; Haruno et al., 2001; Smith et al., 2006) or

neural structures involved in this process (Deuschl et al., 1996; Smith and Shadmehr, 2005; Morton and Bastian, 2006). However, there are inherent limitations to laboratory-based studies that bring into question the extent to which principles governing motor adaptation apply to motor learning in the real-world.

Specifically, there are task-constraints in laboratory-based studies that limit our ability to investigate factors that are critical for motor learning outside the laboratory setting. For example, laboratory-based protocols challenge the study of extended practice, which is a critical aspect of motor learning (Ericsson and Pool, 2016; Haith and Krakauer, 2018). There are several efforts to investigate the effect of extended practice on motor behavior by bringing participants to the laboratory multiple times (Day et al., 2018; Leech et al., 2018; Hardwick et al., 2019). This research effort would be facilitated if individuals could practice outside the laboratory setting. Further, we constrain movements by for example making people walk at a constant speed (Dietz et al., 1994), or repeatedly reach to a certain direction (Krakauer et al., 2000). This is done to simplify the control variables affecting the studied behavior, and at the extreme, this could yield to the study of unnatural behaviors, whose underlying mechanisms might not apply to realistic situations. A byproduct from task-constraints is the context-specificity of motor patterns learned in the laboratory that is movements adapted with the device only partially carry over to movements without the training device (Kluzik et al., 2008; Torres-Oviedo and Bastian, 2010). This is detrimental not only because it limits our capacity for studying the generalization of motor learning across distinct situations, but also because it limits the possibility for using laboratory-based tasks for motor rehabilitation. Notably, it is well-accepted that the generalization of motor patterns from trained to untrained situations can be improved when the two contexts are more similar to one another (Tulving and Thomson, 1973; Spear, 1978; Bouton et al., 1999). Thus, there could be more generalization of laboratory-based knowledge to realistic situations when the tasks studied in the laboratory are more similar to those observed under naturalistic conditions.

Portable devices may offer the possibility to overcome the limitations of laboratory-based studies of motor learning. For example, portable devices allow us to investigate motor learning in real-life settings, such as studies of surgical training with the same tools that are used at the clinic (Sharon et al., 2017). In addition, the portability of training devices also enables the study of extended practice since individuals are not constrained to only train in the laboratory setting (Hardwick et al., 2019). Further, portable devices might allow for more complex movements that involve the whole body (Haar et al., 2019), which might lead to greater motor variability – a key factor for motor learning (Kelly and Sober, 2014; Wu et al., 2014; Therrien et al., 2016). In the context of locomotion there have been efforts to develop portable devices to study motor adaptation (Handzic et al., 2011; Handzic and Reed, 2013; Lahiff et al., 2016). However, the previous devices were passive, lacking the control over the speed difference between the feet. In addition, gait adjustments induced by these devices are not as robust as the ones observed with laboratory-based apparatus such as split-belt treadmills. Thus, we asked if a pair of motorized shoes could induce locomotor adaptation comparable to split-belt walking, which is a well-established sensorimotor adaptation paradigm in locomotion.

We specifically hypothesized that introducing a speed difference between participant's feet with the motorized shoes would result in adaptation of spatiotemporal gait patterns similar to split-belt walking. To test this hypothesis, we compared locomotor adaptation at comparable speed differences imposed by either a pair of motorized shoes or a split-belt treadmill. If the locomotor adaptation with the motorized shoes is similar to the one observed during split-belt walking paradigm, participants could start wearing these shoes outside the laboratory, which would offer the exciting possibility to study locomotor learning under more realistic situations.

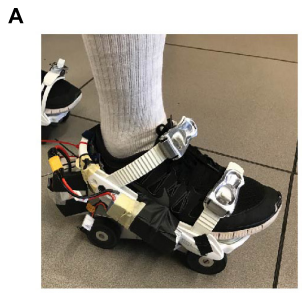
MATERIALS AND METHODS

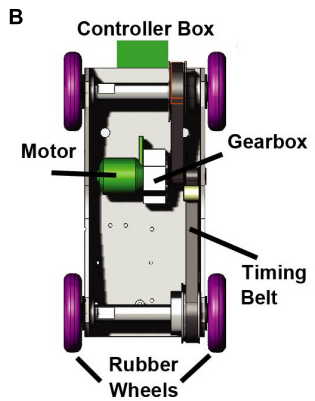
Participants

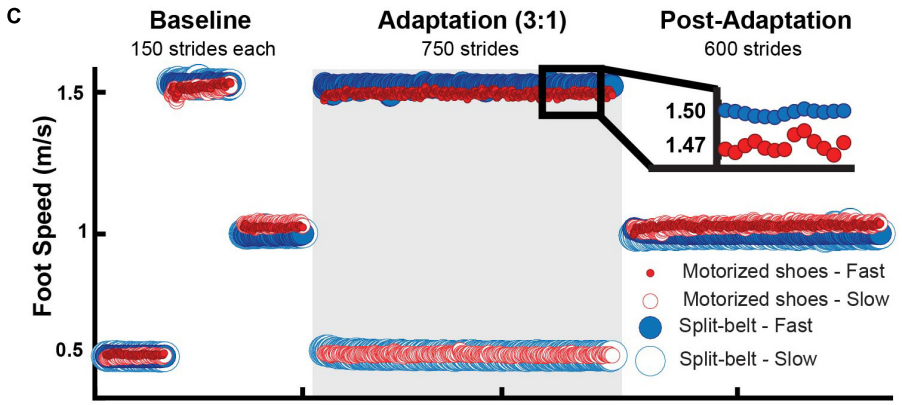
We investigated if a pair of motorized shoes could induce locomotor adaptation and after-effects similar to a split-belt treadmill. To this end, a group of 18 young, healthy, and naïve adults were adapted using either (1) the motorized shoes that imposed speed differences between the feet using actuated wheels under the shoe (motorized shoes group: n = 9; three females: 26.6 ± 3.5 years) or (2) a split-belt treadmill, in which belts moved at different speeds (split-belt group: n = 9; four females: 25.3 ± 4.3 years). The Institutional Review Board at the University of Pittsburgh approved our experimental protocol and all participants gave their written informed consent before being tested.

Set Up

The motorized shoes group walked on the treadmill while wearing the custom made motorized shoes (Nimbus Robotics, Pittsburgh, PA, United States) as shown in **Figure 1A** on top of their normal walking shoes. In brief, the shoes were designed to move an individual (weighing <100 kg) up to 1 m/s in the forward direction only (i.e. wheels cannot be actuated to rotate backward). Each of the motorized shoe (∼1.7 kg) consisted of a motor, a controller box, a gearbox, two toothed timing belts, and four rubber wheels (**Figure 1B**). Lithium batteries (3V) were used to power the motor, which rotated the timing belts via a gearbox connecting the two. The feet moved at different speeds with the motorized shoes by locking the wheels of one foot and actuating the wheels of the other foot, such that the combined effect of the treadmill's belt moving the foot backward and the motorized shoe moving the foot forward would result in the desired foot speed of 0.5 m/s (**Figure 1B**). To this end, the timing belts and rubber wheels were coupled to rotate the wheels such that they locked the non-actuated shoe during stance (∼0 m/s) and moved the actuated shoe forward at a linear speed of 1 m/s. The controller boxes received signals through a remote controller operated by the experimenter. All software for the controller boxes and the remote controller were written in Python. Details on the control software are published in Zhang (2017) and a detailed description of the motorized shoes will be revealed in the full utility patent (currently in provisional status). The split-belt







Speed Implementation

Split-belt	*Motorized Shoes*
Net Speed = Treadmill Speed	*Net Speed = Shoe Speed + Treadmill Speed*
Slow = 0.5 m/s = 0.5 m/s	***Slow = 0.5 m/s*** = 1 m/s (forward) + 1.5 m/s (backward)
Fast = 1.5 m/s = 1.5 m/s	***Fast = 1.5 m/s*** = 0 m/s (locked) + 1.5 m/s (backward)
Medium = 1 m/s = 1 m/s	***Medium = 1 m/s*** = 0 m/s (locked) + 1 m/s (backward)

FIGURE 1 | (A) A motorized shoe involving proprietary technology was used to induce adaptation in the motorized shoes group. **(B)** Schematic of the motorized shoe. This consists of a motor, a controller box, a gearbox, two toothed timing belts, and four rubber wheels. **(C)** Mean time courses for foot speed across participants for the motorized shoes and the split-belt groups. The white background indicates experimental epochs of "tied" walking when both feet moved at the same speed, whereas the gray background indicates the epoch of "split" walking when the dominant leg moved three times faster than the non-dominant leg. The table summarizes the procedure used to set the slow, fast, and medium speeds for each foot. The same procedure was used in all epochs. It is worth pointing out that the treadmill always moved at 1.5 m/s during adaptation in the motorized shoes group. The speed difference between feet was achieved by locking the wheels on the fast side and moving the slow foot forward at 1 m/s to obtain a net speed of 0.5 m/s on the slow side. Of note, the foot's speed on the fast side was slightly slower on the motorized shoes than the split-belt group.

group did not wear the motorized shoes and walked with their regular shoes on an instrumented split-belt treadmill (Bertec, Columbus, OH, United States).

General Paradigm

All participants adapted following a conventional sensorimotor adaptation paradigm that consisted of three walking conditions: baseline, adaptation, and post-adaptation (**Figure 1C**, Top). During these periods, participants' feet moved at one of three possible speeds: slow (0.5 m/s), medium (1 m/s), or fast (1.5 m/s). The implementation of these speeds is displayed in **Figure 1C**. Participants in the motorized shoes group wore these shoes throughout the experimental protocol, whereas participants in the split-belt group wore regular sneakers. Thus, the net foot speed in the motorized shoes group was the sum of the treadmill's speed (moving the foot backward) and the shoe's speed (moving the foot forward), whereas the foot speed in the split-belt group was only dependent on the treadmill's speed (**Figure 1C**, Bottom). For example, in the motorized shoes group the slow foot speed (0.5 m/s) resulted from the combined effect of the treadmill moving the foot at 1.5 m/s (backward) and the motorized shoe moving the foot at 1 m/s (forward) (i.e. 1.5 −1 = 0.5 m/s). The motorized shoes were OFF and wheels were locked (0 m/s) at the fast and medium speeds; thus, the foot's net seed at those velocities was only determined by the treadmill's speed. This was done to maximize the experiment's duration for a given battery life. Our approach also enabled us to implement the same feet speed's in both groups while participants in the motorized shoes group walked on a regular treadmill (i.e. both belts moving at the same speeds).

A baseline period was collected during which both feet moved at either slow, fast, or medium speeds for 150 strides each (**Figure 1C**, Top). The baseline behavior during the slow and fast speeds served as a reference for the adaptation condition when the feet moved at different speeds, whereas the medium speed served as a reference for the post-adaptation period when the two feet move at the same medium speed. Moreover, the baseline speed was matched not only in the speed at which the feet moved, but also on how this speed was implemented. For example, in the motorized shoes group, the shoe was actuated in the slow side (net speed = 0.5 m/s) and it was OFF (wheels locked) in the fast side (net speed = 1.5 m/s) during the adaptation period. Accordingly, both motorized shoes were either actuated or OFF in the slow and fast baselines, respectively. The adaptation period

lasted 750 strides (approx. 15 min) and the dominant leg (self-reported leg to kick a ball) walked fast. The speed difference and period duration was selected to match other split-belt walking studies showing robust gait adaptation (Sombric et al., 2019). Following the adaptation block, all participants experienced a post-adaptation period of 600 strides during which both feet moved at 1 m/s, which was the average speed of the fast and slow feet. The purpose of this phase was to measure the adaptation effects and its washout when the speed perturbation induced by different devices was removed.

Data Collection

All participants walked on an instrumented treadmill either with or without the motorized shoes, while kinematic and kinetic data were collected to characterize participants' gait. Kinematic data were collected at 100 Hz with a passive motion capture system (Vicon Motion Systems, Oxford, United Kingdom) and kinetic data were collected at 1000 Hz using force plates embedded in the treadmill. Gaps in raw kinematic data due to marker occlusion were filled by visual inspection of each participant in Vicon Nexus software. Positions from the toe (5th metatarsal), ankle (lateral malleolus), knee (lateral epicondyles), and the hip (greater trochanter) were collected bilaterally (**Figure 2B**). Heel-strikes (i.e. foot landing) and toe-offs (i.e. foot lift off) were identified using the ground reaction force (Fz) perpendicular to the walking surface. More specifically, heel-strike was defined as the instance when Fz $>$ 30 N and toe-off as the instance when Fz $<$ 30 N. We used this force threshold to have equivalent event detection (i.e. heel strike and toe off) on the treadmill for both groups since each of the motorized shoe weighted 17 N ($\sim$1.7 kg in mass).

Data Analysis

We compared the gait pattern between the motorized shoes and split-belt groups in terms of spatial and temporal symmetry measures that are known to adapt on the split-belt treadmill (**Figure 2A**; Finley et al., 2015). Specifically, we used step length asymmetry as a robust measure of adaptation. Step length asymmetry was defined as the difference between step lengths (i.e. distance between ankles) with the slow leg vs. the fast leg (Eq. 1). A zero value of step length asymmetry indicated that both step lengths were equal and a positive value indicated that the step length of the fast (dominant) leg was longer than the slow (non-dominant) leg. Step length asymmetry was further decomposed into StepPosition, StepTime, and StepVelocity because these parameters have been shown to be adapted differently during split-belt walking (Finley et al., 2015). The StepPosition quantified the difference in positions of the leading leg (i.e. leg in front of the body) between two consecutive steps (Eq. 2). The StepTime quantified the difference in the duration of each of these steps (Eq. 3). Lastly, the StepVelocity

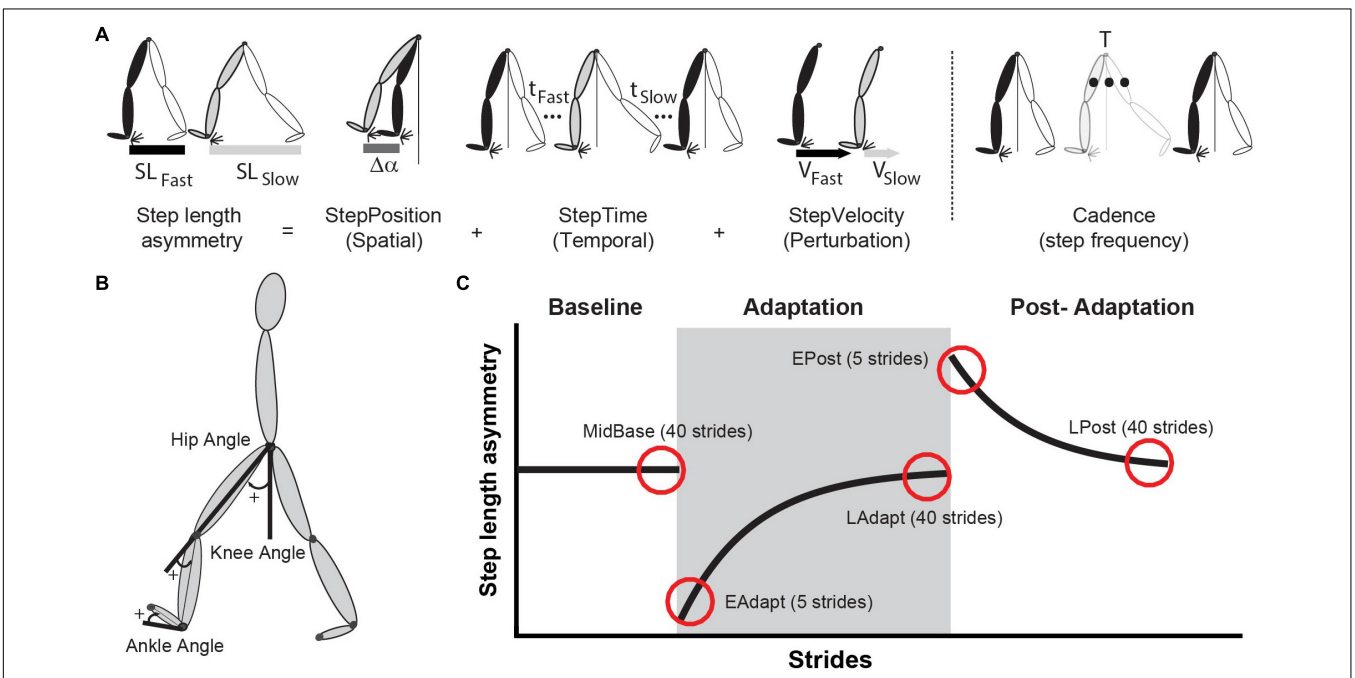


FIGURE 2 | (A) This schematic illustrates step length asymmetry and its decomposition into StepPosition, StepTime, and StepVelocity. Step length asymmetry is quantified as the difference between fast and slow step lengths, normalized by stride length. The equation and decomposition are explained in detail in the section "Materials and Methods" of this manuscript. In brief, (StepPosition) differences between the fast (black leg) and the slow (gray leg) leading leg's positions contribute to step length asymmetry. Similarly, differences in the trailing leg's positions (white legs) also contribute to step length asymmetry. The trailing leg's position depends on step time and step velocity. Consequently, differences in step times (t_{fast} and t_{slow}) or step velocity (V_{fast} and V_{slow}) leads to step length asymmetry. We also show a schematic of Cadence, which is computed as the inverse of the gait period (T). **(B)** Illustration of reflective marker positions and joint angle conventions. **(C)** Epochs of interest are illustrated by the red circles placed over a schematic of step length asymmetry. Shaded gray area represents the adaptation period when the feet move at different speeds ("split" walking), whereas white areas represent when the feet move at the same speed.

quantified the difference in the velocities of each foot with respect to the body for these two steps (Eq. 4). Since participants take steps with different sizes, we normalized the differences in step length, StepPosition, StepTime, and StepVelocity by their stride length, quantified as the sum of two step lengths. This allowed us to avoid inter-subject variability. For visualization purposes, these parameters were smoothed with a five-step running average.

$$Step\ legnth\ asymmetry = \frac{Fast\ Step\ Length - Slow\ Step\ Length}{SL} \quad (1)$$

$$StepPosition = \frac{(\Delta\alpha_{fast} - \Delta\alpha_{slow})}{SL} \quad (2)$$

$$StepTime = \frac{\frac{v_{slow}+v_{fast}}{2}(t_{slow} - t_{fast})}{SL} \quad (3)$$

$$StepVelocity = \frac{\frac{t_{slow}+t_{fast}}{2}(v_{slow} - v_{fast})}{SL} \quad (4)$$

In these equations, $\Delta\alpha$ indicates the difference between each foot's position (i.e. ankle marker) and the body (i.e. mean position of the two hip markers) at ipsilateral heel strike (**Figure 2A**); In addition, t indicates the step time defined as the duration between the heel-strike of ipsilateral leg to the contralateral leg; and v indicates the step velocity quantified as the relative velocity of the foot with respect to the body. When walking on the treadmill, v_{slow} and v_{fast} approximated the speeds of the slow and fast belt, respectively. Therefore, StepVelocity was mostly reflective of belt speed difference, rather than participants' behavior. Finally, note that all measures were normalized by each participant's stride length (SL, sum of both step lengths) to account for inter-subject differences in step sizes.

We also computed joint angles and cadence to determine the impact of the motorized shoes on each foot's motion and step frequency. Ankle, knee, and hip angles were computed on the sagittal plane (2D) to directly contrast our results to previous reports of joint angles during split-walking (Reisman et al., 2005). Joint angles were calculated such that flexion/dorsiflexion was positive and extension/plantarflexion was negative (**Figure 2B**). We also defined all angles to have value of 0° at the neutral standing position (i.e. full extension for knee and hip and approximately 90° angle between shank and foot for the ankle). More specifically, ankle angles were calculated as the angle between the foot (ankle marker to toe marker vector) and the shank (ankle marker to knee marker vector) subtracted from each participant's neutral position (i.e. mean and standard deviation: $88.4 \pm 3.7°$ for the group wearing the motorized shoes and $91.2 \pm 0.95°$ for the split-belt group). Knee angles were calculated as the angle between the shank and the thigh (knee marker to hip marker vector) subtracted from 180°. Lastly, we computed the hip angles as the angle between the thigh and the vertical unit vector. Angle data was time-aligned and binned to compute mean angle values over six intervals of interest during the gait cycle. This was done to focus on changes in angles within the gait cycle, rather than on changes due to differences in cycle duration across the distinct walking conditions (Dietz et al., 1994; Reisman et al., 2005). More specifically, we computed averaged angle values over six phases of interest (Perry, 2010): double support (DS1 and DS2), single stance (SS1 and SS2), and the swing phases (SW1 and SW2). Double support during early stance (DS1) was defined as the period from heel strike to contralateral toe off. Single stance (from contralateral toe-off to contralateral heel strike) was divided into two equal phases (SS1 and SS2). Double support during late stance (DS2) was defined as the interval from contralateral heel strike to ipsilateral toe off. Finally, the swing phase (from ipsilateral toe-off to ipsilateral heel-strike) was divided into two equal phases (SW1 and SW2). Joint angles were assessed in eight participants per group since the remaining two participants (one per group) was missing essential marker data. Lastly, we computed cadence (i.e. number of strides per second) to determine if this gait feature was altered by wearing the motorized shoes.

Outcome Measures

Each gait parameter was analyzed during four experimental epochs of interest (early adaptation, late adaptation, early post-adaptation, and late post-adaptation) to compare the adaptation and after-effects between the motorized shoes and the split-belt treadmill groups. We computed the averaged value of each parameter over these epochs as follows. First, we removed the five strides at the beginning and at the end of each trial to eliminate effects of holding on to the handrail when starting and stopping the treadmill. This was done to characterize people's movement when no individuals were holding on to the safety rail. Then, we computed the average value for each epoch as follows: early adaptation (EAdapt, average of five strides: 6^{th}–10^{th} stride), late adaptation (LAdapt, average of 40 strides: 706^{th}–745^{th} stride), early post-adaptation (EPost, average of five strides: 6^{th}–10^{th} stride), and late post-adaptation (LPost, average of 40 strides: 546^{th}–595^{th} stride) (**Figure 2C**). All of the parameters were corrected by any baseline biases (MidBase, average of 40 strides: 106^{th}–145^{th} stride). EAdapt gave us information about the induced perturbation by the "split" condition, while the LAdapt provided information regarding the steady-state behavior at the end of the adaptation trial. The behavior during EPost was quantified to assess how much participants adapted to the new walking pattern (e.g. after-effects). Finally, we assessed LPost behavior to ensure that participants returned to their baseline walking behavior (e.g. washout). Moreover, we used joint angle measures to determine the effect of the motorized shoes on the overall gait pattern. This analysis was intended to determine if participants were actually walking with the motorized shoes (i.e. not dragging their feet or sliding their feet). To this end, we computed the averaged value over the last 40 strides (after removing the very last five strides, as in the other kinematic parameters) for each one of the four experimental epochs of interest (i.e. SBase, FBase, MidBase, and LAdapt).

Statistical Analysis

We performed one-sample Kolmogorov–Smirnov tests to determine if each parameter (i.e. Step length asymmetry, Step lengths, StepPosition, StepTime, StepVelocity, and Cadence) was

normally distributed in every epoch of interest (i.e. EAdapt, LAdapt, EPost, and LPost). We found that all parameters were normally distributed, thus we ran separate two-way repeated measures ANOVAs to test the effects of epochs and groups (i.e. motorized shoes vs. split-belt) on each of our gait parameters. Statistical analysis was done with unbiased data (i.e. MidBase was subtracted from all the epochs) to focus on changes that occurred beyond those due to distinct group biases. In case of significant main or interaction effects, we used Fisher's *post hoc* testing to determine whether values were different between groups. We chose this *post hoc* testing to be more sensitive to potential group differences. Lastly, we performed a one-sided one sample *t*-test to determine whether early post-adaptation values were different from zero. This was done to determine if after-effects were significant in each group. Comparisons between post-adaptations values across groups were only done when we found significant interactions between group and epoch.

Two separate multiple linear regressions were performed to determine if the individual variation in two independent variables: (1) StepPosition and (2) StepTime in late adaptation could be predicted by two regression coefficients and their interaction: group (categorical factor), StepVelocity (continuous variable), and group#StepVelocity (interaction). We also performed two separate multiple linear regressions to determine if the individual variation in after-effects in StepPosition and StepTime (two independent variables) were predicted by group or each respective steady state (StepPosition LAdapt or StepTime LAdapt). This was done because we observed speed differences between the groups (**Figure 1C**, Top) that could impact the extent of adaptation and after-effects on spatial and temporal measures.

Joint angles were compared across groups using unpaired *t*-test for each of the gait phases. We reasoned this was an appropriate statistical test to compare the behavior across groups given that joint angles are highly temporally correlated within the gait cycle and spatially correlated across segments. We subsequently corrected the significance threshold for each epoch using a Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995), setting a false discovery rate of 5% (FDR correction). The reason for choosing this correction was due to higher number of comparisons that we made.

A significance level of $\alpha = 0.05$ was used for all statistical tests. Stata (StataCorp., Collage Station, TX, United States) was used to perform the ANOVAs, whereas MATLAB (The MathWorks, Inc., Natick, MA, United States) was used for all other analyses.

RESULTS

Motorized Shoes Can Induce Robust Sensorimotor Adaptation of Locomotion

Our results show that the motorized shoes were able to induce similar adaptation of step length asymmetry compared to the split-belt treadmill. Specifically, there were no significant group ($F_{(1,48)} = 0.21$, $p = 0.65$) or group by epoch interaction effects ($F_{(3,48)} = 1.26$, $p = 0.29$) on the adaptation of step length asymmetry, indicating that this parameter was similarly modulated throughout the experiment between the motorized shoes and split-belt groups (**Figure 3A**). We observed a significant main effect of epoch ($F_{(3,48)} = 94.91$, $p < 0.001$) in step length asymmetry and found that both groups had significant after-effects (motorized shoes: $p < 0.001$; split-belt: $p < 0.001$; **Figure 3A**). While modulation of step length asymmetry was indistinguishable between groups, we observed small differences in the adaptation of the fast leg's step length. Specifically, we found a group by epoch interaction effect in the fast step length ($F_{(3,48)} = 3.18$, $p = 0.032$; **Figure 3B**) driven by between-group differences during the early adaptation phase ($p = 0.012$). While significant, this between-group difference might not be meaningful given that the values that observed in both groups fall within the range of those previously reported (Sombric et al., 2019). Moreover, after-effects in this parameter were significant in the motorized shoes group ($p = 0.013$), but not in the split-belt group ($p = 0.15$). In contrast, the adaptation of the slow leg's step length was similar across groups throughout the experiment (group: $F_{(1,48)} = 0.63$, $p = 0.44$; group by epoch interaction: $F_{(3,48)} = 0.69$, $p = 0.49$; **Figure 3C**). We only found a significant epoch effect on slow step length ($F_{(3,48)} = 70.47$, $p < 0.001$) and substantial after-effects in both groups (motorized shoes: $p < 0.001$; split-belt: $p < 0.001$). In summary, fast leg's step length exhibited small differences between the motorized shoes and split-belt groups that did not impact the adaptation of step length asymmetry, which was indistinguishable between these groups.

Smaller Speed Difference With the Motorized Shoes Reduced the Adaptation of StepPosition

We observed between-group differences in the adaptation of StepPosition (quantifying spatial asymmetry), but not StepTime (quantifying temporal asymmetry). This was indicated by the significant group by epoch interaction found in StepPosition ($F_{(3,48)} = 3.47$, $p = 0.023$), but not in StepTime ($F_{(3,48)} = 2.39$, $p = 0.09$) (**Figure 4**). *Post hoc* analyses indicated that these differences in StepPosition were driven by distinct early and late adaptation values of this parameter in the motorized shoes group compared to the split-belt group (early adaptation: $p = 0.031$; late adaptation: $p = 0.036$). Yet, after-effects in StepPosition were significant in both groups (motorized shoes: $p < 0.001$; split-belt: $p < 0.001$) and after-effects in StepTime were only significant in the motorized shoes group (motorized shoes: $p = 0.017$; split-belt: $p = 0.087$) Interestingly, we also found a group effect ($F_{(1,48)} = 6.58$, $p = 0.021$) on StepVelocity and a group by epoch interaction trending effect ($F_{(1,48)} = 2.78$, $p = 0.051$) (**Figure 4C**). In particular, the StepVelocity was smaller in the group with motorized shoes than in the split-belt group during late adaptation ($p = 0.001$), which we thought could impact the motor adaptation of the motorized shoes group. Thus, we performed multiple linear regression analysis on the late adaptation epoch with either StepTime or StepPosition as the dependent variable and StepVelocity as the predictor. StepVelocity was indeed related to StepTime ($R^2 = 0.59$; $p = 0.005$; StepTime $= -1.19 * \text{StepVelocity} - 0.32$) and

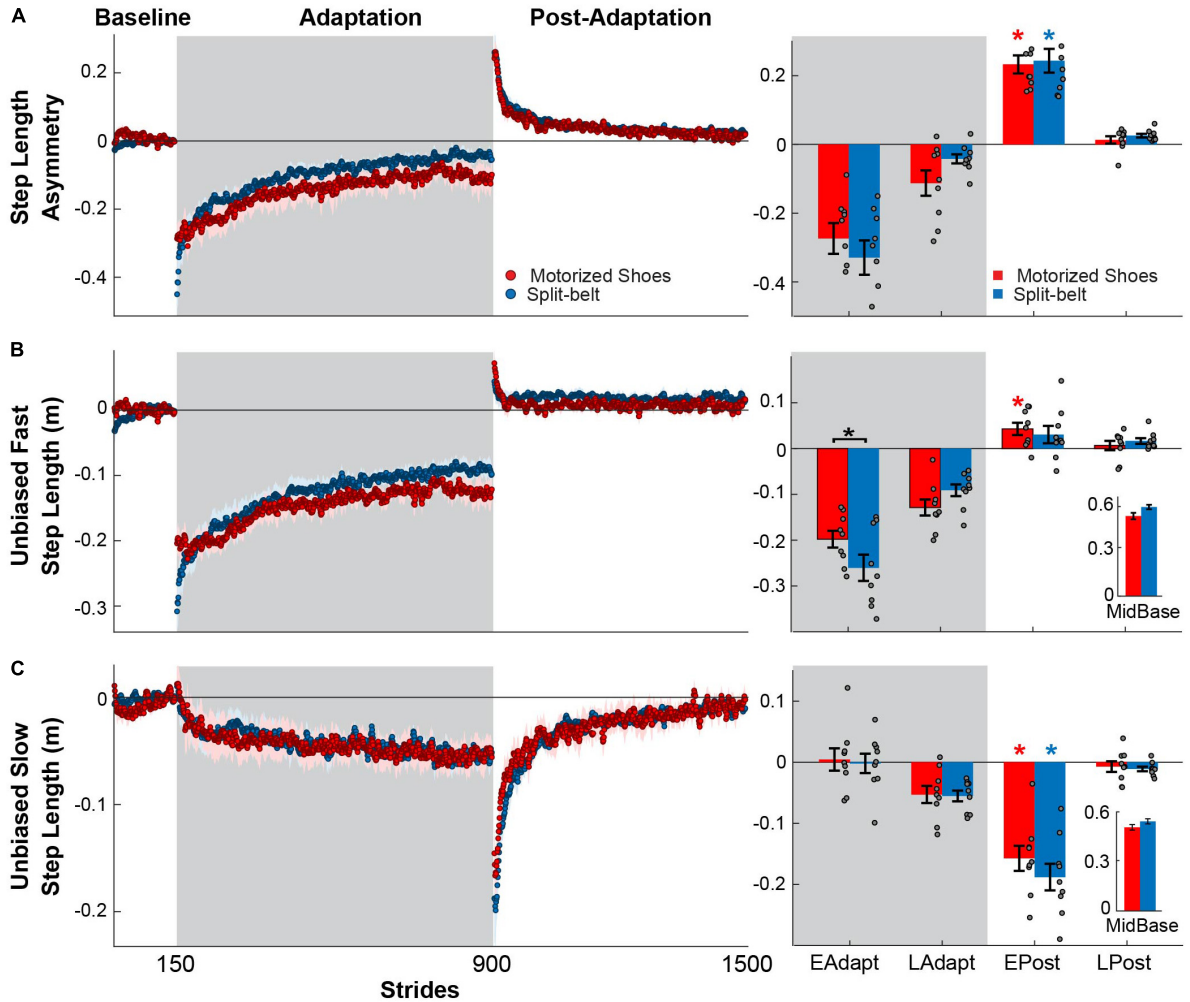


FIGURE 3 | Modulation of step length asymmetry and step lengths. (**A–C**, Left panel) Time courses for step length asymmetry and individual step lengths during medium baseline, adaptation, and post-adaptation. Shaded gray area represents the adaptation period when the feet move at different speeds ("split" walking), whereas white areas represent when the feet move at the same speed. Colored dots represent the group average of five consecutive strides and colored shaded regions indicate the standard error for each group (motorized shoes: red; split-belt: blue). (**A–C**, Right panel) Bar plots indicate the mean ± standard errors for step length asymmetry and step lengths for each group and epoch of interest. Note that the reported step lengths are unbiased. This was done by subtracting the averaged step length values during baseline at medium speed in each participant. Significant differences for *post hoc* tests were indicated as follows. Black asterisks over the bracket above each epoch represent statistical significant differences between the motorized shoes and the split-belt groups ($p < 0.05$). Colored asterisks over the bars indicate significant after-effects (i.e. early post-adaptation is significantly different from baseline; $p < 0.05$) for each of the groups (motorized shoes: red; split-belt: blue). The small bar plots on the right indicate the mean ± standard errors for the step lengths for each group during medium baseline.

StepPosition (R^2 = 0.55; p = 0.009; StepPosition = −0.82 * StepVelocity − 0.15). However, individual StepVelocity values were only a predictor of StepTime values [Group: p_group = 0.19, regression coefficient = 0.44, 95% CI = (−0.25, 1.13); StepVelocity: p_velocity = 0.001, regression coefficient = −1.99, 95% CI = (−3.08, −0.91); Interaction: p_group#velocity = 0.16, regression coefficient = 1.14, 95% CI = (−0.49, 2.78)], whereas the relation between StepVelocity and StepPosition was driven by a group effect (Group: p_group = 0.047, regression coefficient = 0.71, 95% CI = (0.0092, 1.4); StepVelocity: p_velocity = 0.068, regression coefficient = −1.01, 95% CI = (−2.1, 0.086); Interaction: p_group#velocity = 0.069, regression coefficient = 1.5, 95% CI = (−0.13, 3.16)] (**Figure 4D**). We also found that the inter-subject variability in steady-state values was not associated to individual after-effects in neither StepPosition (R^2 = 0.23; p = 0.29), nor StepTime (R^2 = 0.12; p = 0.59) (**Figure 4E**). To sum up, the reduced speed difference in the motorized shoes group limited the adaptation of StepPosition, but we still observed group after-effects with the motorized shoes in the spatial and temporal domains.

Similar Cadence Is Observed Between the Groups Throughout the Experiment

We found that the motorized shoes did not alter the modulation of cadence throughout the experiment compared to split-belt

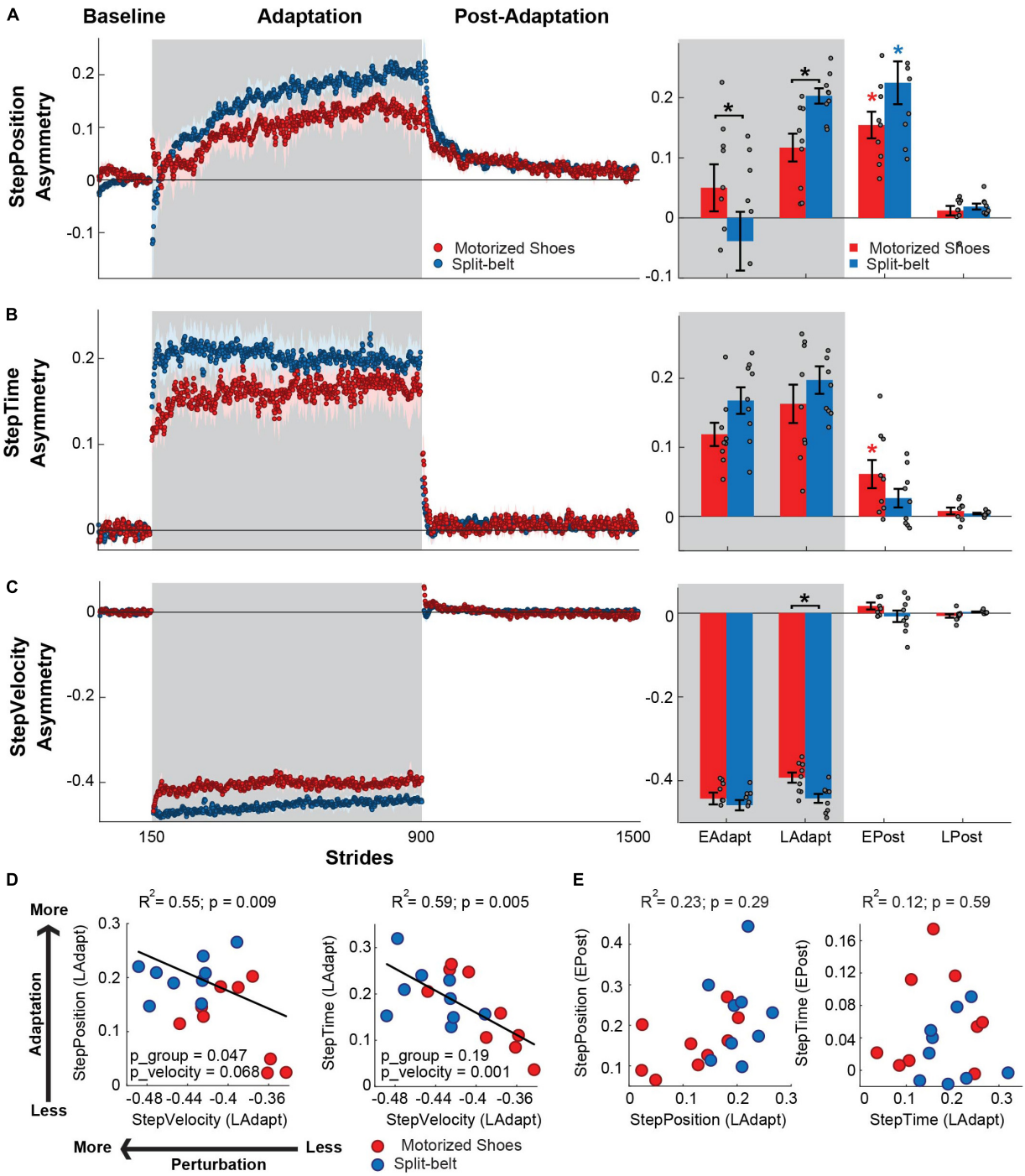


FIGURE 4 | Adaptation of spatiotemporal components of step length asymmetry. (**A–C**, Left panel) Time courses for StepPosition, StepTime, and StepVelocity before, during, and after adaptation. Shaded gray area represents the adaptation period when the feet move at different speeds ("split" walking), whereas white areas represent when the feet move at the same speed. Colored dots represent the group average of five consecutive strides and colored shaded regions indicate the standard error for each group (motorized shoes: red; split-belt: blue). (**A–C**, Right panel) The bar plots indicate the mean ± standard errors for StepPosition, StepTime, and StepVelocity for each group and epoch of interest. Gray dots represent individual participants. Note that the values were corrected for baseline biases. Significant differences for *post hoc* tests were indicated as follows. Black asterisks over the bracket above each epoch represent statistical significant differences between the motorized shoes and the split-belt groups ($p < 0.05$). Colored asterisks over the bars indicate significant after-effects (i.e. early post-adaptation is significantly different from baseline; $p < 0.05$) for each of the groups (motorized shoes: red; split-belt: blue). **(D)** Scatter plots illustrate the association between the StepVelocity at steady state and either the StepPosition or StepTime at steady-state during adaptation (i.e. LAdapt). We present the *p*-values for the multiple regression model (p), for the continuous variable (StepVelocity, p_velocity) and for the categorical variable (group, p_group). **(E)** Scatter plots illustrate the association between the LAdapt and EPost for StepPosition and StepTime. No significant relations were observed for neither StepPosition nor StepTime.

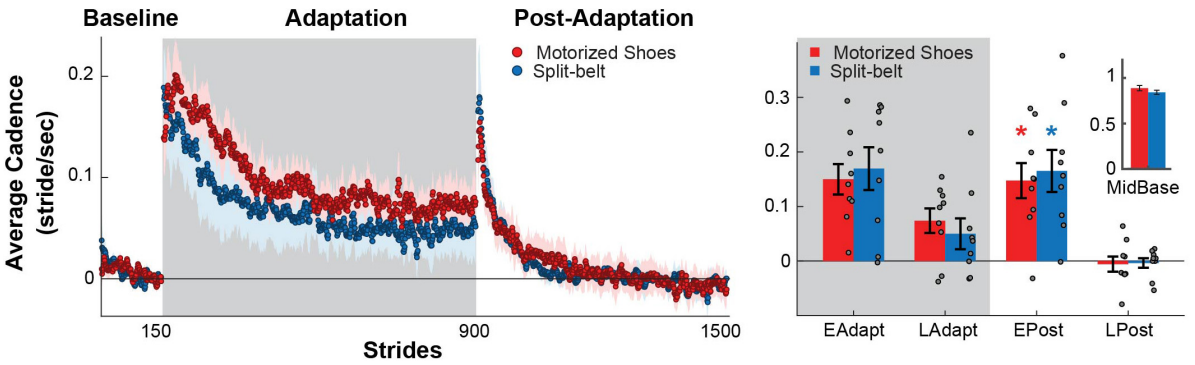


FIGURE 5 | Modulation of cadence. **(Left)** Time courses during medium baseline, adaptation, and post-adaptation for the average cadence is shown for each group. Shaded gray area represents the adaptation period when the feet move at different speeds ("split" walking), whereas white areas represent when the feet move at the same speed. Colored dots represent the group average of five consecutive strides and colored shaded regions indicate the standard error for each group (motorized shoes: red; split-belt: blue). **(Right)** Bar plots indicate the mean ± standard errors for cadence for each group and epoch of interest. Note that the values were corrected for baseline biases (i.e. MidBase). Colored asterisks over the bars indicate significant after-effects (i.e. early post-adaptation is significantly different from baseline; $p < 0.05$) for each of the groups (motorized shoes: red; split-belt: blue). The small bar plot on the right indicates the mean ± standard errors for the Cadence for each group during medium baseline.

walking (**Figure 5**, left). Specifically, there were no significant group ($F_{(1,48)} = 0.02$, $p = 0.88$) or group by epoch interaction effects on cadence ($F_{(3,48)} = 0.32$, $p = 0.81$), indicating that the adaptation and after-effects of cadence were similar between groups (**Figure 5**, right). We also found that both groups exhibited increased cadences during early post-adaptation compared to baseline (motorized shoes: $p = 0.002$; split-belt: $p = 0.003$). In sum, individual wearing the motorized shoes modulate cadence similarly to individuals in the split-belt group.

Effect of Wearing Motorized Shoes on Gait Kinematics

Overall, the gait pattern with and without the motorized shoes was similar. **Figure 6A** illustrates the joint angles over the gait cycle for the ankle, knee, and hip joints for the group wearing the motorized shoes (red) and the group wearing regular shoes (blue) during medium baseline walking. We found joint angles were the same between groups for most phases of the gait cycle, in which significance was determined with an FDR controlling procedure (18 comparisons, $p >$ Pthreshold, Pthreshold = 0.0055, see the section "Materials and Methods") (**Figure 6A**). There were only a few differences in specific phases of the gait cycle. Specifically, the motorized shoes group demonstrated reduced ankle dorsiflexion following ipsilateral heel strike and during late swing (double support DS1: $p = 0.004$, effect size = 3.3°; late swing SW2: $p = 0.004$, effect size = 4.1°). Moreover, the motorized shoes group exhibited reduced knee flexion compared to the split-belt group during early swing (SW1: $p = 0.004$, effect size = 7.8°), followed by slightly more knee extension in late swing (SW2: $p = 0.001$, effect size = 9.6°). Lastly, the motorized shoes group had larger hip flexion during stance of baseline walking ($p = 0.005$, effect size = 4.1°). While these between-group differences were significant, they should be interpreted consciously given the reliability of kinematic measurements. Namely, one can find significant changes in joint angles that are greater than 5° when measured across sessions within the same cohort of healthy, young participants (Wilken et al., 2012). Therefore, the differences that we find, ranging from 3.3° to 9.6°, might not be meaningful. In addition to baseline joint kinematics, we also compared late adaptation kinematics across groups (**Figure 6B**). Specifically, we contrasted the changes in joint angles during late adaptation relative to the speed-specific baseline for each of the six phases of the gait cycle. We found no differences between the groups (36 comparisons, $p >$ Pthreshold), suggesting that joint angles were modulated similarly in the split condition with the motorized shoes or the split-belt treadmill. Thus, our results demonstrated that walking with the motorized shoes had only minor effects on joint kinematics and did not alter the adaptation of individual joint angles during split walking.

DISCUSSION

Summary

We investigated if a pair of motorized shoes could induce split-like locomotor adaptation. We found that the adaptation effects induced by the motorized shoes moving at different speeds were as robust as those observed with a split-belt treadmill. Moreover, we found that the gait pattern was largely similar between walking with the motorized shoes or on the split-belt treadmill. Specifically, step length asymmetry, cadence, and step lengths were similar across groups during and after the split condition with either device. We only observed subtle differences in individual joint angles during the baseline condition with the motorized shoes compared to walking with regular shoes, which might be due to the greater height and weight of the motorized shoes. Taken together, our results suggest motorized shoes can induce robust sensorimotor adaptation in locomotion, opening the exciting possibility to study locomotor learning under more realistic situations outside the laboratory setting.

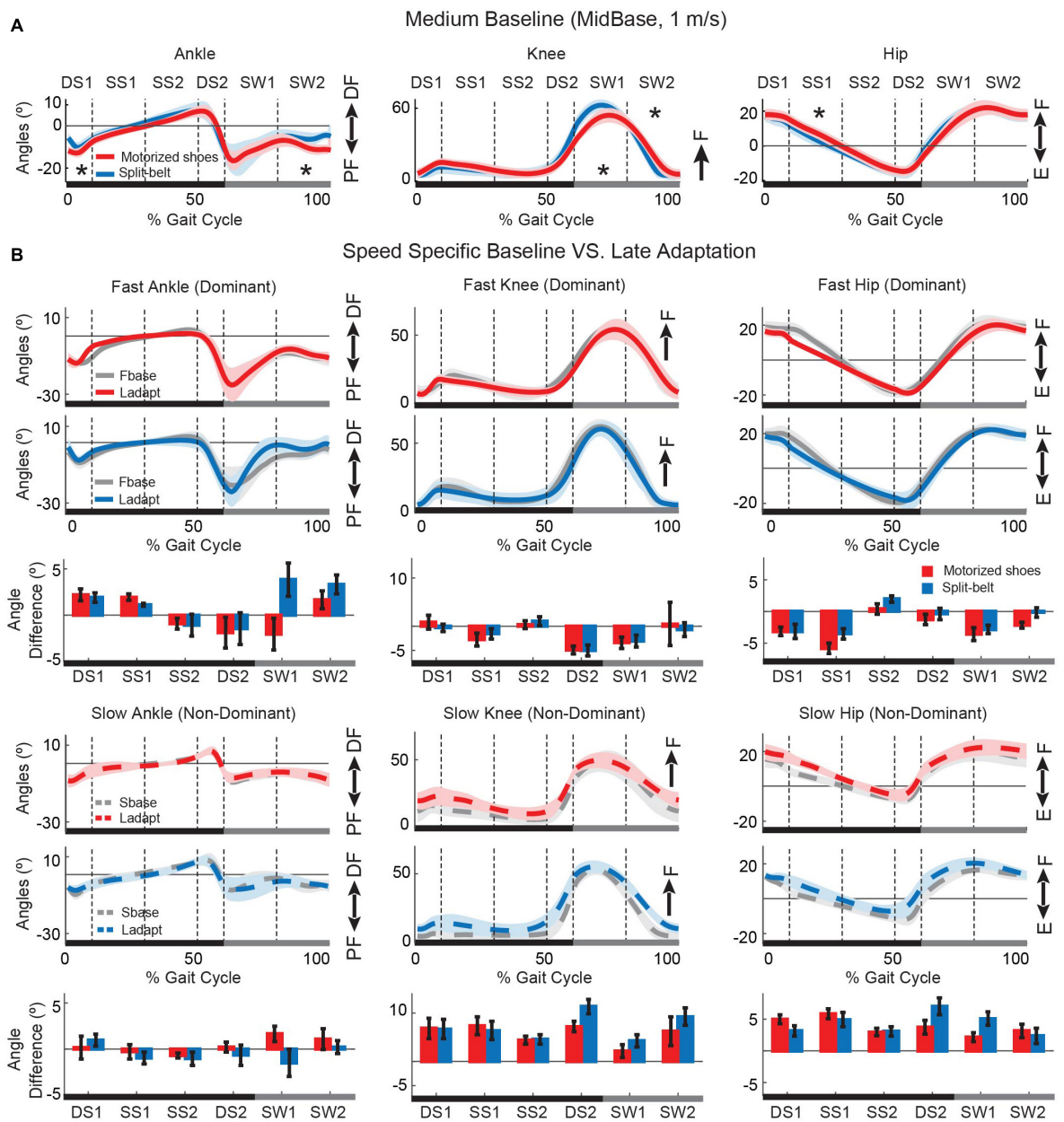


FIGURE 6 | Joint angles over the gait cycle during baseline and adaptation. **(A)** Baseline joint angles are shown for the group walking with regular sneakers (i.e. blue trace) and the group walking with the motorized shoes (i.e. red trace). Solid lines represent the group average and shaded areas represent standard errors. Asterisks indicate instances during the gait cycle when joint angles were significantly different across groups. The overall motion for all joints was similar across groups, but hip flexion, knee flexion, and ankle dorsiflexion were smaller when wearing the motorized shoes. **(B)** Speed specific baseline (gray) and steady-state angle trajectories during adaptation for the motorized shoes (red) and the split-belt (blue) groups. Solid lines represent the motion of the leg walking fast in the split condition (colored lines) and in the fast baseline (gray) condition. The dashed lines represent the motion of the leg walking slow in the split condition (colored lines) and in the slow baseline (gray) condition. The bars represent the change from the speed-specific baseline to late adaptation in joint angles during different phases of the gait cycle. DS, double support; SS, single stance; SW, swing; DF, dorsiflexion; PF, plantarflexion; F, flexion; E, extension.

Similar Walking and Adaptation With Split-Belt Treadmill and With Motorized Shoes

We demonstrated that the motorized shoes can induce locomotor adaptation largely similar to the adaptation induced with the split-belt treadmill. This was shown by the comparable adaptation across groups of gait parameters, such as step length asymmetry, and the same modulation of joint angles from baseline to adaptation for both groups. Namely, the initial and steady state values during the split condition for the split-belt group and motorized shoes group were consistent with values

previously reported for joint angle kinematics (Winter, 1987; Reisman et al., 2005) and asymmetries in step length (Malone and Bastian, 2010; Finley et al., 2015), step position (Sombric et al., 2017), and step time (Gonzalez-Rubio et al., 2019). We found between-group differences in the fast step length during early adaptation, such that participants with the motorized shoes placed the fast leg closer to the body. This distinct behavior might also be explained by the fact that the balance is perturbed in the beginning of the split condition (Buurke et al., 2018; Iturralde and Torres-Oviedo, 2019) and it might be further challenged when stepping with the motorized shoes by augmenting the center of mass' height, increasing even further gait instabilities while walking. However, this between-group differences might not be very meaningful and should be interpreted cautiously given than the range of these step length values fall within those previously reported (Sombric et al., 2019).

Participants with the motorized shoes reached lower steady state values of StepPosition (spatial) and slightly lower steady state values of StepTime (temporal) relative to the split-belt group. Our multiple regression analysis indicated that smaller speed differences (i.e. perturbation) were predictive of smaller steady state values for StepTime, but not StepPosition. Thus, perturbation size regulated the extent to which participants adapted in our temporal measure, as observed in other sensorimotor adaptation protocols of reaching (Morehead et al., 2015; Marinovic et al., 2017) or walking (Finley et al., 2015; Yokoyama et al., 2018). We did not find a direct relation between perturbation size and the reached steady state of StepPosition at an individual level, indicating that there are other factors, such as navigation strategies (Matthis et al., 2017) or practice (Day et al., 2018), influencing "where" people place their feet. Despite the subtle differences during adaptation, we saw similar after-effects between groups during early post-adaptation in all gait parameters. For example, cadence exhibited comparable changes between the groups during early adaptation and early de-adaptation, which is consistent with previous literature showing that stride time (i.e. inversely related to cadence) decreases in the beginning of adaptation (Reisman et al., 2005) and post-adaptation (MacLellan et al., 2014). In summary, our portable device induced significant adaptation and after-effects of gait asymmetries in space and time opening the door for studying locomotor adaptation outside of the laboratory.

We did not find a direct correspondence between adaptation and after-effects in neither the spatial nor the temporal domains. The positive relation between steady state values and after-effects is commonly found in reaching or saccadic movements with well-defined performance errors (Chen-Harris et al., 2008). This relation between steady-state values during the adaptation period and after-effects is, however, elusive in split-belt protocols. For example, gait parameters such as StepTime asymmetry can change dramatically during the Adaptation period (i.e. split condition) without showing any significant after-effects (Long et al., 2015; Gonzalez-Rubio et al., 2019). A recent study has also shown that changes in motor patterns during steady state split-belt walking and post-adaptation are not related and might be mediated by different neural substrates (de Kam et al., 2020). Taken together our findings further support the idea that gait adjustments during and after split-belt walking are governed by different mechanisms.

Study Implications

We found a few differences in joint motions when walking with our motorized shoes during regular walking, which will be useful for future designs of this portable device. Notably, we observed gait changes during baseline walking (i.e. both feet moving at the same speed) with the motorized shoes that were consistent with other studies showing that shoe weight (Ochsmann et al., 2016) and height (McDonald et al., 2019) alter walking movements. In addition, the rigidity of the motorized shoes' soles (Chiou et al., 2012) is another factor that might contribute to the differences that we observed in joint angles during baseline walking. Thus, our gait analysis enabled us to identify key shoe features that we will modify to reduce the effect of the motorized shoes on the regular walking pattern. This is important because contextual differences when wearing the motorized shoes could limit the extent of generalization of movements from walking with them to walking without this portable device. Locomotor adaptation with the motorized shoes overground could certainly reduce context-specific difference that limit the generalization of treadmill movements, such as visual flow (Torres-Oviedo and Bastian, 2012), walking speed (Dingwell et al., 2001), and step initiation. However, it remains to be determined whether contextual cues due to the height, weight, and rigidity of the motorized shoes would also limit the generalization of locomotor learning with them.

It is worth emphasizing that both groups were tested on a treadmill. This was done to track the movements of participants throughout the experiment, which we could not do with the motorized shoes outside the laboratory. Nevertheless, our results are promising because body-worn sensors, also referred to as wearables, now provide an inexpensive opportunity for the continuous monitoring of ambulatory activity in free-living environments (Wang and Adamczyk, 2019), which is a match to our technology. The actuation of the motorized shoes can add up to 1 m/s to the speed of each foot. Thus, we are certain that we can evoke speed differences comparable to split-belt studies (Reisman et al., 2005; Sombric et al., 2019) with these motorized shoes while walking over ground. In sum, the combination of these technologies can enable gait adaptation studies in realistic settings outside the laboratory. However, future studies with systems including adequate sensing mechanisms are needed to test this possibility.

Our results are also exciting because this portable device could also offer the possibility to study gait under more realistic situations, such as walking with self-regulated and variable gait speeds. It is well-accepted that motor variability can impact motor learning (Wu et al., 2014; Ulman et al., 2019), and walking on a treadmill is less variable compared to overground walking (Dingwell et al., 2001). Thus, having a device that can induce locomotor adaptation overground would help us gain more understanding about the relationship between variability and motor adaptation in walking. Moreover, learning a new task involves generation of new neural activity patterns, which appears after several days of practice (Oby et al., 2019).

Our device will enable training over longer periods of time because individuals will be able to train at home and gain much more practice in the altered split environment than what is currently available. This can help us contribute to recent efforts to investigate the effect of long-term practice (Hardwick et al., 2019).

There have been efforts to develop portable rehabilitation devices (Handzic et al., 2011; Afzal et al., 2015; Lahiff et al., 2016; Calabrò et al., 2018) and assistive devices (Rao et al., 2008; Awad et al., 2017; Bae et al., 2018) to improve walking patterns in individuals with gait asymmetries, such as individuals post-stroke. While these apparatus could reduce the metabolic cost associated to gait in this clinical population (Awad et al., 2017) and improve walking speed (Rao et al., 2008; Buesing et al., 2015; Calabrò et al., 2018), these devices were unsuccessful in modifying the step length asymmetry (Handzic et al., 2011), which is an important parameter in rehabilitation of post-stroke patients (Patterson et al., 2008, 2014). For example, Lahiff et al. (2016) were able to modify push-off and breaking forces, but their device was unable to change step length of the participants. Similarly, Handzic and colleagues designed a device to passively induce a speed difference between the feet (Handzic et al., 2011; Handzic and Reed, 2013). However, this passive device induced limited changes in step length asymmetry post-adaptation (i.e. ∼5% of the after-effect size observed with the split-belt treadmill and motorized shoes). In sum, our study indicates that motorized shoes could tackle previous limitations altering gait asymmetries with portable devices and thus could be potentially used to correct asymmetric steps post-stroke.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Pittsburgh Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YA was involved in the acquisition, analysis, and interpretation of the data, drafting the work, and agreement to be accountable for all aspects of the work. XZ and RS were involved in the development of the motorized shoes and providing technical expertise for using the motorized shoes. GT-O was involved in the conception and design of the work, revising the work, and agreement to be accountable for all aspects of the work. All authors contributed to revising the manuscript and providing a final approval of the version to be published.

FUNDING

This research was funded by an National Institutes of Health (NIH) grant to GT-O (K01 NS092785-01A1) and a fellowship to YA from the Department of Education, Graduate Assistantships in Areas of National Need (GAANN) program P200A150050.

ACKNOWLEDGMENTS

This manuscript has been released as a Pre-Print at https://www.biorxiv.org/content/10.1101/788943v2 (Aucie et al., 2019).

REFERENCES

Afzal, M. R., Oh, M. K., Lee, C. H., Park, Y. S., and Yoon, J. (2015). A portable gait asymmetry rehabilitation system for individuals with stroke using a vibrotactile feedback. *Biomed Res. Int.* 2015:375638. doi: 10.1155/2015/375638

Aucie, Y., Zhang, X., Sargent, R., Torres-oviedo, G., Robotics, N., and Torres-oviedo, G. (2019). Motorized shoes induce robust sensorimotor adaptation in walking. *bioRxiv [Preprint]*

Awad, L. N., Bae, J., O'Donnell, K., De Rossi, S. M. M., Hendron, K., Sloot, L. H., et al. (2017). A soft robotic exosuit improves walking in patients after stroke. *Sci. Transl. Med.* 9:eaai9084. doi: 10.1126/scitranslmed.aai9084

Bae, J., Siviy, C., Rouleau, M., Menard, N., Odonnell, K., Geliana, I., et al. (2018). "A lightweight and efficient portable soft exosuit for paretic ankle assistance in walking after stroke," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Brisbane, 2820–2827. doi: 10.1109/ICRA.2018.8461046

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bouton, M. E., Nelson, J. B., and Rosas, J. M. (1999). Stimulus generalization, context change, and forgetting. *Psychol. Bull.* 125, 171–186. doi: 10.1037/0033-2909.125.2.171

Buesing, C., Fisch, G., O'Donnell, M., Shahidi, I., Thomas, L., Mummidisetty, C. K., et al. (2015). Effects of a wearable exoskeleton stride management assist system (SMA®) on spatiotemporal gait characteristics in individuals after stroke: a randomized controlled trial. *J. Neuroeng. Rehabil.* 12:69. doi: 10.1186/s12984-015-0062-0

Buurke, T. J. W., Lamoth, C. J. C., Vervoort, D., Van Der Woude, L. H. V., and den Otter, R. (2018). Adaptive control of dynamic balance in human gait on a split-belt treadmill. *J. Exp. Biol.* 221:jeb174896. doi: 10.1242/jeb.174896

Calabrò, R. S., Naro, A., Russo, M., Bramanti, P., Carioti, L., Balletta, T., et al. (2018). Shaping neuroplasticity by using powered exoskeletons in patients with stroke: a randomized clinical trial. *J. Neuroeng. Rehabil.* 15:35. doi: 10.1186/s12984-018-0377-8

Chen-Harris, H., Joiner, W. M., Ethier, V., Zee, D. S., and Shadmehr, R. (2008). Adaptive control of saccades via internal feedback. *J. Neurosci.* 28, 2804–2813. doi: 10.1523/JNEUROSCI.5300-07.2008

Chiou, S. S., Turner, N., Zwiener, J., Weaver, D. L., and Haskell, W. E. (2012). Effect of boot weight and sole flexibility on gait and physiological responses of firefighters in stepping over obstacles. *Hum. Factors* 54, 373–386. doi: 10.1177/0018720811433464

Day, K. A., Leech, K. A., Roemmich, R. T., and Bastian, A. J. (2018). Accelerating locomotor savings in learning: compressing four training days to one. *J. Neurophysiol.* 119, 2100–2113. doi: 10.1152/jn.00903.2017

de Kam, D., Iturralde, P. A., and Torres-Oviedo, G. (2020). Cerebral contribution to the execution, but not recalibration, of motor commands in a novel walking environment. *eNeuro* [Epub ahead of print].

Deuschl, G., Toro, C., Zeffiro, T., Massaquoi, S., and Hallett, M. (1996). Adaptation motor learning of arm movements in patients with cerebellar disease. *J. Neurol. Neurosurg. Psychiatry* 60, 515–519. doi: 10.1136/jnnp.60.5.515

Dietz, V., Zijlstra, W., and Duysens, J. (1994). Human neuronal interlimb coordination during split-belt locomotion. *Exp. Brain Res.* 101, 513–520. doi: 10.1007/BF00227344

Dingwell, J. B., Cusumano, J. P., Cavanagh, P. R., and Sternad, D. (2001). Local dynamic stability versus kinematic variability of continuous overground and treadmill walking. *J. Biomech. Eng.* 123, 27–32. doi: 10.1115/1.1336798

Ericsson, A., and Pool, R. (2016). *Peak: Secrets from the New Science of Expertise*. Boston, MA: Houghton Mifflin Harcourt.

Finley, J. M., Long, A., Bastian, A. J., and Torres-Oviedo, G. (2015). Spatial and temporal control contribute to step length asymmetry during split-belt adaptation and hemiparetic gait. *Neurorehabil. Neural Repair* 29, 786–795. doi: 10.1177/1545968314567149

Gonzalez-Rubio, M., Velasquez, N. F., and Torres-Oviedo, G. (2019). Explicit control of step timing during split-belt walking reveals interdependent recalibration of movements in space and time. *Front. Hum. Neurosci.* 13:207. doi: 10.3389/fnhum.2019.00207

Haar, S., van Assel, C. M., and Faisal, A. A. (2019). Neurobehavioural signatures of learning that emerge in a real-world motor skill task. *bioRxiv [Preprint]*

Haith, A. M., and Krakauer, J. W. (2018). The multiple effects of practice: skill, habit and reduced cognitive load. *Curr. Opin. Behav. Sci.* 20, 196–201. doi: 10.1016/j.cobeha.2018.01.015

Handzic, I., and Reed, K. B. (2013). Comparison of the passive dynamics of walking on ground, tied-belt and split-belt treadmills, and via the Gait Enhancing Mobile Shoe (GEMS). *IEEE Int. Conf. Rehabil. Robot.* 2013:6650509. doi: 10.1109/ICORR.2013.6650509

Handzic, I., Vasudevan, E. V., and Reed, K. B. (2011). Motion controlled gait enhancing mobile shoe for rehabilitation. *IEEE Int. Conf. Rehabil. Robot.* 2011, 1–6. doi: 10.1109/ICORR.2011.5975417

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., and Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nat. Hum. Behav.* 3, 1252–1262. doi: 10.1101/201095

Haruno, M., Wolpert, D. M., and Kawato, M. (2001). Mosaic model for sensorimotor learning and control. *Neural Comput.* 13, 2201–2220. doi: 10.1162/089976601750541778

Iturralde, P. A., and Torres-Oviedo, G. (2019). Corrective muscle activity reveals subject-specific sensorimotor recalibration. *eNeuro* 6:ENEURO.0358-18.2019. doi: 10.1523/ENEURO.0358-18.2019

Kelly, C. W., and Sober, S. J. (2014). A simple computational principle predicts vocal adaptation dynamics across age and error size. *Front. Integr. Neurosci.* 8:75. doi: 10.3389/fnint.2014.00075

Kluzik, J., Diedrichsen, J., Shadmehr, R., and Bastian, A. J. (2008). Reach adaptation: what determines whether we learn an internal model of the tool or adapt the model of our arm? *J. Neurophysiol.* 100, 1455–1464. doi: 10.1152/jn.90334.2008

Krakauer, J. W., Pine, Z. M., Ghilardi, M., and Ghez, C. (2000). Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J. Neurosci.* 20, 8916–8924. doi: 10.1523/jneurosci.20-23-08916.2000

Lahiff, C. A., Ramakrishnan, T., Kim, S. H., and Reed, K. (2016). "Knee orthosis with variable stiffness and damping that simulates hemiparetic gait," in *Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Orlando, FL, 2218–2221. doi: 10.1109/EMBC.2016.7591170

Leech, K. A., Day, K. A., Roemmich, R. T., and Bastian, A. J. (2018). Movement and perception recalibrate differently across multiple days of locomotor learning. *J. Neurophysiol.* 120, 2130–2137. doi: 10.1152/jn.00355.2018

Long, A. W., Finley, J. M., and Bastian, A. J. (2015). A marching-walking hybrid induces step length adaptation and transfers to natural walking. *J. Neurophysiol.* 113, 3905–3914. doi: 10.1152/jn.00779.2014

MacLellan, M. J., Ivanenko, Y. P., Massaad, F., Bruijn, S. M., Duysens, J., and Lacquaniti, F. (2014). Muscle activation patterns are bilaterally linked during split-belt treadmill walking in humans. *J. Neurophysiol.* 111, 1541–1552. doi: 10.1152/jn.00437.2013

Malone, L. A., and Bastian, A. J. (2010). Thinking about walking: effects of conscious correction versus distraction on locomotor adaptation. *J. Neurophysiol.* 103, 1954–1962. doi: 10.1152/jn.00832.2009

Malone, L. A., Bastian, A. J., and Torres-Oviedo, G. (2012). How does the motor system correct for errors in time and space during locomotor adaptation? *J. Neurophysiol.* 108, 672–683. doi: 10.1152/jn.00391.2011

Marinovic, W., Poh, E., De Rugy, A., and Carroll, T. J. (2017). Action history influences subsequent movement via two distinct processes. *eLife* 6:e26713. doi: 10.7554/eLife.26713

Matthis, J. S., Barton, S. L., and Fajen, B. R. (2017). The critical phase for visual control of human walking over complex terrain. *Proc. Natl. Acad. Sci. U.S.A.* 114, E6720–E6729. doi: 10.1073/pnas.1611699114

McDonald, K. A., Devaprakash, D., and Rubenson, J. (2019). Is conservation of center of mass mechanics a priority in human walking? Insights from leg-length asymmetry experiments. *J. Exp. Biol.* 222(Pt 9):jeb195172. doi: 10.1242/jeb.195172

Morehead, J. R., Qasim, S. E., Crossley, M. J., and Ivry, R. (2015). Savings upon re-aiming in visuomotor adaptation. *J. Neurosci.* 35, 14386–14396. doi: 10.1523/JNEUROSCI.1046-15.2015

Morton, S. M., and Bastian, A. J. (2006). Cerebellar contributions to locomotor adaptations during splitbelt treadmill walking. *J. Neurosci.* 26, 9107–9116. doi: 10.1523/JNEUROSCI.2622-06.2006

Oby, E. R., Golub, M. D., Hennig, J. A., Degenhart, A. D., Tyler-Kabara, E. C., Yu, B. M., et al. (2019). New neural activity patterns emerge with long-term learning. *Proc. Natl. Acad. Sci. U.S.A.* 116, 15210–15215. doi: 10.1073/pnas.1820296116

Ochsmann, E., Noll, U., Ellegast, R., Hermanns, I., and Kraus, T. (2016). Influence of different safety shoes on gait and plantar pressure: a standardized examination of workers in the automotive industry. *J. Occup. Health* 58, 404–412. doi: 10.1539/joh.15-0193-OA

Patterson, K. K., Mansfield, A., Biasin, L., Brunton, K., Inness, E. L., and McIlroy, W. E. (2014). Longitudinal changes in poststroke spatiotemporal gait asymmetry over inpatient rehabilitation. *Neurorehabil. Neural Repair* 29, 153–162. doi: 10.1177/1545968314533614

Patterson, K. K., Parafianowicz, I., Danells, C. J., Closson, V., Verrier, M. C., Staines, W. R., et al. (2008). Gait asymmetry in community-ambulating stroke survivors. *Arch. Phys. Med. Rehabil.* 89, 304–310. doi: 10.1016/j.apmr.2007.08.142

Perry, J. B. J. (2010). *Gait Analysis Normal and Pathological Function*, 2nd Edn. Thorofare, NJ: Slack Inc.

Rao, N., Chaudhuri, G., Hasso, D., D'Souza, K., Wening, J., Carlson, C., et al. (2008). Gait assessment during the initial fitting of an ankle foot orthosis in individuals with stroke. *Disabil. Rehabil. Assist. Technol.* 3, 201–207. doi: 10.1080/17483100801973023

Reisman, D. S., Block, H. J., and Bastian, A. J. (2005). Interlimb coordination during locomotion: what can be adapted and stored? *J. Neurophysiol.* 94, 2403–2415. doi: 10.1152/jn.00089.2005

Savin, D. N., Tseng, S. C., and Morton, S. M. (2010). Bilateral adaptation during locomotion following a unilaterally applied resistance to swing in nondisabled adults. *J. Neurophysiol.* 104, 3600–3611. doi: 10.1152/jn.00633.2010

Shadmehr, R., and Mussa-ivaldi, F. A. (1994). Adaptive task of dynamics during learning of a motor. *J. Neurosci.* 14, 3208–3224. doi: 10.1523/jneurosci.14-05-03208.1994

Sharon, Y., Lendvay, T. S., and Nisky, I. (2017). Instrument orientation-based metrics for surgical skill evaluation in robot-assisted and open needle driving. *arXiv [Preprint]*. Available at: http://arxiv.org/abs/1709.09452

Smith, M. A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* 4:e179. doi: 10.1371/journal.pbio.0040179

Smith, M. A., and Shadmehr, R. (2005). Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J. Neurophysiol.* 93, 2809–2821. doi: 10.1152/jn.00943.2004

Sombric, C. J., Calvert, J. S., and Torres-Oviedo, G. (2019). Large propulsion demands increase locomotor adaptation at the expense of step length symmetry. *Front. Physiol.* 10:60. doi: 10.3389/fphys.2019.00060

Sombric, C. J., Harker, H. M., Sparto, P. J., and Torres-Oviedo, G. (2017). Explicit action switching interferes with the context-specificity of motor memories in older adults. *Front. Aging Neurosci.* 9:40. doi: 10.3389/fnagi.2017.00040

Spear, N. (1978). Retrieval of memory in animals. *Psychol. Rev.* 80, 163–194. doi: 10.1037/h0034326

Therrien, A. S., Wolpert, D. M., and Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain* 139, 101–114. doi: 10.1093/brain/awv329

Thoroughman, K. A., and Shadmehr, R. (2000). Learning of action through adaptive combination of motor primitives. *Nature* 407, 742–747. doi: 10.1038/35037588

Torres-Oviedo, G., and Bastian, A. J. (2010). Seeing is believing: effects of visual contextual cues on learning and transfer of locomotor adaptation. *J. Neurosci.* 30, 17015–17022. doi: 10.1523/JNEUROSCI.4205-10.2010

Torres-Oviedo, G., and Bastian, A. J. (2012). Natural error patterns enable transfer of motor learning to novel contexts. *J. Neurophysiol.* 107, 346–356. doi: 10.1152/jn.00570.2011

Tulving, E., and Thomson, D. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 352–373. doi: 10.1037/h0020071

Ulman, S., Ranganathan, S., Queen, R., and Srinivasan, D. (2019). Using gait variability to predict inter-individual differences in learning rate of a novel obstacle course. *Ann. Biomed. Eng.* 47, 1191–1202. doi: 10.1007/s10439-019-02236-x

Wang, W., and Adamczyk, P. G. (2019). Analyzing gait in the real world using wearable movement sensors and frequently repeated movement paths. *Sensors* 19:1925. doi: 10.3390/s19081925

Wilken, J. M., Rodriguez, K. M., Brawner, M., and Darter, B. J. (2012). Reliability and minimal detectible change values for gait kinematics and kinetics in healthy adults. *Gait Posture* 35, 301–307. doi: 10.1016/j.gaitpost.2011.09.105

Winter, D. A. (1987). *The Biomechanics and Motor Control of Human Gait*, 2nd Edn. Waterloo: University of Waterloo Press. doi: 10.1016/s0031-9406(10)63713-3

Wu, H. G., Miyamoto, Y. R., Gonzalez Castro, L. N., Ölveczky, B. P., and Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* 17, 312–321. doi: 10.1038/nn.3616

Yokoyama, H., Sato, K., Ogawa, T., Yamamoto, S. I., Nakazawa, K., and Kawashima, N. (2018). Characteristics of the gait adaptation process due to split-belt treadmill walking under a wide range of right-left speed ratios in humans. *PLoS One* 13:e0194875. doi: 10.1371/journal.pone.0194875

Zhang, X. (2017). *A Control System for Bionic Shoes*. Available at: https://www.ri.cmu.edu/wp-content/uploads/2017/08/XunjieZhang_CMU-RI-TR-17-39.pdf (accessed July 2017).

Conflict of Interest: XZ was employed by the company Nimbus Robotics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fpubh.2020.479431

Mental Condition Monitoring Based on Multimodality Biometry

Masashi Kiguchi[1]*, *Stephanie Sutoko*[1], *Hirokazu Atsumori*[1], *Ayako Nishimura*[1], *Akiko Obata*[1], *Tsukasa Funane*[1], *Hiromitsu Nakagawa*[2], *Masashi Egi*[2] and *Hiroyuki Kuriyama*[3]

[1] Center for Exploratory Research, Hitachi, Ltd., Kokubunji, Japan, [2] Center for Technology Innovation, Hitachi, Ltd., Kokubunji, Japan, [3] Global Center for Social Innovation, Hitachi, Ltd., Kokubunji, Japan

We have developed a system with multimodality that monitors objective biomarkers for screening the mental distress in the office. A field study using a prototype of the system was performed over four months with 39 volunteers. We obtained PC operation patterns using a PC logger, sleeping time and activity levels using a wrist-band-type activity tracker, and brain activity and behavior data during a working memory task using optical topography. We also administered two standard questionnaires: the Brief Job Stress Questionnaire (BJS) and the Kessler 6 scale (K6). Supervised machine learning and cross validation were performed. The objective variables were mental scores obtained from the questionnaires and the explanatory variables were the biomarkers obtained from the modalities. Multiple linear regression models for mental scores were comprehensively searched and the optimum models were selected from 2,619,785 candidates. Each mental score estimated with each optimum model was well correlated with each mental score obtained with the questionnaire (correlation coefficient = 0.6–0.8) within a 24% of estimation error. Mental scores obtained by means of questionnaires have been in general use in mental health care for a while, so our multimodality system is potentially useful for mental healthcare due to the quantitative agreement on the mental scores estimated with biomarkers and the mental scores obtained with questionnaires.

Keywords: multimodality, multivariate linear regression, PC logger, activity tracker, near-infrared spectroscopy (fNIRS)

OPEN ACCESS

Edited by:
Berthold Langguth, University of Regensburg, Germany

Reviewed by:
Mitsunori Ogihara, University of Miami, United States
Franziska C. Weber, University of Regensburg, Germany

***Correspondence:**
Masashi Kiguchi
masashi.kiguchi.py@hitachi.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *19 June 2019*
Accepted: *01 September 2020*
Published: *22 October 2020*

Citation:
Kiguchi M, Sutoko S, Atsumori H, Nishimura A, Obata A, Funane T, Nakagawa H, Egi M and Kuriyama H (2020) Mental Condition Monitoring Based on Multimodality Biometry. Front. Public Health 8:479431. doi: 10.3389/fpubh.2020.479431

INTRODUCTION

Mental disorders are of significant concern in terms of not only public health but also economic development and social welfare (1). Depression affects over 120 million people and causes long absences from work and increased risk of suicide. Organization for Economic Co-operation and Development (OECD) reported that mild-to-moderate mental disorders affect around 20% of the working-age population in the average OECD country and predominantly include highly treatable disorders such as anxiety and depression. Although treating depression in primary care is feasible and very cost-effective, studies have shown that 56.3% of patients do not get sufficient care (2). Furthermore, U.S. workers suffering from depression cost employers an additional 31 billion dollars each year in lost productive time (3). A key management issue facing enterprises today is ensuring that they prevent depression, support the return to work, and prevent recurrence. Many companies have services for managing employees' mental condition, such as the employee assistance program (EAP)[1]. Because the mental distress is a risk factor of depression, early detection of risk factors

[1] *Web Page of International Employee Assistance Professionals Association (EAPA)*. Available online at: https://www.eapassn.org/FAQs

could be exploited for prevention purposes. Therefore, the accurate and low-cost monitoring of mental condition is required for screening distress in office.

Information and Communication Technology (ICT) and Internet of Things (IoT) have the potential to provide a low-cost condition monitoring system with high accuracy for mental healthcare. Especially, the mental distress levels observed using questionnaires are widely used for screening. A variety of devices for monitoring mental stress have already been developed (4). We previously developed a PC logger, a wrist-band type activity tracker, and a wearable optical topography as a non-invasive brain activity sensor. These have been independently used in several studies on mental healthcare application. The fractal dimensions obtained from the PC logger data are expected to be related with mood states (5). The usage of the wrist-band type activity tracker for judgments of reinstatement reduced the ratio of re-leave (6). The brain activity measured by optical topography during working memory tasks was affected by mood state (7–9). The optical topography measurement was applied to return-to-work trainees in remission of mental disorders with depressive symptoms (10). In this paper, we introduce a multimodality data acquisition system we developed to combine these devices for monitoring multilateral conditions of mental health. The combination of behavioral and brain measurements is a novel approach.

METHOD

Multimodality Measurements

The equipment and data acquisition system used in this study are shown in **Figure 1**. The PC logger is our original software developed for the Windows OS. It was installed on each participant's PC and hooked key- and mouse-events with time stamps as a background process. In order to avoid an information security risk, the kind of tapped key, i.e., alphabet key, number key, or special key, was recorded instead of key characters. Mouse button click and mouse movement distance were recorded as well. A cumulative distribution of the key/mouse events frequently showed power of event intervals. These power exponents are related to the total number of event and operator mood states. Using the key/mouse data recorded over the course of a day, a fractal dimension was obtained as a slope of fitted line for cumulative distribution vs. time-interval graph. Because the fractal dimension also depends on the workload in a day, the key/mouse index $d_{key/mouse}$ for the day was obtained as a deviation from the fractal dimension stemming from workload (5).

The wrist-band type activity tracker, Life Scope (LS, Hitachi, Ltd.), is equipped with a triaxial accelerator. All raw signals of the accelerator are stored in its own memory. A variety of features were calculated using the raw data in the server (11). Among

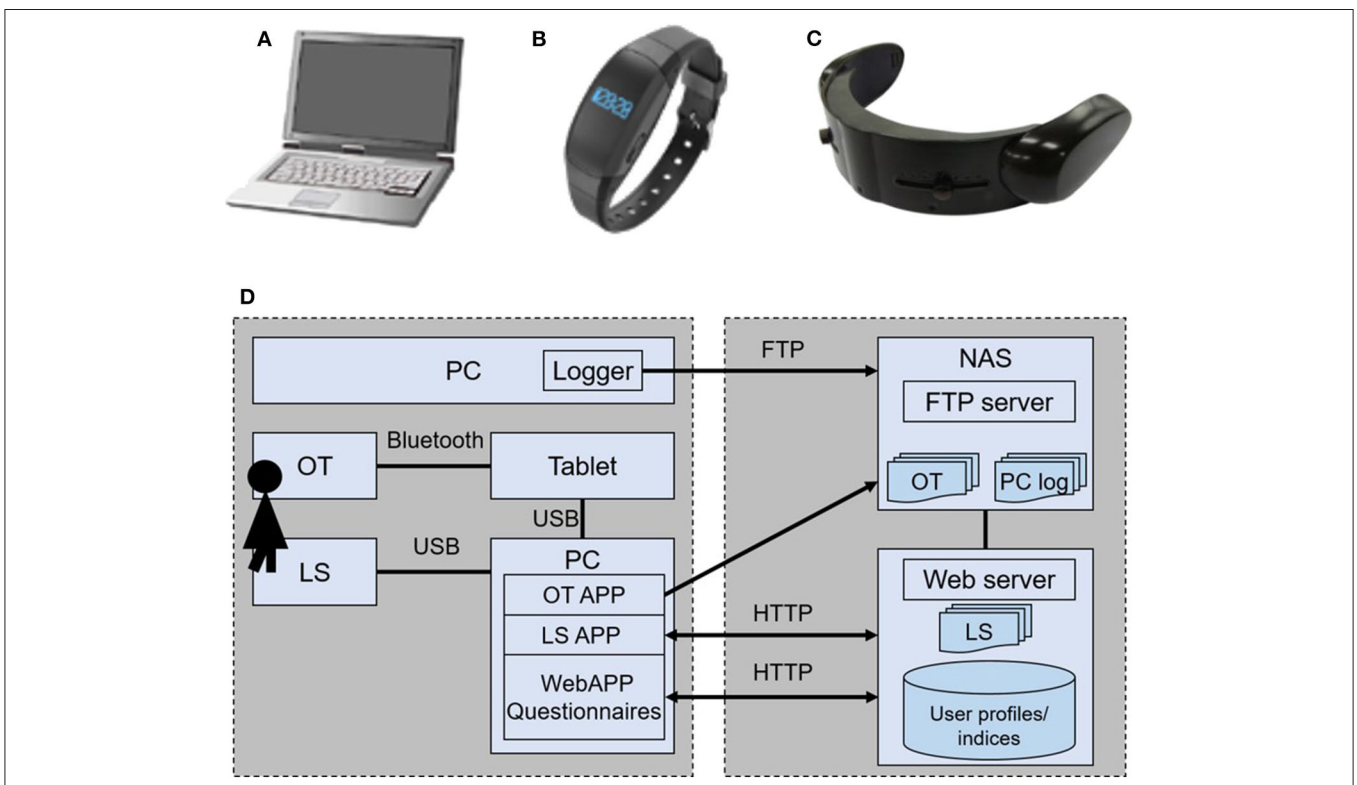


FIGURE 1 | Devices used for monitoring mental conditons: **(A)** PC logger, **(B)** Life Scope (LS), and **(C)** HOT-1000 optical topography (OT). **(D)** The data acquisition and database system.

them, we used steps, activity strength in metabolic equivalents (METs), and sleeping hours for the analysis.

Optical topography (OT) is a tool for measuring cerebral blood volume change associated with brain activity on the basis of Near-infrared spectroscopy (NIRS). Two regions on the forehead were covered by a handy and wireless model, HOT-1000 (Hitachi High-technologies Corporation). The headset was connected through Bluetooth with a tablet PC that provided a task for activating brain function, acquired data, and displayed the results. The brain activity during a working memory (WM) task (namely, a delayed matching task; see **Figure 2**), was observed by OT for estimating mood states (10). Increases of oxygenated hemoglobin during spatial and verbal working memory tasks at the left and right sides of the forehead were measured. The response time to answer and the rate of correct answers were also recorded.

Each item of time-series data obtained by the PC logger, wristband type activity tracker, and optical topography was stored in the time-series analysis database through the local network. The query object graph analysis method was used to enable analysis sharing among analysts by preventing duplicate processing and data explosion (12).

Questionnaires

The Kessler psychological distress scale (K6) (13) and a part of the Brief Job Stress Questionnaire (BJS) (14)[2] were used for obtaining mental scores. The K6 score is the standard questionnaire for screening the mental distress. When the K6 score is lower, the condition is better. The cutoff is 5. BJS consists of 52 questions concerning job, health, and people surrounding the respondent. As the situations concerning job and people did not frequently change during the field test, 29 questions concerning health were used for determining subjective mental scores of lassitude, irritation, fatigue, anxiety, depression, and physical stress. All mental scores of BJS were converted into values from 1 to 5 in accordance with the manual[3]. When the mental score is higher, the condition is better. A total value of mental scores lower than 12 is regarded as stressful.

Trial in Office

Thirty-nine healthy volunteers (32 males, 7 females, 43.7 ± 8.9 years old) with no history of mental disorders participated in the measurement over four months. Neither measured data nor information were fed back to participants so as to avoid any effects on the trial. The data from the volunteers were obtained according to the regulations set forth by the internal review board at the Central Research Laboratory, Hitachi, Ltd., following receipt of their written informed consent.

The PC loggers were installed on all PCs used by participants in the office. The keyboard and mouse operations were recorded the entire time PCs were running. All logs were combined for each participant according to the timestamp. The key index and the mouse index were calculated using data from one day and averaged over one week.

The participants were asked to wear the LS wristband all day except when bathing. Data from the LS was sent to the server

[2]*The Brief Job Stress Questionnaire English version*. Available online at: https://www.mhlw.go.jp/bunya/roudoukijun/anzeneisei12/dl/160621-1.pdf

[3]Available online at: http://www.tmu-ph.ac/topics/pdf/sotenkansan.pdf (in Japanese).

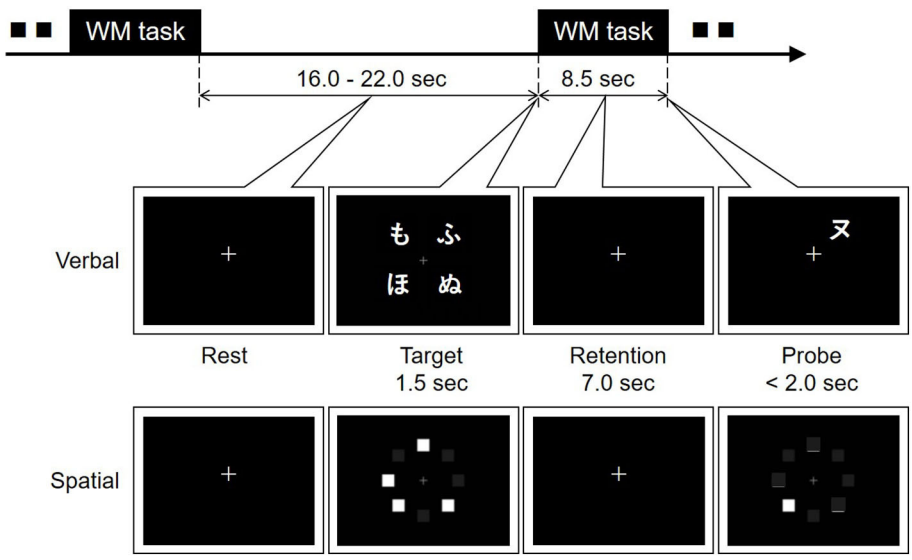


FIGURE 2 | Verbal and spatial delayed matching tasks for optical topography measurement. Four Japanese characters (*hiragana*) or four squares were displayed for 1.5 s on a tablet PC as a target in the verbal or spatial working memory task, respectively. After memory retention for 7.0 s, another Japanese character (*katakana*) or one square was shown as a probe. In the verbal task, participants were asked to click the circle button or the cross button displayed on the tablet PC when the probe character had the same pronunciation as one of the target characters or when it did not. In the spatial task, the agreement of the position between the probe and one of the targets was answered in the same way.

TABLE 1 | The number of explanatory variable and model for each target variable.

Target variable	Explanatoy variable	Model
Lassitude	24	28230
Irritation	31	108551
Fatigue	27	56103
Anxiety	19	6239
Depression	16	2288
Physical stress	20	10354
Total score	16	2816
K6	15	2777
Total	–	217358
Average	21	27169.75

through a local area network once a week. The values of steps, METs, and sleeping hours were obtained as averages of one week.

Once a week at the same time, the OT measurements were performed and the questionnaires were administered. While the dates and times for each participant were set in principle, they were nevertheless flexible in order to ensure that enough data were gathered.

Analysis

A linear multiple regression model was obtained for each mental score using multimodality indices, as

$$y[i] = c_0 + \sum_j^m c_j x_j[i] \tag{1}$$

$$\varepsilon^2 = \frac{\sum_{i=1}^{N} \left(y_o[i] - y[i]\right)^2}{N}, \tag{2}$$

with target variable y_o, explanatory variable x_j, regression coefficients c_j, square error ε^2, the number of explanatory variables m, and the number of data N. Target variables were scores of lassitude, irritation, fatigue, anxiety, depression, physical stress, total, and K6. The obtained multimodality indices are listed in **Table 1**. In order to check the possibility of prediction, each index was observed one week and two weeks before the score sheet was tallied. In total, 51 explanatory variables were obtained.

Here, we investigated three cases of $m = 3$, 4, and 5. The total combinations of choosing 3, 4, and 5 indices out of 51 indices equal to 2,619,785 for one mental score. More than twenty million cases of calculation were required for eight mental scores. In order to reduce the number of calculations, each explanatory variable that had a correlation coefficient with each target variable of <0.1 was rejected as a target variable. Also, combinations of the same index for a different week were not permitted. As a result, the number of combinations was reduced to one-hundredth.

The number of the explanatory variables may be different across participants because the measurement schedules did not match the participants' ones (e.g., business trips, days off). When the temporally shifted explanatory variables (i.e., one, two weeks in prior to the measured target variables) were used in regression models, the temporal data of the target variables consequently decreased. Therefore, missing some temporal data inevitably occurred; only data having the complete temporal information were included in the regression analysis. Neither data normalization nor further elimination were performed because there was no observed improvement (data not shown).

The three-fold cross validation was performed to select each optimum model for each target variable. The fold number was determined to ensure that a sufficient amount of data was included in each subset. The subsets were created based on the participant-wise approach to validate participant dependency on models. No data measured from a participant at different measurement times were assigned in different subsets; models with the minimum participant-dependent effect were selected. In order to avoid dropping the models with a larger correlation coefficient r and slightly larger ε^2 than the model with minimum ε^2, we define an evaluation index V for selecting candidates of the optimum model as follows:

$$V = \frac{r - < r >}{\sigma_r} - \frac{\varepsilon^2 - < \varepsilon^2 >}{\sigma_{\varepsilon^2}}, \tag{3}$$

where $<r>$, $<\varepsilon^2>$, σ_r, and σ_{ε^2} are mean values and standard deviations of r and ε^2, respectively. Both r and ε^2 were standardized to equally contribute to V. The optimum model for each mental score whose ε^2 was the smallest among ten candidate models identified using V was chosen.

RESULTS AND DISCUSSION

Table 1 shows the valid target variables and the numbers of models for each target variable obtained after the reduction described above. The numbers of significant explanatory variables (x_{sig}) and regression models $\left(C^3_{x_{sig}} + C^4_{x_{sig}} + C^5_{x_{sig}}\right)$ were varied across the target variables. The total number of models was 217,358, which means we were able to reduce the computation cost by 1/1000 compared to the initial one.

The explanatory variables are shown in **Table 2**. The postfix, "_n" means that the explanatory variable was obtained n weeks before the measurement of target variables. For example, Ped_*2* is the mean value of steps in a week obtained two weeks in prior to the measurement of mental scores. The number of data (N; the complete temporal data) ranged from about 30 to 70.

Figure 3 shows the relationship between mental scores and values estimated using each optimum model shown in **Table 3**. The best combination of explanatory variables was selected for each target variable. For example, the best model for depression is described as below.

$$\begin{aligned}\text{Depression} = {} & 1.25 + 6.53^{*}\text{keylog\_0} + 4.06^{*}\text{mlog\_2} \\ & + 2.43^{*}\text{ot\_cr\_2} + 0.000149^{*}\text{ot\_s\_rt\_2}.\end{aligned} \tag{4}$$

No explanatory variables taken one week before the measurement of depression score was used in the above model.

TABLE 2 | Explanatory variables.

Name	Modality	Description
Ped_*n*	LS	Steps in a day
mets_*n*	LS	Metabolic equivalents in a day
sleep_*n*	LS	Sleeping time in a day
keylog_*n*	BM1	Fractal dimension of key operation in a day
mlog_*n*	BM1	Fractal dimension of mouse operation in a day
ot_s_l_*n*	OT	Left PFC activity during spatial working memory task
ot_s_r_*n*	OT	Right PFC activity during spatial working memory task
ot_v_l_*n*	OT	Left PFC activity during verbal working memory task
ot_v_r_*n*	OT	Right PFC activity during verbal working memory task
ot_sv_l_*n*	OT	Left PFC activity (spatial—verbal)
ot_sv_r_*n*	OT	Right PFC activity (spatial–verbal)
ot_s_rl_*n*	OT	Laterality of PFC activity during spatial working memory task
ot_v_rl_*n*	OT	Laterality of PFC activity during verbal working memory task
ot_s_cr_*n*	OT	Correction rate of spatial working memory task
ot_v_cr_*n*	OT	Correction rate of verbal working memory task
ot_s_rt_*n*	OT	Response time for spatial working memory task
ot_v_rt_*n*	OT	Response time for verbal working memory task

"_n" indicates that the values were obtained n weeks before the mental score.

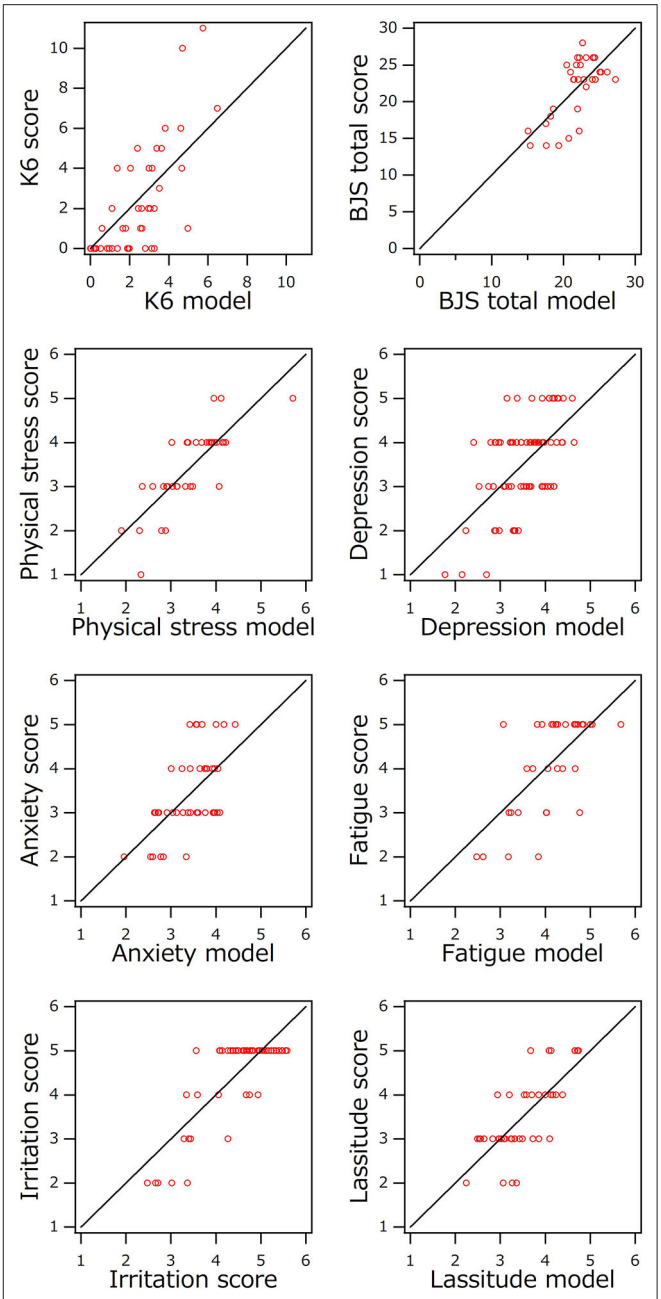


FIGURE 3 | Comparison of mental scores and values estimated using optimized multivariate linear regression model.

According to **Table 3**, the correlation coefficients were 0.6 for depression and anxiety and 0.7–0.8 for others. Each error ε for each mental score, lassitude, irritation, fatigue, anxiety, depression, or physical stress was about 1, which is almost the same as the minimum scale of score. The errors for total score and K6 were 5.2 and 2.9, and the full scores for total score and K6 were 30 and 24, respectively. Each error was within 24% of each full score. Considering the accuracy of the subjective score sheet, these errors seem acceptable for practical use.

The coefficients of variation (CVs) of mean squared errors across participants for the target variables were around one (**Figure 4**). When CV is one, the standard deviation is equal to the mean value. The effect of participant dependency on models was remarkable, although it had been tried to be controlled through the participant-wise cross validation. Other participant features, such as gender and age, were not controlled in this study due to the small number of data. Case analysis potentially improves the robustness of model. Increasing the number of participants data possibly provide the models with much smaller participant dependency.

We should point out here that the optimum model for K6 consisted of explanatory variables measured one or two weeks before the mental score. As such, these multimodality measurements predicted the mental condition before participants made a subjective complaint. Because K6 is used for screening mental distress, the system shows potential as an early screening technique.

Major participants were healthy volunteers. Neither clinical diagnosis nor intervention were performed according to the results obtained in this study. Therefore, the effectiveness of this system for screening and preventing mental distress has not been confirmed yet. This system remains to be investigated further. Even though the current results were preliminary, this system showed a promising function to replace the conventional mental health care services based on manual questionnaire sheets.

TABLE 3 | Optimum model for each mental score with correlation coefficient r, estimation error ε, and percentage of ε to full score.

Mental score	c0	c1/var	c2/var	c3/var	c4/var	c5/var	*r*	*ε*	*ε [%]*
Lassitude	4.17	−0.00341	0.00316	−0.0257	0.204	−7.82	0.765	0.935	18.7
		ot_s_rt_0	ot_v_rt_0	sleep_1	ot_s_r_1	mlog_2			
Irritation	7.76	0.305	0.00184	−4.36	−2.21	−0.192	0.798	0.780	15.6
		ot_v_l_1	ot_v_rt_1	mets_2	mlog_2	ot_sv_r_2			
Fatigue	3.93	0.352	0.177	−0.0483	−0.000195	−0.233	0.688	1.19	23.8
		ot_v_rt_0	ot_s_r_1	ot_s_rt_1	sleep_2	mlog_2			
Anxiety	−1.65	−0.598	3.53	−0.00108	0.00290	−0.0554	0.624	0.961	19.2
		mlog_0	ot_v_cr_0	ot_s_rt_1	ot_v_rt_1	sleep_2			
Depression	1.25	6.53	4.06	2.43	0.000149		0.627	1.09	21.8
		keylog_0	mlog_2	ot_s_cr_2	ot_s_rt_2				
Physical stress	1.23	0.273	−11.4	−0.137	−0.504	−0.0228	0.826	0.874	17.5
		sleep_0	mlog_1	ot_s_r_1	ot_s_rl_1	ot_s_cr_2			
Total score	−4.79	−0.135	1.28	14.3	0.0116	0.252	0.762	5.22	17.4
		sleep_0	ot_v_r_0	ot_s_cr_1	ot_v_rt_1	ot_s_l_2			
K6	8.14	−0.007028	0.339	−27.5	49.6		0.738	2.88	12.0
		ot_v_rt_1	sleep_2	keylog_2	mlog_2				

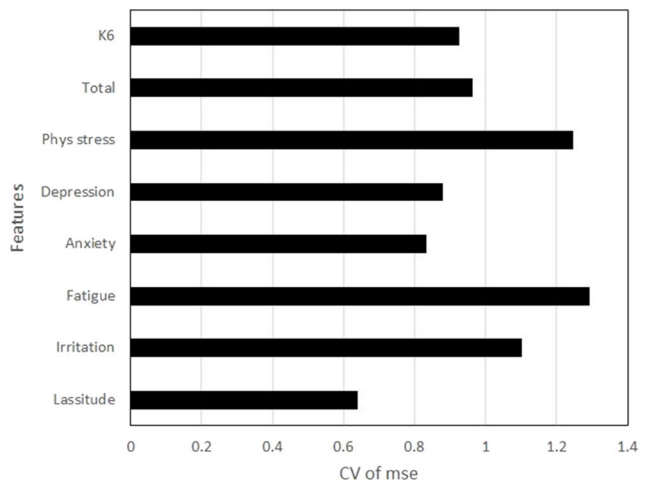


FIGURE 4 | Coefficients of variation of mean squared errors (mse) across participants for the target variables.

CONCLUSION

We developed a monitoring system for mental condition involving the PC logger, the activity tracker, and Optical Topography (OT). We collected the biometric data from thirty-nine healthy volunteers in office for more than four months. The multivariate linear models for mental scores of BJS and K6 were obtained by using the supervised machine learning and the cross-validation. Those models included several variables from the collected biometric data such as the fractal dimensions of PC operation obtained from the PC log, steps, METs, and sleeping time from the activity tracking log, brain activities, laterality, correction rates and response time during working memory tasks from the brain activity and performance log. Each mental score estimated by each model was well agreed with each score of questionnaire. Especially, K6 score was estimated by using the biometric data collected from one or two weeks before. The system is potentially useful for the mental healthcare including the prevention.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. It was not included in the ethics approval to submit the datasets.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the internal review board of Central Research Laboratory, Hitachi, Ltd. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MK conceived the concept and led the project. SS and MK analyzed the data. HA, AN, and AO performed the field trial and

the data preprocessing. HN developed the platform and database. ME and HK developed the PC logger and the activity tracker and analyzed each data, respectively. MK, HA, and TF developed the wearable OT. All authors contributed to the article and approved the submitted version.

FUNDING

The authors declare that this study received funding from the research fund contributed by Hitachi Group companies. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors gratefully thank Drs. Keiro Muro and Takeshi Tanaka at the Center for Technology Innovation, Hitachi, Ltd. for their significant contributions to developing the data-acquisition and database system.

REFERENCES

1. *Investing In Mental Health:Evidence For Action*. Kokubunji: World Health Organization (2013).
2. Hewlett E, Moran V. *Making Mental Health Count*. OECD Publishing (2014).
3. Stewart W, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. *JAMA*. (2003) 289:3135–45. doi: 10.1001/jama.289.23.3135
4. Kamišalić A, Fister I Jr, Turkanović M, Karakatič S. Sensors and functionalities of non-invasive wrist-wearable devices: a review. *Sensors*. (2018) 18:E1714. doi: 10.3390/s18061714
5. Egi M, Ishikawa H. Scale-free dynamics of human behavior in personal computer operations. In: *10th International Conference on Management of Digital EcoSystems*. Tokyo (2018).
6. Hayashi T. A study on estimation method of business performance using wristband type Life monitor. In: *the 70th National Industrial Safety and Health Convention 2011*. Tokyo (2011).
7. Aoki R, Sato H, Katura T, Utsugi K, Koizumi H, Matsuda R, et al. Relationship of negative mood with prefrontal cortex activity during working memory tasks: an optical topography study. *Neurosci Res*. (2011) 70:189–96. doi: 10.1016/j.neures.2011.02.011
8. Sato H, Aoki R, Katura T, Matsuda R, Koizumi H. Correlation of within-individual fluctuation of depressed mood with prefrontal cortex activity during verbal working memory task: optical topography study. *J Biomed Opt*. (2011) 16:126007. doi: 10.1117/1.3662448
9. Sato H, Dresler T, Florian B, Andreas HJ, Ann-Christine E. Replication of the correlation between natural mood states and working memory-related prefrontal activity measured by near-infrared spectroscopy in a German sample. *Front Hum Neurosci*. (2014) 8:37. doi: 10.3389/fnhum.2014.00037
10. Atsumori H, Obata A, Sato H, Funane T, Yamaguchi R, Kiguchi M. Prefrontal cortex activation of return-to-work trainees in remission of mental disorders with depressive symptoms compared to that of healthy controls. *J Biomed Optics*. (2019) 24:056008. doi: 10.1117/1.JBO.24.5.056008
11. Sasaki S, Ukawa S, Okada E, Wenjing Z, Kishi T, Sakamoto A, Tamakoshi A. Comparison of a new wrist-worn accelerometer with a commonly used triaxial accelerometer under free-living conditions. *BMC Res Notes*. (2018) 11:746. doi: 10.1186/s13104-018-3849-9
12. Nakagawa H, Muro K. Analysis sharing method by managing provenance of query. In: *IEEE Systems Conference 2017*. Montreal, QC (2017).
13. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psycol Med*. 32:959-76. doi: 10.1017/s0033291702006074
14. Ando E, Kawakami N, Shimazu A, Shimomitsu T, Odagiri Y. Reliability and validity of the English version of the New Brief Job Stress Questionnaire. In: *Presented at the 31st International Conference on Occupational Health*. Seoul (2015).

Conflict of Interest: All authors were employed by company Hitachi, Ltd.

frontiers in Public Health

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fpubh.2021.490604

Effective Adoption of Tablets for Psychodiagnostic Assessments in Rural Burundi: Evidence for the Usability and Validity of Mobile Technology in the Example of Differentiating Symptom Profiles in AMISOM Soldiers 1 Year After Deployment

Roland Weierstall[1*†], *Anselm Crombach*[2,3†], *Corina Nandi*[2], *Manassé Bambonyé*[3], *Thomas Probst*[4] *and Rüdiger Pryss*[5]

[1] Department of Psychology, Medical School Hamburg, Hamburg, Germany, [2] Department of Psychology, University of Konstanz, Konstanz, Germany, [3] Department of Clinical Psychology, University Lumière de Bujumbura, Bujumbura, Burundi, [4] Department for Psychotherapy and Biopsychosocial Health, Danube University Krems, Krems, LA, Austria, [5] Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany

OPEN ACCESS

Edited by:
Andreea Molnar,
Swinburne University of Technology, Australia

Reviewed by:
Sheila A. M. Rauch,
Emory University, United States
Eric Vermetten,
Leiden University, Netherlands

***Correspondence:**
Roland Weierstall
roland.weierstall@medicalschool-hamburg.de

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: 18 September 2019
Accepted: 19 March 2021
Published: 15 April 2021

Citation:
Weierstall R, Crombach A, Nandi C, Bambonyé M, Probst T and Pryss R (2021) Effective Adoption of Tablets for Psychodiagnostic Assessments in Rural Burundi: Evidence for the Usability and Validity of Mobile Technology in the Example of Differentiating Symptom Profiles in AMISOM Soldiers 1 Year After Deployment. Front. Public Health 9:490604. doi: 10.3389/fpubh.2021.490604

Research on the use of mobile technology in health sciences has identified several advantages of so-called mHealth (mobile health) applications. Tablet-supported clinical assessments are becoming more and more prominent in clinical applications, even in low-income countries. The present study used tablet computers for assessments of clinical symptom profiles in a sample of Burundian AMISOM soldiers (i.e., African Union Mission to Somalia; a mission approved by the UN). The study aimed to demonstrate the feasibility of mHealth-supported assessments in field research in Burundi. The study was conducted in a resource-poor setting, in which tablet computers are predestined to gather data in an efficient and reliable manner. The overall goal was to prove the validity of the obtained data as well as the feasibility of the chosen study setting. Four hundred sixty-three soldiers of the AMISOM forces were investigated after return from a 1-year military mission in Somalia. Symptoms of posttraumatic stress disorder (PTSD) and depression were assessed. The used data-driven approach based on a latent profile analysis revealed the following four distinct groups, which are based on the soldiers' PTSD and depression symptom profiles: Class 1: moderate PTSD, Class 2: moderate depression, Class 3: low overall symptoms, and Class 4: high overall symptoms. Overall, the four identified classes of soldiers differed significantly in their PTSD and depression scores. The study clearly demonstrates that tablet-supported assessments can provide a useful application of mobile technology in large-scale studies, especially in resource-poor

settings. Based on the data collected for the study at hand, it was possible to differentiate different sub-groups of soldiers with distinct symptom profiles, proving the statistical validity of the gathered data. Finally, advantages and challenges for the application of mobile technology in a resource-poor setting are outlined and discussed.

Keywords: tablet computer, application, post-deployment aggression, PTSD, depression, latent-profile-analysis, soldiers, mobile data collection

INTRODUCTION

In recent years, more and more studies and meta-analyses have investigated and shown the usefulness of mobile technology in psychological research, clinical assessments, and therapeutic interventions [e.g., (1, 2)]. In particular, digital assessments and interventions offer several advantages but also challenges [e.g., (3, 4)]. On the one hand, the collection of data can be improved by reducing time efforts and resources, especially when large data sets have to be processed in a rather short period of time (5). On the other hand, a suitable infrastructure has to be provided, additional development costs have to be covered, and data security issues must be addressed [e.g., (6, 7)].

However, the implementation of mobile health (mHealth) applications is not solely bound to practical issues. The validity of the gathered data via mobile devices has to be considered carefully. As the proliferation of mobile technology has increased by orders of magnitude, more and more researchers address validity issues in the context of mHealth-collected data. In the beginning of the development of mHealth applications, no guidelines existed on how interventions based on mHealth have to be reported in scientific applications, making it difficult to compare the quality of research designs. The World Health Organization (WHO) mHealth Technical Evidence Review Group therefore developed the mHealth evidence reporting and assessment (mERA) checklist in 2016, covering 16 items to be addressed when reporting mHealth applications in scientific publications (8).

Although extant research on mHealth mostly stems from western countries, several studies in resource-poor countries have emerged in recent years that used mobile technology; e.g., in the East African countries Burundi, Rwanda, and Uganda [e.g., (9–11)]. The latter studies covered aspects on interventions as well as diagnostic questions. This demonstrates that mHealth does not have to be limited to studies in first-world countries, but can be applied all over the world. The validity of the psychodiagnostic assessments, however, is a substantial and necessary prerequisite for any further clinical research for example on interventions based on mHealth applications and mobile technology. Therefore, the use of mobile technology is demonstrated in this paper based on psychodiagnostic assessments in a sample of Burundian soldiers. The conducted study utilized tablet computers to carry out standardized clinical interviews digitally. Instead of solely focusing on the users' willingness to use a mobile application, this paper aims to demonstrate - on the basis of a data-driven approach - that the gathered data is valid and suitable for further clinical research questions.

There is a considerable amount of public attention on psychological consequences of mental health problems in soldiers in the aftermath of their deployment (12, 13). Among the various disorders reported in the literature, the post-traumatic-stress disorder (PTSD) and depression are among the most common ones (14). While some studies primarily focus on either depression (15) or PTSD symptoms (16), others acknowledge the high comorbidity between the two disorders (17). Although the relation between PTSD and depression in soldiers after deployment is still under debate (18), it is reasonable to assume that different subgroups of soldiers exist. In these subgroups, in turn, soldiers might either display predominant PTSD or depression symptoms, or suffer from both of them (comorbidity). The present study therefore investigated whether or not distinct depression and PTSD symptom profiles can be identified in AMISOM (African Union Mission to Somalia) soldiers, based on a tablet-supported data collection solution.

The AMISOM (19, 20) is a UN Security Council authorized peacekeeping mission, sending military troops based in Uganda, Kenya, Burundi, Djibouti, and Ethiopia to stabilize Somalia and try to reclaim territories from the Al-Shabaab militia. AMISOM troops frequently cope with attacks and armed fights in a hostile environment. Beside mission related hassles, most Burundian soldiers experience a civil war that already started in 1993 and lasted more than a decade (21–24). During this period, the currently active Burundian soldiers often must fight against each other, including fights with former government troops or rebel movements. As a consequence of these particular circumstances, the affected population is challenged with severe traumas. Therefore, it is promising – and required - to investigate trauma- and deployment related mental-health consequences in these people.

The aim of this study was not only to demonstrate the general feasibility of structured clinical interviews in a low-income country that are accomplished through the use of tablet computers. On top of that, privacy concerns are discussed (25), which might have an impact on the participation in tablet computer guided interviews, leading to a general tendency to either disclose information or producing meaningless or invalid data. As an approach to validate the experts' ratings conducted in this study, it was tested, whether analyses of the collected data are capable to differentiate distinct symptom profiles in PTSD and depression symptoms. Altogether, a sample of 463 participants from the AMISOM was investigated 1 year after their deployment. A latent profile analyses (LPA) was conducted to separate different symptom profile groups. We expected to identify distinct diagnostic groups according to PTSD and depression.

MATERIALS AND METHODS

Participants

The tabled-based diagnostic procedure was accomplished in a larger project aiming to improve the mental health status of Burundian soldiers of the AMISOM mission [for further details on PTSD rates and specific types of trauma-exposure pre- and peri-deployment see (17)]. The composition of the survey was limited to PTSD and depression as main psychiatric disorders and was primarily concerned with the assessment of deployment-associated risk factors. In general, this sample was exposed to various traumatic pre-deployment events [Median = 11 (17)], many of them thereby related to the Burundian civil war. Many of the sample have faced traumatic incidents during their deployment [Median = 5 (17)], including being attacked by an enemy, experiencing suicide attack, or witnessing comrades being killed.

For the present analysis, only full data sets of 463 participants were included who had been assessed 1 year after returning from their AMISOM deployment. No data imputation method was used in order to avoid any bias due to modeling data prior to the actual data-driven analysis. All participants were male. Mean age was 35 (SD = 5 years) at the time of interview (i.e., post deployment). Out of the 463 soldiers, 81% reported to be married, whereas the rest was not in a stable relationship. Seventy-nine percentage of the soldiers had at least one child. On average, the soldiers had received 6 years of formal education (SD = 2 years). Out of the 463 soldiers who reported their military rank, the following main ranks were obtained: major: corporal: 219 (48%); chief corporal: 173 (38%); other: 66 (14%).

The participation was on a voluntary basis. Participants received no incentives for participation. All participants gave written informed consent. Oral informed consents were collected in case of illiteracy. The Ethical Review Boards of the University of Konstanz and the University of Bujumbura, Burundi approved the study. The study was conducted in cooperation with the Force de la Défense National, Burundi. All parties involved granted strict confidentiality, acknowledging the specific vulnerability of the target population.

Study Procedure

Clinical symptoms were assessed using standardized clinical interviews, guided by a survey implemented for tablet computers (i.e., Apple ipads). All participants were released from their routine duty for the interviews. To ensure anonymity and confidentiality, electronic coding and storage of the data was utilized, which fulfilled the highest and most secure data encryption standards (7). Before their application in the interviews, all questionnaires had been translated into Kirundi, using back and forth translations (26). Trained mental health experts from Burundi and Germany conducted all interviews in Kirundi, so that literacy was not an issue. Bi-lingual local interpreters supported the German mental health experts. All questionnaires were translated from English into Kirundi using back-and-forth translations. All translations were discussed in an experts' panel consisting of bi-lingual translators as well as mental health experts from Burundi and Germany. Assessments were conducted in different military camps of the Burundian army and lasted about 2 h. Separate barracks were provided for the implementation of the research project by the Burundian army and interviews were conducted individually to prevent any undue influence or the issue of stigma, which could have resulted from group-based assessments. Interviewers entered the participants' responses into the iPad and probed the responses prior to the rating.

All questionnaires and scales were administered on tablet computers, using the software and technical equipment described below. Only written informed consent was collected by paper-pencil mode. Interviews were carried out in a private space between the participant, the clinical interviewer, and if necessary, by a local interpreter. Clinical interviewers had to rate symptoms and responses using the tablet computers. Experienced international and local clinical psychologists, and Burundian psychology students, who had been – just like the interpreters - excessively trained in mental health concepts, were continuously supervised during the assessment period, and carried out the diagnostic interviews. Ongoing intervision and rotating supervisors, which attended the interviews at random, ensured a high quality of the interviews.

Assessment of Posttraumatic Stress Symptoms

The fifth version of the PTSD Symptom Scale Interview [PSS-I; (27)] was administered for the assessment of PTSD symptoms. It is a 20-item interview that assesses each symptom of the DSM-5 during the past month for severity and frequency. However, due to the necessity to keep the results comparable to previous assessments, the response options for each item ranged in accordance with the DSM-IV version on a four-point Likert scale from 0 (not at all) to 3 (five or more times per week/almost always), instead of the newly adapted five-point Likert scale of the DSM-5 version. The PSS-I has proven its validity already in an application with soldiers from the Burundian Army prior to their deployment and in a sample of former Burundian combatants (28). Homogeneity in the present sample was satisfying (Cronbach's Alpha = 0.89). To conduct latent profile analysis (LPA), individual item scores were used. To distinguish class profiles, sub-scores for the DSM B (PSS-I re-experiencing), C (avoidance), D (PSS-I negative changes in cognition and mood) and E (PSS-I increased arousal and reactivity), criteria across classes were compared. Mean PSS-I score in the present study was 4.4 (SD = 6.1); clusters: reexperiencing (M = 1.1, SD = 2.0), avoidance (M = 0.6, SD = 1.0), negative changes in cognition and mood (M = 1.1, SD = 2.1), and increased arousal and reactivity (M = 1.3, SD = 2.1).

Assessment of Depression Symptoms

Depression symptoms were assessed with the Patient Health Questionnaire-9 (PHQ-9), a well validated, and short severity measure of depression [cf. (29)]. The PHQ-9 was originally designed as a self-rating instrument, but has been successfully implemented in clinical interviews as well (14). For the identification of symptom profiles in the LPA, individual items were included in the analysis. However, for the specification of

class characteristics, class differences in the PHQ-9 sum score were analyzed. Therefore, the item scores for the assessment of symptom frequency were summed up ($M = 2.7$, $SD = 3.7$). Cronbach's Alpha for the entire scale was 0.85.

Mobile Devices

The data collection procedure was performed using mobile devices. As a new iOS tablet application with particular characteristics was developed for this study, some aspects of more general interest are shortly discussed. The overall time to develop the mobile application was rather tight (i.e., roughly 8 weeks), therefore an approach from computer science was chosen that is called rapid prototyping [cf. (30)]. However, only using this well-known approach was not sufficient enough in the end, more ideas had to be created and technically carried out to cope with the challenges of the study. As it turned out that the application had to be changed frequently on-site, having in mind that often no internet connection is available and the computer scientists were not present in Africa, a procedure had to be found to transfer application changes from Germany to Burundi. The reasons for these change demands were mainly due to language issues, new interview functions (e.g., feature to quickly jot down notes), or user interface changes. Beside on-site changes, it was challenging to cope with requirements pertaining to the provided procedure how questionnaires are filled out by the psychologists. They wanted features to navigate through questionnaires that required to implement individual features by the computer scientists. In the light of the short implementation time, while preserving validity and integrity of the collected data at the same time, the implementation phase was challenging before the study as well as during the study. As another important aspect, the procedure how data was stored on the used iPads as well as securely transferred for statistical analyzes, was also challenging and required new ideas, procedures, and features. This included implementation efforts as well as training efforts between the psychologists and the computer scientists. Finally, note that the mobile application was not installed to the used iPads using the official App Store from Apple. Instead, the mobile application was directly installed to the iPads; i.e., before the interviewers left Germany to Burundi.

Mobile Data Assessment and Data Security

The collection procedure for the study at hand was accomplished using the aforementioned mobile application. At the time of the study, 3rd generation iPads have been used. During the collection procedure, three aspects were particularly relevant. First, all data must be locally stored on the iPads to properly consider the local circumstances. In addition, data must be locally secured. For this purpose, data was anonymized and encrypted. For encryption purposes, the AES-256 encryption algorithm was used (31). Second, a multi-user feature was implemented to distinguish between interviewers and administrators. The latter were the only entitled persons to decrypt all locally store data. Therefore, administrators had their own area within the mobile application, which was also secured with a password. To secure the data transfer procedure even more strictly, the data transfer was only possible from the iPads to a stationary PC in this administrator area using iTunes. Third, applications adaptions during the data collection procedure became actually necessary. Technically, if adaptions had to be carried out, they were accomplished using an SVN server (32) to which the psychologists in Burundi had remote access. To properly use this access method, the psychologists were taught by the computer scientists in Germany how to deploy a new version of the mobile application. The final collected and anonymized data was secured according to data protection regulations in Germany and stored for 10 years.

Data Analysis

The data analysis was conducted in two steps. First, a latent profile analysis (LPA) (33) was performed to identify subgroups of soldiers based on the PHQ-9 and PSS-I items, accounting for depression and PTSD symptoms simultaneously. The LPA was conducted using Mplus 7 for Mac. A LPA uses latent categorical variables to identify groups of individuals with similar symptom patterns (classes) on a set of clinical variables. In comparison to other statistical approaches that aim to identify groups of participants within a dataset, like a cluster analysis, LPA has several advantages, in particular the "availability of more rigorous empirical criteria for determining the number of clusters" (34). Due to positively skewed, over-dispersed, and non-normally distributed outcome data, a negative binomial model was preferred over linear or Poisson regression models. Applying a zero-inflated negative binomial model to the data was discarded due to unacceptable fit indices. For the appropriate assignment of class labels, one-way analyses of variance (ANOVAs) were conducted with group membership as the independent variable and depression as well as PTSD symptom severity as dependent variables. For the selection of the appropriate number of classes, Lo–Mendell–Rubin-adjusted likelihood ratio tests (LMR-A) as well as bootstrap likelihood ratio test (BLRT) were calculated, indicating the superiority of the final model in comparison to models with a different number of classes [cf. (33)]. Additionally, the Bayesian Information Criterion (BIC) was chosen as a model fit indicator. Analyses were conducted using R statistics, applying a cutoff-level for significance of $p = 0.05$.

RESULTS

Class Assignment by Symptom Profiles

In a first step, latent profile analyses were calculated for two- to eight-classes models, using negative binomial models with automatic starting values and random starts. For the two-class

TABLE 1 | Fit indices for the seven different latent profile analyses.

Modell	Log-likelihood	BIC	Entropy	LMR-A *p*	BLRT *p*
2 classes	−7,828.89	14,184.64	0.914	<0.001	<0.001
3 classes	−6,826.47	13,905.54	0.895	0.232	0.231
4 classes	**−6,591.97**	**13,762.56**	**0.881**	**0.103**	**0.103**
5 classes	−6,428.32	13,818.59	0.888	0.502	0.503
6 classes	−6,396.03	13,981.80	0.891	0.735	0.736
7 classes	−6,352.41	14,173.28	0.901	0.655	0.656
8 classes	−6,306.72	14,258.65	0.911	0.340	0.340

BIC, bayesian information criterion; LMRA-A, Lo-Mendell-Rubin adjusted likelihood ratio test; BLRT, bootstrap likelihood ratio test. The bold values represent the finally selected model.

model, LMR-A and BLRT tests were significant on a $p < 0.05$ level. For the comparison between the three- and four classes model, the LMR-A test did not reach statistical significance. However, the BIC value favored a four classes model. Thus, according to the recommendations by the authors of (10), and in line with the results demonstrating that models with more than five classes did not further improve the model fit, the four-class model was selected for all further analyses on the identification of latent profiles (Cluster 1: $n = 194$, 41.9 %; Cluster 2: $n = 91$, 19.7 %; Cluster 3: $n = 115$, 24.8 %; Cluster 4: $n = 63$, 13.6 %). **Table 1** and **Figure 1** give an overview and illustration of the seven different models' fit indices.

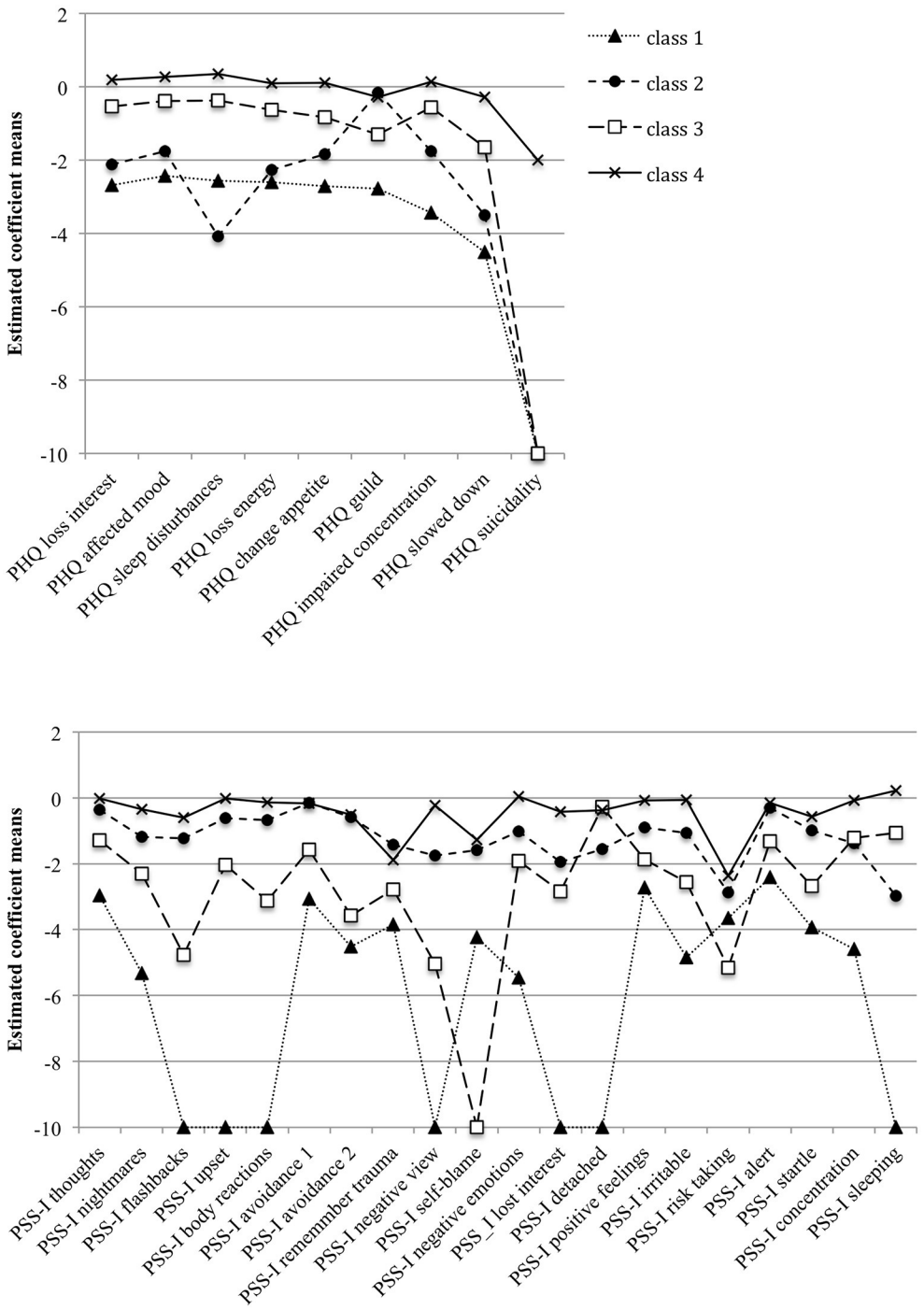


FIGURE 1 | Estimated coefficient means in the PHQ-9 and PSS-I items across the four classes derived from the Latent Profile Analyses. For a better visual presentation, coefficients smaller than−10 were set to−10.

TABLE 2 | Average posterior probabilities for the 4-class model.

class	N	1	2	3	4
class 1	194	**0.95**	0.01	0.04	0.00
class 2	91	0.02	**0.93**	0.04	0.02
class 3	115	0.05	0.04	**0.89**	0.02
class 4	63	0.00	<0.01	<0.01	**0.99**

Posterior probabilities represent the probability that an individual belongs to the respective assigned class. The bold values highlight the correct classification.

Results in **Table 2**, in turn, indicate a number of cases in each of the four classes between 63 and 194 participants. The posterior probabilities that participants belonged to their assigned class ranged between 0.89 and 0.99. Therefore, the model selection produced a meaningful class assignment with four distinguishable classes.

Differences in Psychopathology Between Classes

In a second step (for a better illustration, see **Figure 2**), mean differences between classes were compared using one-way ANOVAs for the (1) PHQ-9 scale [$F(3, 459) = 291.50$, $p < 0.001$, $\eta p2 = 0.66$], the (2) PSS-I re-experiencing scale [$F(3, 459) = 161{,}19$, $p < 0.001$, $\eta p2 = 0.51$], the (3) PSS-I avoidance scale [$F(3, 459) = 115.23$, $p < 0.001$, $\eta p2 = 0.43$], the (4) PSS-I changes in mood and cognition scale [$F(3, 459) = 161.64$, $p < 0.001$, $\eta p2 = 0.51$], and the (5) PSS-I hyper-arousal scale accordingly [$F(3, 459) = 215.04$, $p < 0.001$ $\eta p2 = 0.58$]. Almost all *post-hoc* Tamhane t2 tests for pairwise comparisons were also statistically significant (all $p < 0.001$), except for the difference between (1) PHQ-9: Class 1 vs. Class 2, (2) PSS-I avoidance scale: Class 1 vs. Class 3 and Class 2 vs. Class 4, and (3) PSS-I changes in mood and cognition scale: Class 1 vs. Class 3.

Thus, the results revealed four classes with distinct symptom profiles. Class 1 (low overall symptoms) had the lowest scores on all mental health measures, whereas Class 4 (high overall symptoms) had the highest symptoms scores on both, depression and PTSD scales. Participants in Class 2 (moderate PTSD) showed moderate PTSD symptoms, and in comparison, to the other three groups, low to almost no depression symptoms. Class 3 (moderate depression) was characterized by moderate depression symptoms and low to almost no PTSD symptoms.

DISCUSSION

The results of this study emphasize the feasibility of tablet-assisted clinical interviews assessing mental health symptoms in resource-poor post-conflict regions, such as Burundi. Using a latent class analysis, we identified four different symptom cluster groups amongst Burundian soldiers in a 1-year follow-up after the deployment in AMISOM. Those cluster groups included a low overall symptom profile, a moderate PTSD profile, a moderate depression profile, and within a minority, a high overall-symptom profile.

These findings indicate that mental health symptoms related to depression and PTSD are clustered in a similar way as we would expect amongst soldiers from high-income countries after deployment (35, 39–42). This result could indicate that mental health symptoms as a reaction to traumatic and/or daily stressors might be indeed similar between different cultures (13). The results add to the mounting evidence that mental health concepts and assessment tools developed in high-income cultures can be successfully adapted to different cultural backgrounds (13). These implications would be in line with evidence suggesting that many symptoms of trauma-related disorders and also depressive symptoms result from universal physiological reactions to stress, e.g., the ways traumatic memories are processed within the brain (36, 37).

The majority of the participants had little to no experience with tablets prior to this mental health project, and many of them had little school education. Nevertheless, the use of tablets in the clinical interviews seems not to have affected their willingness to talk about their mental health problems with clinical mental health experts in a way, which provides meaningful results. This conclusion is very promising and might allow researchers and mental health services in resource-poor countries to use mobile technology for meaningful assessments and service provisions. As technology will get more accessible and its' use more usual in every culture in the future, overcoming possible remaining obstacles, such as unfamiliarity with use for self-assessments, and/or lack of alphabetisation, seems very likely.

However, while we could identify four clearly distinct symptom profiles, the mental health symptoms were not that pronounced in individuals of our sample. Taking the mean values of PTSD symptoms ($M = 4.4$) and depression symptoms ($M = 2.7$) into account, we have to acknowledge that we either assessed a highly resilient group of individuals, the majority of whom did not develop severe mental symptoms despite their significant exposure to traumatic stress and violence, or that the soldiers underreported some of their mental health symptoms. Most likely, both of these potential explanations contribute to the low reported symptom scores. The soldiers we assessed remained in the Burundian army after the end of the civil war, when many of those severely affected by injuries had been demobilized. The soldiers continuously took benefit from unit support and relative income stability resulting from their status. Furthermore, they have been less exposed to traumatic experiences than their demobilized colleagues, and might have been particularly adapted to traumatic and violent environments (28). However, mental health problems are also associated with stigma, particularly amongst soldiers. For some of the soldiers, symptoms might be underreported due to the necessity being regarded as strong and functional soldiers, and to avoid any risk of demobilization. Even though, we informed them that no individual information would be passed on to superiors, nevertheless, comprehensibly, a certain mistrust remained. The fact that we successfully identified the four symptoms clusters despite the low symptom Scores, might indicate that those symptoms have been in fact underreported. Hence, sophisticated statistical methods, such as latent class analysis, help to better understand underlying properties of data that would otherwise not be detectable, thereby confirming reliability and validity of the data gathered by the use of mobile technology

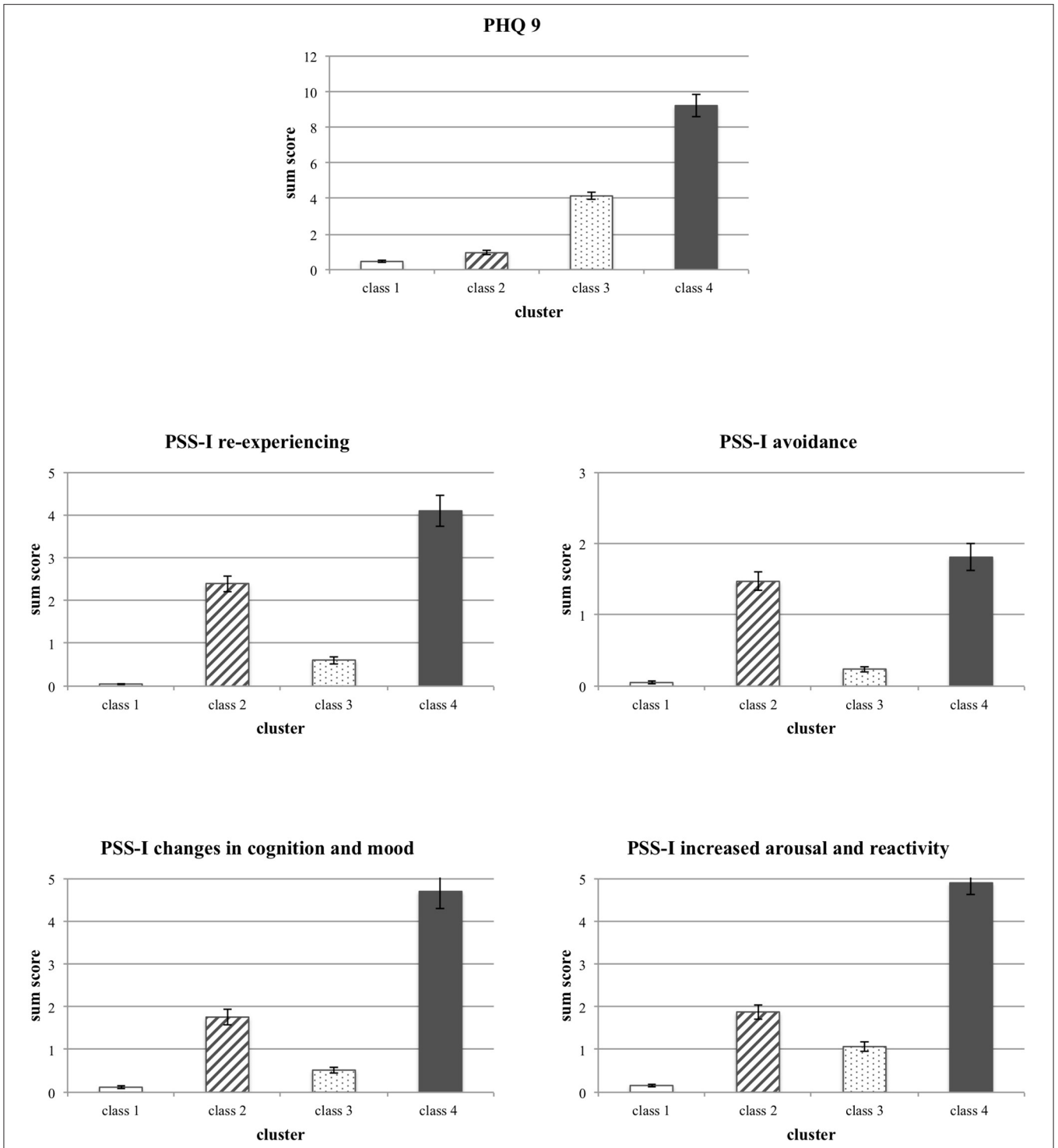


FIGURE 2 | Differences in the mean scores between the four classes of participants for the PHQ-9 sum score as well as the PSS-I sub-scales. Means and Standard Deviations are displayed.

Other analyses methods could be applied to expand the results of the latent class analysis. For example, machine-learning methods could be used to re-evaluate the classes. Based on the number of variables and participants, several machine learning approaches could be a valuable target (e.g., support vector machines). One limitation of this study is that professionals interviewed and entered the data for the soldiers. In consequence, participants might have been less open to report about their mental health symptoms, as research from high-income countries indicates that participants generally tend

to be more honest when providing their answers directly to a digital solution [cf. (38)]. Possibly, this circumstance could be improved in the future with more literate samples, although it remains unclear if the reporting bias identified within high-income countries toward more openness is the same within cultures less exposed to technology when reporting health issues to a machine. Another shortcoming relates to the fact that the same measures were used for identifying and validating the profiles. After a set of PTSD and depression measures is used to empirically determine profiles, using a new set of PTSD and depression measures to validate the profiles would provide a stronger validation approach.

However, it was striking during this study that the mobile application has several benefits compared to a traditional paper-based study. A higher amount of collected data in a rather short time and with higher data quality could be achieved. Regarding the data quality, transcription errors are minimized since a procedure to digitize the data is no longer necessary. In addition, by easily switching between different languages for a questionnaire, the collection procedure could be also improved since translation issues could be mitigated. For example, a psychologist can always toggle between languages if needed, which eases the understanding of questions at hand. Finally, being a challenge from the software engineering perspective, mobile applications that are used for studies must ensure that even when changes are applied to the implementation, the study results must be still valid and comparable. Therefore, data sets must be always tagged with a questionnaire version if substantial changes have been applied to the structure of the questionnaire or the general app implementation.

CONCLUSION

With the present study, it could be demonstrated that mobile technology can enable clinical studies in a new, reliable, and innovative way, especially when studies are carried out in challenging environments. In particular, studies can be conducted in a rather short time with many advantages compared to traditional paper-based studies. For example, by gathering larger amounts of data and with less required resources when using tablet computers or smartphones. In addition, the application of recently emerging analysis methods like machine learning become more easily possible. This study has demonstrated that mobile technology is able to produce data sets, which are valuable and feasible for innovative analysis methods. However, the use of mobile technology also causes challenges that must be considered carefully. As this study showed that the implemented mobile application for the Apple iPad is able to reveal new and valuable research insights in the context of a large-scale study in a resource-poor setting, the general use of mobile technology for clinical studies, especially in challenging environments and with large-scale demands, seems to be promising.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethical Review Boards of the University of Konstanz and the University of Bujumbura, Burundi. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RW, AC, and RP wrote the manuscript. RW, AC, CN, and MB defined the study design, were responsible for the data collection, and led the project. RW and AC conducted the data analysis and the pre-processing. RP developed the technical solution. CN, MB, and TP provided critical feedback on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by the Volkswagen foundation.

ACKNOWLEDGMENTS

We are pleased to acknowledge the valuable contribution and excellent support of the Burundian National Defense Force (FDN). We are most grateful to the Institute of Databases and Information Systems at Ulm University for providing mobile data collection and storage technology. Furthermore, we are grateful to all the Burundian and German students and psychologists who have been involved in this project and without whom interviewing such a large sample of participants would have been impossible.

REFERENCES

1. Carpenter RW, Wycoff AM, Trull TJ. Ambulatory assessment: new adventures in characterizing dynamic processes. *Assessment*. (2016) 23:414–24. doi: 10.1177/1073191116632341
2. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. *Behav Mod*. (2015) 39:785–804. doi: 10.1177/0145445515595198
3. Gruszka P, Burger C, Jensen MP. Optimizing expectations via mobile apps: a new approach for examining and enhancing placebo effects. *Front Psychiatry*. (2019) 10:365. doi: 10.3389/fpsyt.2019.00365
4. Lui JH, Marcus DK, Barry CT. Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. *Pro Psychol Res Pra*. (2017) 48:199. doi: 10.1037/pro0000122
5. Schobel J, Pryss R, Probst T, Schlee W, Schickler M, Reichert M. Learnability of a configurator empowering end users to create mobile

data collection instruments: Usability study. *JMIR Mhealth Uhealth.* (2018) 6:e148. doi: 10.2196/mhealth.9826
6. Areán PA, Ly KH, Andersson G. Mobile technology for mental health assessment. *Dial Clin Neuro.* (2016) 18:163. doi: 10.31887/DCNS.2016.18.2/parean
7. Schobel J, Pryss R, Schickler M, Reichert M. Towards Flexible Mobile Data Collection in Healthcare. In: *29th IEEE International Symposium on Computer-Based Medical Systems (CBMS 2016)* Dublin: IEEE (2016). p. 181–82. doi: 10.1109/CBMS.2016.43
8. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ.* (2016) 352:i1174. doi: 10.1136/bmj.i1174
9. Mercader HFG, Kabakyenga J, Katuruba DT, Hobbs AJ, Brenner JL. Female respondent acceptance of computer-assisted personal interviewing (CAPI) for maternal, newborn and child health coverage surveys in rural Uganda. *Int J Med Inf.* (2017) 98:41–6. doi: 10.1016/j.ijmedinf.2016.11.009
10. Ndayizigamiye P, Maharaj M. Determinants of mobile health adoption in burundi. *Afri J Inf Syst.* (2017) 9:4.
11. Rusatira JC, Tomaszewski B, Dusabejambo V, Ndayiragije V, Gonsalves S, Sawant A, et al. Enabling access to medical and health education in rwanda using mobile technology: needs assessment for the development of mobile medical educator apps. *JMIR Med Edu.* (2016) 2:e7. doi: 10.2196/mededu.5336
12. Heavey SC, Homish DL, Goodell EA, Homish GG. US reserve soldiers' combat exposure and intimate partner violence: not more common but it is more violent. *Stress Health.* (2017) 33:617–23. doi: 10.1002/smi.2748
13. Weiss MG, Saraceno B, Saxena S, van Ommeren M. Mental health in the aftermath of disasters: consensus and controversy. *J Nerv Ment Dis.* (2003) 191:611–5. doi: 10.1097/01.nmd.0000087188.96516.a3
14. Wulsin L, Somoza E, Heck J. The feasibility of using the Spanish PHQ-9 to screen for depression in primary care in Honduras. *Prim Care Com J Clin Psychiatry.* (2002) 4:191–5. doi: 10.4088/PCC.v04n0504
15. Bonde JP et al. Risk of depressive disorder following disasters and military deployment: systematic review with meta-analysis. *Bri J Psychiatry.* (2016) 208:330–6. doi: 10.1192/bjp.bp.114.157859
16. Eekhout I, Reijnen A, Vermetten E, Geuze E. Post-traumatic stress symptoms 5 years after military deployment to afghanistan: an observational cohort study. *Lancet Psychiatry.* (2016) 3:58–64. doi: 10.1016/S2215-0366(15)00368-5
17. Nandi C, Crombach A, Elbert T, Bambonye M, Pryss R, Schobel J, et al. The cycle of violence as a function of PTSD and appetitive aggression: a longitudinal study with burundian soldiers. *Aggr Behav.* (2020) 46:391–9. doi: 10.1002/ab.21895
18. Ginzburg K, Ein-Dor T, Solomon Z. Comorbidity of posttraumatic stress disorder, anxiety and depression: a 20-year longitudinal study of war veterans. *J Aff Dis.* (2010) 123:249–57. doi: 10.1016/j.jad.2009.08.006
19. Williams PD. Fighting for peace in somalia: AMISOM's seven strategic challenges. *J Int Peac.* (2013) 17:222–47. doi: 10.1163/18754112-1704004
20. Williams PD. Stabilising somalia: the african union mission and the next stage in the war against Al-Shabaab. *RUSI J.* (2014) 159:52–60. doi: 10.1080/03071847.2014.912803
21. Bundervoet T, Verwimp P, Akresh R. Health and civil war in rural burundi. *J Hum Res.* (2009) 66–88. doi: 10.1596/1813-9450-4500
22. Daley P. Ethnicity and political violence in Africa: the challenge to the burundi state. *Pol Geog.* (2006) 25:657–79. doi: 10.1016/j.polgeo.2006.05.007
23. Daley P. Unearthing the local: hegemony and peace discourses in central Africa. *Geog Peace.* (2014) 66:66–86. doi: 10.5040/9780755619900.ch-004
24. Uvin P. *Life after Violence. A peoples Story of Burundi*. New York, NY: Zed Books (2009).
25. Bhuyan SS, Kim H, Isehunwa OO, Kumar N, Bhatt J, Wyant DK, et al. Privacy and security issues in mobile health: current research and future directions. *Health Policy Technol.* (2017) 6:188–91. doi: 10.1016/j.hlpt.2017.01.004
26. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine.* (2000) 25:3186–91. doi: 10.1097/00007632-200012150-00014
27. Foa E, Capaldi SI. *Manual for the Administration and Scoring of the PTSD Symptomscale–Interview for DSM-5 (PSS-I-5).* (2013). Available online at: https://www.ptsd.va.gov/professional/assessment/adult-int/pss-i.asp (accessed June 4, 2021).
28. Nandi C, Crombach A, Bambonye M, Elbert T, Weierstall R. Predictors of post-traumatic stress and appetitive aggression in active soldiers and former combatants. *Eur J Psychotraumatol.* (2015) 6:26553. doi: 10.3402/ejpt.v6.26553
29. Kroenke K, Spitzer RL, Williams JB. The Phq-9. *J Gen Int Med.* (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x
30. Najjar LJ. *Rapid Prototyping (TR 52.0020).* (1990). Available online at: http://www.lawrence-najjar.com/papers/Rapid_prototyping.html (accessed September 12, 2012).
31. Biryukov A, Dunkelman O, Keller N, Khovratovich D, Shamir A. Key recovery attacks of practical complexity on AES-256 variants with up to 10 rounds. In: Gilbert H, editor. *Advances in Cryptology – EUROCRYPT 2010.* Berlin; Heidelberg: Springer. (2010). p. 299–319.
32. Collins-Sussman B, Fitzpatrick BW, Pilato CM. *Version Control with Subversion*. (2008). p. 299.
33. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte carlo simulation study. *Str Equ Mod.* (2007) 14:535–69. doi: 10.1080/10705510701575396
34. Meyer JP, Stanley LJ, Vandenberg RJ. A person-centered approach to the study of commitment. *Hum Res Manag Rev.* (2013) 23:190–202. doi: 10.1016/j.hrmr.2012.07.007
35. MacManus DK, Dean M, Jones RJ, Rona N, Greenberg L, Hull T et al. Violent offending by UK military personnel deployed to iraq and afghanistan: a data linkage cohort study. *Lancet.* (2013) 381:907–17. doi: 10.1016/S0140-6736(13)60354-2
36. Brewin CR, Gregory JD, Lipton M, Burgess N. Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications. *Psychol Rev.* (2010) 117:210–32. doi: 10.1037/a0018113
37. Elbert T, Schauer M. Burnt into memory. *Nature.* (2002) 419:883. doi: 10.1038/419883a
38. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views JMIR ment health. (2017) 4:e52 doi: 10.2196/mental.8311
39. Wilk JE, Quartana PJ, Clarke-Walper K, Kok BC, Riviere LA. Aggression in US soldiers post-deployment: associations with combat exposure and PTSD and the moderating role of trait anger. *Aggr Behav.* (2015) 41:556–65. doi: 10.1002/ab.21595
40. Thomas JL, Wilk JE, Riviere LA, McGurk D, Castro CA, Hoge CW. Prevalence of mental health problems and functional impairment among active component and national guard soldiers 3 and 12 months following combat in Iraq. *Archiv Gen Psychiatry.* (2010) 67:614–23. doi: 10.1001/archgenpsychiatry.2010.54
41. Nichter B, Norman S, Haller M, Pietrzak RH. Psychological burden of PTSD, depression, and their comorbidity in the US veteran population: suicidality, functioning, and service utilization. *J Aff Dis.* (2019) 633–40. doi: 10.1016/j.jad.2019.06.072
42. Milliken CS, Auchterlonie JL, Hoge CW. Longitudinal assessment of mental health problems among active and reserve component soldiers returning from the Iraq war. *JAMA.* (2007) 298:2141–8. doi: 10.1001/jama.298.18.2141

Conflict of Interest: The authors declare that this study received funding from the Volkswagen Foundation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership