Full-Length Research Report

Behavioral Activation by Wearable Devices in Patients with Late-Life Depression

A Randomized Crossover Study

Alexandra Wuttke[1,2], Annika Steinmetz[3], Kristina Endres[3], Perikles Simon[4], Andreas Fellgiebel[1,3,5], and Nils Haller[4,6]

[1]Center for Mental Health in Old Age, Landeskrankenhaus (AöR), Mainz, Germany
[2]Department of Psychiatry, Psychosomatics and Psychotherapy, Center of Mental Health, University Hospital Würzburg, Germany
[3]Department of Psychiatry and Psychotherapy, University Medical Center, Johannes Gutenberg-University Mainz, Germany
[4]Department of Sports Medicine, Rehabilitation and Disease Prevention, Johannes Gutenberg University of Mainz, Germany
[5]Hospital for Psychiatry, Psychosomatic and Psychotherapy, Agaplesion Elisabethenstift, Darmstadt, Germany
[6]Department of Sport and Exercise Science, University of Salzburg, Austria

Abstract. *Background*: We examine the potential of behavioral activation by wearable devices in inpatients with late-life depression (LLD). *Methods*: In randomized order, 17 inpatients with LLD wore a GPS watch that either allowed monitoring their activity level or had a disabled activity display. The patients answered questionnaires repeatedly in addition to completing an ambulatory assessment of biopsychological stress. *Results*: Acceptance and satisfaction were high. We found short-term differential effects of the randomization order, with patients first in the control group being more active but also more stressed. A high number of missing values limits the informative value of ambulatory assessment. *Conclusions*: The intervention was feasible and well accepted. However, the effect of continuous feedback on activity levels in a crossover design needs to be critically reassessed.

Keywords: aging, antidepressant therapy, physical activity tracker, biofeedback, e-mental health

Mental disorders are highly prevalent in older adults (aged 65–84 years), with an estimated one in three suffering from a mental illness in the past 12 months (Andreas et al., 2017). A recent cross-sectional study in four European regions reported an overall prevalence of late-life depression (LLD) of 29% (Horackova et al., 2019). Chronic conditions, pain, limitations in daily living, and cognitive impairment were found to be strongly associated with LLD (Horackova et al., 2019). The main treatment options for LLD are psychotherapy and/or antidepressants. However, antidepressants are less effective in older patients than in younger adults (Alexopoulos, 2019), and psychotherapy is rarely prescribed in this population, with older patients being underrepresented in outpatient psychotherapy (Kessler & Blachetta, 2020). Therefore, it is essential to develop easily accessible and effective treatment strategies that patients accept.

Recently, there has been increasing interest in exercise and/or physical activity (PA) as a safe adjunct or alternative treatment option for major depressive disorder (MDD). Recent meta-analyses have highlighted the efficacy of exercise for depressive symptoms in general (Schuch et al., 2016) and of individualized exercise for LLD in particular (Bridle et al., 2012). Current guidelines recommend exercising for 45–60 minutes three times a week (National Collaborating Centre for Mental Health, 2010). However, older patients face several barriers that prevent them from engaging in such high levels of PA, including low self-esteem, loss of energy, interest, and motivation, and fatigue (Catalan-Matamoros et al., 2016). Moreover, MDD patients generally tend to have a low fitness level and a low level of PA, which may be exacerbated in older adults (Boettger et al., 2009). Thus, even performing daily activities such as cleaning or climbing stairs could become difficult for these patients (Haller & Simon, 2017).

Accordingly, it is important to promote PA in this patient group to safely increase their amount of activity over time. Patients with low levels of regular PA might therefore be initially activated by setting simple goals, such as a daily step target. In this regard, Catalan-Matamoros et al. (2016)

formulated requirements to optimize the impact of PA interventions, such as goal-setting, formulation of realistic goals, and regular feedback. These requirements can be met and monitored with the help of *wearable technology* (i.e., *wearables*, defined as "devices that incorporate wireless connectivity for the purposes of seamlessly accessing, interacting with and exchanging contextually relevant information" [Bower & Sturman, 2015]), which have proven their potential to facilitate improvements in health behaviors. Such interventions are broadly applicable, and as highlighted by Blake et al. (2009), they might be more cost-effective than traditional therapeutic approaches or exercise interventions (Lyons et al., 2014). However, according to the technology acceptance model, it is crucial that wearables be easy to use and perceived as useful, which promotes their actual and effective use (Al-Maroof et al., 2021; Marangunić & Granić, 2015). This seems to be particularly relevant in older adults who might be less technically experienced with wearables than younger adults. It is essential to understand the feasibility of wearables in this population in order to develop appropriate interventions.

To date, however, research in patient cohorts in this area has been limited. A recent review underlined the potential of wearables to promote PA (Strath & Rowley, 2018). In another study, wearables proved to have the potential to increase daily activity in patients with serious mental illnesses over several months (Naslund et al., 2016). In a 28-week qualitative study, Chum et al. (2017) demonstrated that the use of Fitbit wearables was perceived as helpful for increasing PA in MDD patients. Interestingly, baseline depression or age did not affect the perceived benefit of the device. To the best of our knowledge, no study has yet examined the effects of wearable use on depressive symptoms in LLD with additional elements of biofeedback in a therapeutic context. The present study, therefore, aimed at exploratorily investigating the feasibility, acceptability, and efficacy of a GPS (global positioning system) PA watch in a patient cohort with LLD, with the goal of improving PA and reducing depressive symptoms.

Furthermore, the mechanisms underlying the effect of PA in patients with LLD remain unclear. According to the rationale of the diathesis-stress model, psychobiological stress increases the vulnerability to developing diseases. In this regard, patients with LLD often present with hypercortisolism, reflecting disturbances in the hypothalamic-pituitary-adrenal (HPA) axis (Tiemeier, 2003). Higher secretion of cortisol – a stress hormone reflecting the activity of the HPA axis – has serious health implications, as cortisol secretion is a risk factor for further health deterioration, for example, neurodegeneration and executive dysfunction (Shindel et al., 2020). Moreover, dysregulations in the autonomic nervous system (ANS) have been reported, with patients with LLD showing reduced heart rate variability (Brown et al., 2018). Changes in the stress-sensitive systems in the body (HPA axis, ANS, immune system) have been associated with treatment trajectories in LLD, in line with the allostatic load theory of stress (Andreescu et al., 2019). At the same time, PA is also associated with stress-reducing effects. Movement-based exercise interventions have attracted increasing interest, particularly in the prevention and treatment of LLD (Laird et al., 2019), and there is growing evidence that these interventions exert a beneficial effect on physiological stress markers as well (Toni et al., 2016). We, therefore, set out to additionally evaluate the feasibility of psychobiological stress markers in patients with LLD.

Methods

Participants

All participants were patients with confirmed LLD who had been recruited at the beginning of their inpatient stay in a gerontopsychiatric ward. The final sample for analysis consisted of 17 patients (8 females; mean age: 67.5 ± 6.5 years). At admission, the mean Geriatric Depression Scale (GDS-15) score was 8.13 ± 2.94, and the mean Quick Inventory of Depressive Symptomatology (QIDS-16) score was 15.25 ± 4.70. The order of conditions was randomized, with 10 patients participating in the intervention group first and seven patients participating in the control group first.

Procedure

We designed this multidisciplinary single-center crossover study as an exploratory feasibility study. Recruitment took place in the gerontopsychiatric ward of the Rheinhessen-Fachklinik Alzey, Germany, between May 2018 and May 2019. Upon admission to the clinic, all patients with a suspected or confirmed primary diagnosis of any depressive disorder (according to the ICD-10) received an information bulletin and were asked to participate in the study. Patients who expressed an interest in participating were given further information in a personal interview, and the following study inclusion criteria were reviewed: primary diagnosis of depression (F32, F33, or F34 according to the ICD-10), age $\geq$ 60 years, being mobile, inpatient or day-clinic treatment for a minimum of 4 weeks, no acute suicidality, no psychotic symptoms, no comorbid substance abuse or substance dependence, no cognitive impairment, no medication with immunological or endocrine effects (except for antidepressants), BMI between 18 and 35 kg/m^2. All participants provided written informed consent prior to study participation.

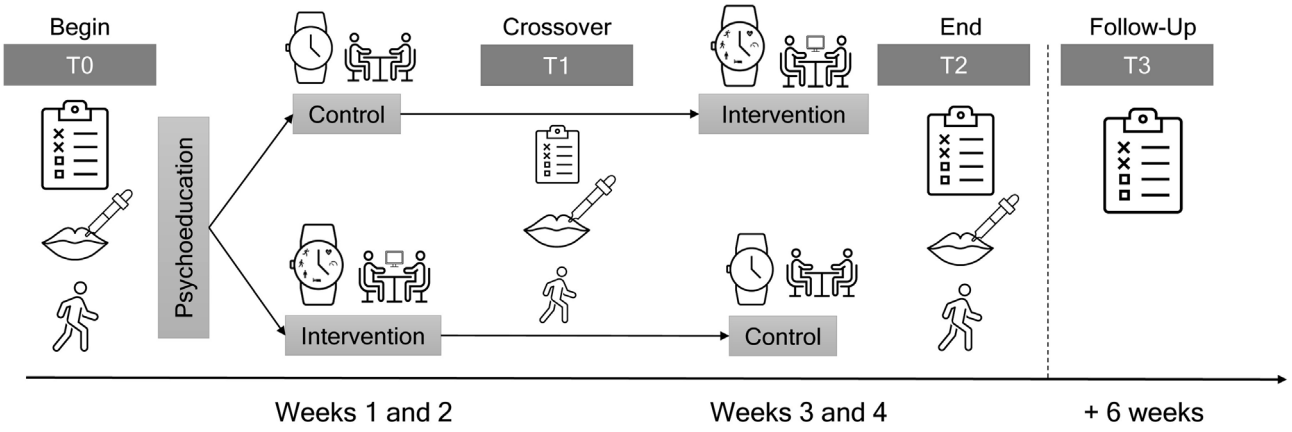

Figure 1. Graphic illustration of the crossover study design.

Figure 1 illustrates the study procedure. The baseline assessment (T0) comprised paper-and-pencil questionnaires, physical and cognitive assessments, and 2 days of ambulatory assessment. After completing this baseline assessment, all participants received individual psychoeducation on the relationship between activity and depression following the rationale of cognitive-behavioral therapy (Hautzinger, 2000). As part of the psychotherapeutic approach, all patients completed a weekly diary, which was used to plan activities. Next, using the research randomizer tool (https://www.randomizer.org/), we randomized patients to start either in the intervention or the control group. In both groups, patients received a GPS PA watch (Polar M430). In the intervention group, patients could use all functions and directly review their activity level; in the control group, patients could only use the GPS watch as a regular watch, and all functions for checking their own activity were disabled. Both groups received twice-weekly therapy sessions of 25 minutes each, during which goals for behavioral activation were reflected upon and set. In the intervention group, objective data on physical activity were evaluated and interpreted with respect to changes in mood. In the control group, activity and mood were reflected without reference to the objective activity data. Thus, the main variation between the control and intervention groups concerns features of the GPS watch and the availability of objective data on the previous activity level, which could be accessed in the intervention group but not in the control group. We kept the psychotherapeutic elements constant between the two groups, the only exception being that behavioral activation was supported by the objective data of the GPS watch in the intervention group, whereas it was based on the subjective perception of the patients during the control group.

After 2 weeks, a crossover took place (T1). The patients completed the paper-and-pencil questionnaires again, repeated the physical assessment, and completed an ambulatory assessment for 2 days. After completing a further 2 weeks of study participation, participants filled out the paper-and-pencil questionnaires a third time, in addition to filling out a self-generated questionnaire on the feasibility and acceptability of the intervention, they repeated the physical and cognitive assessments, after which 2 days of ambulatory assessment took place (T2). Finally, 6 weeks later, the patients completed the paper-and-pencil questionnaires again back at home (T3).

All procedures were approved by the Ethics Committee of the State Chamber of Physicians of Rhineland-Palatinate, Mainz, Germany. Participation in the study did not affect previous and ongoing treatments.

Measures

Feasibility

We analyzed the feasibility of the intervention by evaluating dropouts and missing values concerning the outcome measures. We defined patient dropout when patients completed less than four sessions and thus did not complete one phase. To determine missing values, we evaluated how many therapy sessions the patients had attended, the number of missing values on the paper-and-pencil questionnaires and the number of missing values concerning the biological stress markers.

Acceptability

We analyzed the acceptability of the intervention using a self-designed post-monitoring interview. At the end of the intervention as well as at 6 weeks post-intervention, all participants who had completed the intervention rated self-generated items regarding their satisfaction with the intervention on a scale ranging from 0 (*not at all*) to 3 (*completely*). The items at the end of the intervention covered the overall evaluation of the study treatment, the acceptability of providing saliva samples, and the presumed

effects of the intervention on mood, activity, and general well-being. At follow-up, these items were slightly reformulated and focused additionally on the long-term influence on life from participation. All items assessed at the end of the intervention are presented in the Results section. Concerning the follow-up interview (six weeks later), only those items covering long-term changes are presented.

Efficacy

We evaluated the efficacy of the intervention using a multidimensional approach, combining subjective with physiological markers both in the clinical setting and in patients' daily life: (a) psychometric test battery, (b) continuous assessment of PA, (c) physical and cognitive assessment, and (d) ambulatory assessment.

Psychometric Test Battery

At all four measurement time points (baseline, crossover, end, follow-up), participants completed psychometric scales concerning the following constructs: depressive symptoms using a self-report, the QIDS-16 (Rush et al., 2003) and the GDS-15 (Sheikh & Yesavage, 1986); subjective stress with the German version of the Perceived Stress Scale (PSS-10; Cohen, 1988) and the Screening Scale for Chronic Stress (SSCS) from the Trier Inventory for the Assessment of Chronic Stress (TICS; Schulz et al., 2004). Further, we used the General Self-Efficacy Scale (GSE; Schwarzer & Jerusalem, 2003).

Continuous Assessment of PA

To monitor patients' PA, we used the Polar M430 GPS watch. The display settings varied between the two conditions (see Figure 1). In the intervention group, participants could instantly monitor their level of PA. They received visual and haptic feedback as soon as their activity goal had been achieved. Furthermore, if they were inactive for too long, visual and haptic feedback appeared, stating, "It's time to move" (inactivity alert). In the control condition, the display settings merely showed the time of day, and all inactivity alerts were muted (Figure 1). The following parameters were analyzed each day using the algorithms of the manufacturer: activity time, inactivity markers, activity goal (0–100%), energy expenditure (in kilocalories), steps, and time spent lying down, sitting, standing, walking, and running.

Physical and Cognitive Assessment

To assess the patients' physical fitness, we conducted a 6-minute walk test (6mwt). In this test, patients walk as far as they can on a predetermined course for a duration of 6 minutes. The primary outcome measure was the total distance walked (Enright, 2003).

To evaluate patients' cognitive state, we asked them to complete the two parts of the Trail Making Test (TMT-A, TMT-B; Reitan & Wolfson, 1995) for a brief examination of processing speed (TMT-A) and executive functioning (TMT-B).

Ambulatory Assessment

As part of the ambulatory assessment, we measured psychobiological stress markers 6 times per day for 2 consecutive days at the beginning, crossover, and end of the intervention. In particular, directly after awakening, 30 minutes after awakening, and at 10 am, 2 pm, 6 pm, and 9 pm, patients were asked to collect a saliva sample for the later analysis of salivary cortisol (sCort) and salivary alpha-amylase (sAA) and to subjectively rate their momentary perceived stress on a visual analog scale (VAS) from 0 (*not at all*) to 100 (*extremely*). Participants were required to transfer accumulated saliva into prelabeled vials, which were then stored in a study freezer located at the inpatient clinic. We measured sCort levels using a commercially available enzyme-linked immunoassay (IBL International, a Tecan Group Company, Hamburg, Germany). sAA activity was measured using a kinetic colorimetric test and reagents obtained from DiaSys Diagnostic Systems (Holzheim, Germany). The interassay variance was 14.18% for sAA and 4.5% for sCort, and the intraassay variance was 8.57% for sAA and 1.69% for sCort.

Statistical Analysis

We performed the statistical analyses using SPSS (IBM SPSS Statistics for Windows, Version 23.0, IBM Corporation, Armonk, NY). Differences in the mean scores on the paper-and-pencil questionnaires over time were tested for significance using paired *t*-tests, comparing baseline to crossover (T0-T1), crossover to end (T1-T2), and end to follow-up (T2-T3). Missing values were excluded listwise. We further controlled for the order of randomization and tested whether the changes in the scores on paper-and-pencil questionnaires differed over time depending on the order of randomization. To compare the mean values at T0, T1, and T2, we conducted repeated measures ANOVA, with order of randomization as a between-subject factor. We excluded the 6-week follow-up assessment (T3) because of the high number of missing values. We tested differences in the cognitive assessment using paired *t*-tests, comparing the time required to complete the TMT-A and TMT-B between the beginning and end of the intervention. We tested differences in the 6mwt over time using repeated measures ANOVA, controlling for the order of randomization. We assessed differences in activity levels depending on the control or intervention group using MANOVA, with condition (control vs. intervention group) and order of randomization (intervention group first, control group first) and their interaction as fixed factors. We considered *p*-values < .05 as statistically significant.

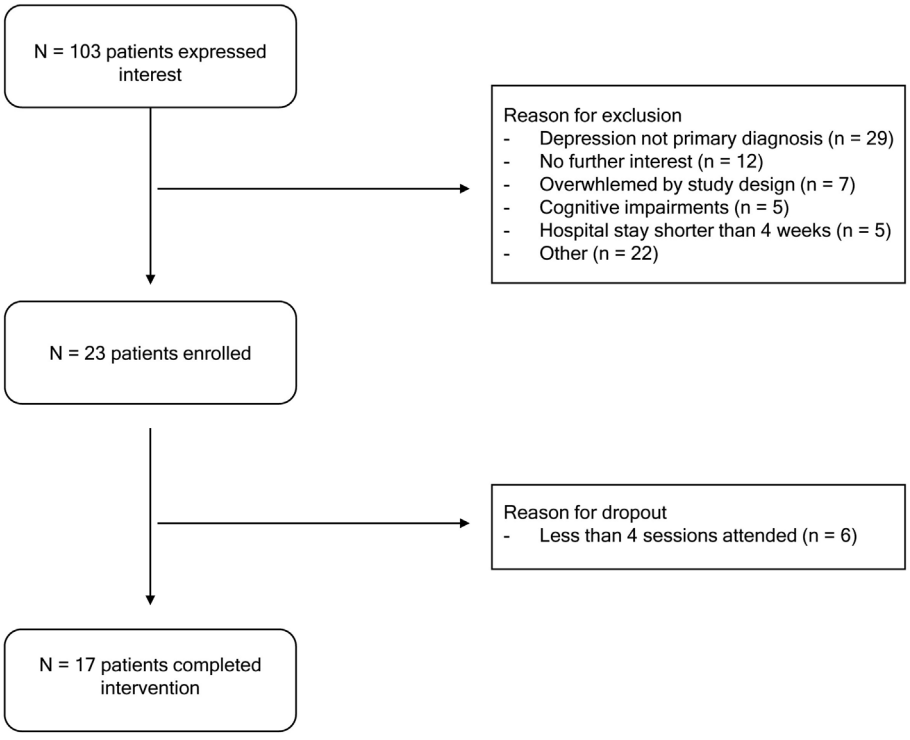

Figure 2. Flow chart diagram.

Results

Feasibility

103 patients expressed interest in study participation after receiving the information bulletin upon admission to the clinic (see Figure 2), of whom 23 were finally enrolled in the study, resulting in an inclusion rate of 22%. Reasons for non-participation were (in descending order): depression not primary diagnosis (29), no interest (12), overwhelmed by study design (7), cognitive impairments (5), short hospital stay (5), and other (22).

Data from 17 of the 23 participants were entered into the analyses, as 6 patients dropped out of the study at less than four therapy sessions, resulting in a dropout rate of 26%. The reasons for dropping out were either earlier discharge from the hospital or transfer to another hospital. Ten patients completed all eight sessions, 2 completed seven sessions, 3 completed five sessions, and 2 completed only four sessions, resulting in a mean attendance rate of 6.88 sessions per patient.

The missing values varied between paper-and-pencil questionnaires and ambulatory assessment data. Concerning the paper-and-pencil questionnaires, between 29.4% and 30.9% of data were missing. The proportion of missing values in the ambulatory assessment was high, with 38.4% of missing data for sCort, 42.3% of missing data for sAA, and 45.4% of missing data for the VAS. Overall, four participants, amounting to 23.5% of the sample, failed to provide any saliva samples.

Acceptability

We rated the acceptability of the intervention directly at T2 (end of intervention) and at T3 (6 weeks post-intervention). A subsample of n = 12 (70.6%) patients completed this questionnaire at T2, and a subsample of n = 7 (41.2%) at T3. The overall acceptance was high. Table 1 (end of intervention) and Table 2 (follow-up) show the results.

Exploratory Efficacy Analysis

Scores on the GDS-15 and SSCS remained stable over time in the total sample. We found significant differences for the PSS-10 and GSE. In particular, the GSE sum score increased from T1 (crossover) to T2 (end of intervention) (t = −2.300, df = 10, p = .044), and the PSS-10 sum score decreased from T2 (end of intervention) to T3 (follow-up) (t = 3.321, df = 5, p = .021). The complete statistics on the pairwise comparisons may be found in the Electronic Supplementary Material (ESM 1). Concerning the order of randomization, there were significant differences in the stress measures PSS-10 and SSCS, as these variables interacted significantly with the order of randomization (PSS × order of randomization: F = 8.839, p = .003;

Table 1. Acceptability of intervention directly at the end of the intervention

Item	*M* ± *SD*	Min	Max	*N*
The study treatment was interesting.	2.75 ± 0.45	2	3	12
I am satisfied with the treatment.	2.75 ± 0.45	2	3	12
The study led to positive changes in daily routine.	2.83 ± 0.39	2	3	12
I found it easy to use the watch.	2.50 ± 0.80	1	3	12
Participation led to negative events.	0.50 ± 0.91	0	3	12
I experienced taking saliva samples as annoying.	1.67 ± 0.99	0	3	12
Participation in the study has led to increased physical activity.	2.67 ± 0.49	2	3	12
The saliva samples were requested too frequently.	1.42 ± 1.00	0	3	12
I feel better than before the treatment.	2.25 ± 0.75	1	3	12
There have been positive effects on my mood.	2.17 ± 0.72	1	3	12
I feel I can better influence the course of my disease.	2.17 ± 0.72	1	3	12
I experience less stress.	1.83 ± 0.72	1	3	12
Study treatment led to increased self-observation, and I behaved differently than usual.	2.42 ± 0.52	2	3	12
I found it difficult to complete the diaries.	1.50 ± 0.91	0	3	12
Something in my life has changed for the better.	2.25 ± 0.62	1	3	12
I would also have agreed to take saliva samples more often.	0.92 ± 0.79	0	2	12
The analyses of activity data were easy to understand.	2.50 ± 0.52	2	3	12
Therapy sessions were useful and filled with interesting content.	2.67 ± 0.49	2	3	12
The time commitment was too high.	0.50 ± 0.67	0	2	12
I have experienced positive changes in my activity and mood.	2.08 ± 0.79	1	3	12

Note. All statements were rated on a scale ranging from 0 (*not at all*) to 3 (*completely*). *M* = mean; *SD* = standard deviation; Min = minimum; Max = maximum; *N* = sample size.

Table 2. Acceptability of intervention at follow-up 6 weeks after the end of the intervention

Item description	*M* ± *SD*	Min	Max	*N*
Participation also led to positive changes in my daily routine in the long term.	2.29 ± 0.76	1	3	7
Participation led to negative events in the long term.	0 ± 0	0	0	7
I feel better than before the treatment.	2.14 ± 0.69	1	3	7
Today, I still try to implement what I learned.	2.29 ± 0.76	1	3	7
I always try to be active to improve my mood.	2.43 ± 0.54	2	3	7
I still experience less stress today.	2.14 ± 0.38	2	3	7
Today, I still complete an activity and mood log.	0.57 ± 0.54	0	1	7
Because of the study treatment, I experience more joy in movement.	2.00 ± 0.58	1	3	7
I often remember the content of the sessions and try to be more active in my daily life.	1.86 ± 0.38	1	2	7
I have noticed positive changes in my activity and mood.	2.00 ± 0.00	2	2	7

Note. All statements were rated on a scale ranging from 0 (*not at all*) to 3 (*completely*). *M* = mean; *SD* = standard deviation; Min = minimum; Max = maximum; *N* = sample size.

SSCS × order of randomization: F = 4.822, p = .026). Patients who participated first in the intervention group experienced a stress reduction from baseline to end, while patients who participated first in the control group experienced increases in stress during this time frame. The stress ratings converged at the 6-week follow-up (T3). These interaction terms were insignificant concerning the GDS-15 and GSE (GDS-15 × order of randomization: $F = 2.694$, $p = .102$; SWE × order of randomization: F = 0.160, p = .853). Figure 3 presents a graphical illustration of the interaction terms.

Notably, the average number of steps per day was 11,792 ± 4152 in the total sample, irrespective of condition and phase. The objectively assessed activity level did not change significantly between the control and intervention condition (all $p \geq .113$, see Table 3), although a main effect of the order of randomization emerged for steps ($F = 14.695$, $p \leq .001$) and time spent sitting ($F = 8.754$, $p = .004$). Moreover, there was a positive interaction of condition × order of randomization regarding time spent lying down (F = 5.120, p = .026). These findings can be interpreted such that the number of steps was higher and the time spent sitting was

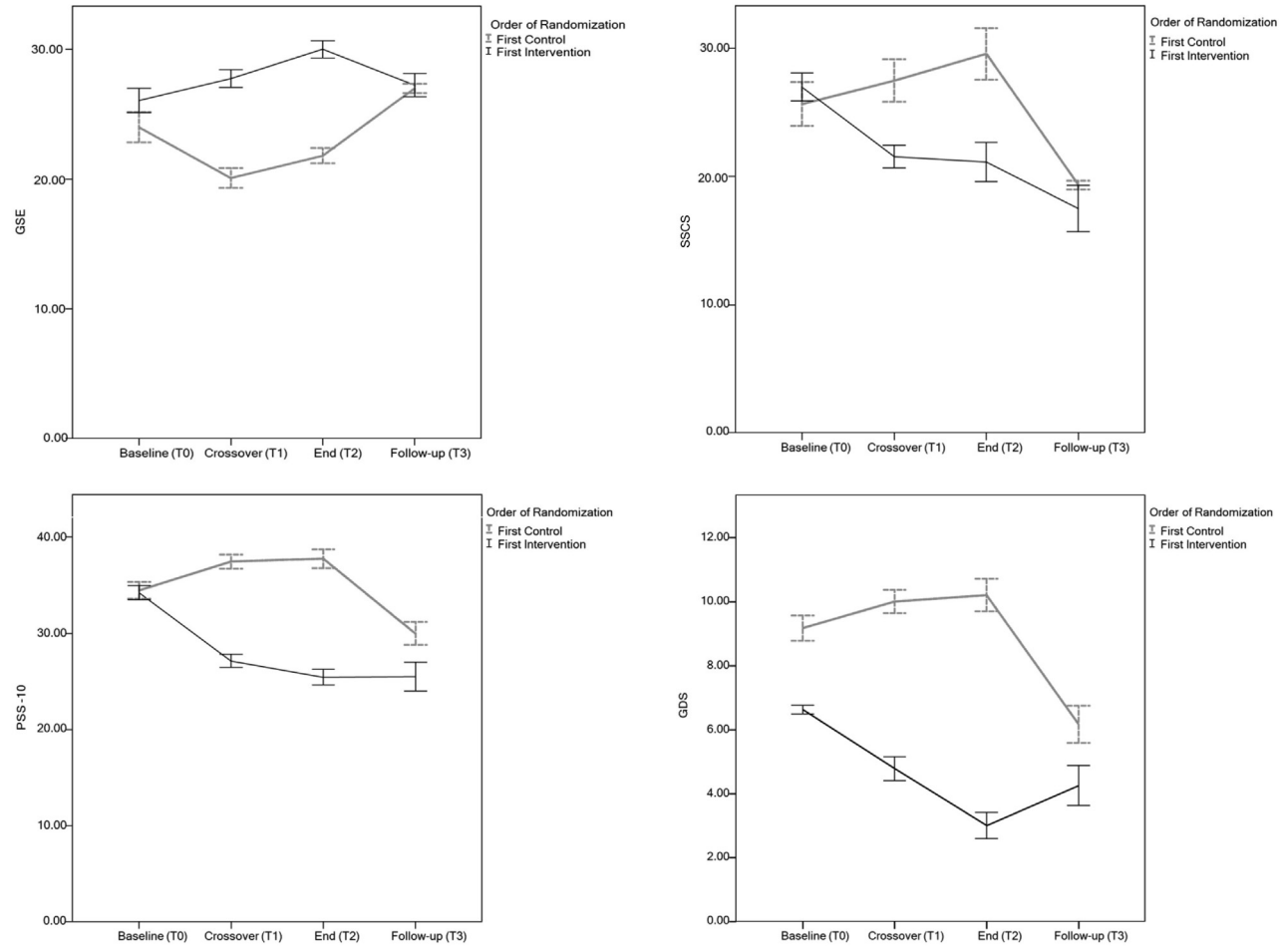

Figure 3. Changes in paper-and-pencil questionnaires over time relative to randomization order. GSE: General Self-Efficacy Scale (upper left), SSCS: Screening Scale for Chronic Stress, taken from the Trier Inventory of Chronic Stress (upper right), PSS-10: Perceived Stress Scale, PSRS (bottom left), GDS: Geriatric Depression Scale (GDS-15; bottom right), Error bars: ± standard error.

lower in the group of patients who underwent the control condition first and did not receive feedback on their activity level. The amount of time spent lying down was lowest in the control condition for patients who underwent the control condition first and was highest when these patients changed to the intervention.

The physical assessment and cognitive functioning improved descriptively in the total sample from T0 to T2. Regarding the 6mwt, the distance covered increased from 433 ± 134 m (T0) to 451 ± 149 m (T1) to 531 ± 132 m (T2), although this did not reach statistical significance (F = 4.306, p = .069). Likewise, the time taken to complete the TMT-A and TMT-B decreased in the total sample from T0 to T2 (TMT-A: 46.96 ± 22.61 s at T0; 37.28 ± 11.90 s at T2; TMT-B: 106.7 ± 42.39 s at T0, 83.11 ± 45.29 s at T2). Although these differences did not reach statistical significance (TMT-A: $T(11)$ = 1.770, p = .104; TMT-B: $T(10)$ = .196, p = .848), they are clinically relevant when compared to the normative value (Tombaugh, 2004), as participants improved from percentile 20 to percentile 50 in the TMT-A and from percentile 20 to percentile 40 in the TMT-B.

Because of the high number of missing values concerning physiological stress markers and particularly the ambulatory assessment, the data are presented descriptively in the ESM 2. We did find characteristic day profiles, although there were no significant differences in salivary stress markers over time.

Discussion

This study exploratorily investigated the acceptability, feasibility, and efficacy of a physical activity GPS watch to increase PA and improve depressive symptoms in inpatients with LLD. Further, it examined the mechanisms underlying the effect of PA on LLD by means of psychobiological stress markers. The findings regarding aspects of feasibility and acceptability were very promising. The mere use of the GPS watch seems to have activating effects:

Table 3. Objective activity tracker data relative to group (control vs. intervention) and order of randomization (intervention first vs. control first)

	Phase: Intervention		Phase: Control		Test statistics: MANOVA		
	First intervention	First control	First intervention	First control	Main effect on phase	Main effect on order	Interaction effect of phase × order
	M ± *SD*	*M* ± *SD*	*M* ± *SD*	*M* ± *SD*			
Activity time	7.12 ± 2.78	7.71 ± 0.90	7.31 ± 1.61	8.04 ± 1.36	$F = 0.449$, $p = .504$	$F = 3.225$, $p = .075$	$F = 0.051$, $p = .823$
Inactivity markers	0.16 ± 0.23	0.14 ± 0.18	0.18 ± 0.26	0.09 ± 0.21	$F = 0.783$, $p = .378$	$F = 1.114$, $p = .294$	$F = 0.273$, $p = .602$
Activity goal	0.70 ± 0.47	0.95 ± 0.23	0.61 ± 0.50	0.92 ± 0.25	$F = 0.156$, $p = .694$	$F = 0.935$, $p = .336$	$F = 0.013$, $p = .911$
Energy expenditure	2473 ± 466	2467 ± 249	2587 ± 386	2607 ± 490	$F = 2.554$, $p = .113$	$F = 0.006$, $p = .939$	$F = 0.023$, $p = .880$
Steps	9967 ± 3954	13931 ± 2756	11251 ± 4418	13128 ± 4003	$F = 0.206$, $p = .651$	$F = 14.695$, $p \leq .001$	$F = 2.375$, $p = .126$
Time spent							
Lying	8.96 ± 2.09	9.55 ± 1.94	9.35 ± 1.57	8.12 ± 2.70	$F = 1.645$, $p = .202$	$F = 0.615$, $p = .435$	$F = 5.120$, $p = .026$
Sitting	7.09 ± 1.74	6.48 ± 1.11	7.15 ± 1.22	6.20 ± 1.18	$F = 0.168$, $p = .682$	$F = 8.754$, $p = .004$	$F = 0.382$, $p = .538$
Standing	4.06 ± 1.29	4.47 ± 0.82	4.45 ± 1.13	4.69 ± 0.91	$F = 2.145$, $p = .146$	$F = 2.472$, $p = .119$	$F = 0.176$, $p = .676$
Walking	2.58 ± 2.54	2.66 ± 0.81	2.13 ± 0.74	2.83 ± 0.70	$F = 0.207$, $p = .650$	$F = 1.696$, $p = .196$	$F = 1.064$, $p = .305$
Running	0.46 ± 0.52	0.67 ± 0.42	0.48 ± 0.65	0.46 ± 0.35	$F = 1.074$, $p = .302$	$F = 0.932$, $p = .336$	$F = 1.525$, $p = .219$

Note. Mean values in physical activity watch data are presented (average daily activity per phase), activity time: time in movement as calculated by manufacturer's algorithm is reported in hours, inactivity markers and activity goal reached in %, energy expenditure in kcal: kilocalories, first intervention/control: order of randomization (intervention group first; control group first). *M* = mean; *SD* = standard deviation.

We found high levels of PA as measured by steps per day in the total sample, irrespective of assignment to intervention or control group. However, our analyses also revealed short-term motivational effects of the intervention based on the order of randomization, insofar as patients who underwent the control group first and did not initially receive activity feedback were more active and more stressed than patients who received activity feedback first. These differential trajectories, which converged at follow-up but not at the end of the intervention, highlight the need to reassess the effect of continuous feedback on activity levels in patients with LLD. It is of note, that in our crossover design, patients with LLD deteriorated in the psychometric test battery from the beginning to the end of the intervention, which they were unable to compensate for at the end of the intervention, even after changing from the control to the intervention group. Although the trajectories converged at follow-up, the effects of being in a control condition first might have overshadowed the beneficial effects of the intervention group. The assessment of underlying psychobiological mechanisms was challenging and yielded a high number of missing values. Thus, again, these mechanisms need to be re-examined in future studies with larger sample size.

Highly Satisfied, but Difficult to Reach

The acceptance questionnaire (Tables 1 and 2) indicated that patients predominantly perceived the study participation as interesting and satisfying. Study involvement was not associated with adverse side effects, highlighting PA as a safe therapeutic approach (DGPPN et al., 2015; Gelenberg et al., 2010; National Collaborating Centre for Mental Health, 2010). It should be noted, however, that only a subsample of 12 patients completed the acceptance questionnaires at T2 and only 7 patients at T3, leading to ambiguities about the acceptance of the other patients. According to the technology acceptance model, usefulness and ease of use contribute to attitudes toward technology and the consequent use thereof (Al-Maroof et al., 2021; Marangunić & Granić, 2015). Thus, if patients perceive meaningful feedback from both the watch and the therapist during use, this may increase the perceived usefulness and in turn positively impact their attitudes toward the GPS watch (Marangunić & Granić, 2015). In this context, we must stress the idea that the use of technological innovations such as wearable devices should be embedded in a therapeutic rationale. On the one hand, that allows lowering the threshold for the use of technological devices, as their purpose is clearly stated; on the other hand, it allows including real-life data into the therapeutic process, which might facilitate the transfer of therapeutic content to daily life. Ultimately, this could lead to wearable devices being an adjuvant treatment option for patients with LLD among a variety of tools that can meaningfully stimulate the therapeutic process (e.g., ambulatory just-in-time interventions). In this regard, it is promising that the high acceptance was maintained at follow-up, with a high agreement on even longer-lasting effects of the intervention. In contrast, the repeated saliva collections and the completion of the diary to assess goal-setting of behavioral activation were perceived as less convenient, thus explaining the high number of missing values, limiting the interpretability of the findings.

Of the 103 patients who initially expressed interest, only approximately a quarter ultimately chose to participate in

our study. This suggests that PA may be particularly appropriate for treating depressive symptoms in a subgroup of patients who are both interested in and physically able to participate in PA (Craft & Landers, 1998). These factors may deteriorate later in life, as comorbidities and physical and psychological barriers increase (Catalan-Matamoros et al., 2016), including "being too tired," "illness," or "feeling unsafe" (Ussher et al., 2007), which are among the main reasons for not engaging in PA. Therefore, low-threshold interventions aimed at behavior change to increase daily PA levels may be of great interest. The approach used in our study of using a GPS watch along with achievable activity goals, including aspects of biofeedback, appears to be appropriate for MDD patients who typically have low levels of PA and/or performance (Boettger et al., 2009; Ussher et al., 2007). Kononova et al. (2019) revealed that older adults are generally open to the use of wearables, and that these devices may be a useful tool to increase PA, even among people who have not previously used them.

The dropout rate of 26% is comparable to previous studies using exercise to treat depressive symptoms (Blumenthal et al., 1999, 2007), slightly lower than that reported in an RCT of older patients with chronic obstructive pulmonary disease (COPD) using a PA app (Vorrink et al., 2016), and higher than in a study of healthy older adults using a wearable for PA enhancement (Lyons et al., 2017). To the best of our knowledge, however, no RCT has been conducted in patients with MDD (Gal et al., 2018), particularly not when the use of the device is part of a therapeutic rationale. In addition, RCTs in this research area generally ranged from 4 weeks to 12 months and can therefore be used only as an approximate comparison (see Gal et al., 2018, for review).

Continuous Activity Feedback as a Double-Edged Sword – Motivational Effects of Randomization Order

Depressive symptoms did not change over time. Although improvements in depressive symptoms are possible within such a short period after starting to exercise (Haller et al., 2018), this is not considered likely. Interestingly, results on measures of stress and self-efficacy changed over time, with increases in self-efficacy and decreases in stress. In particular, self-efficacy typically improves through exercise and is considered one of the main factors for improving mood and depressive symptoms (Blumenthal et al., 2007; Haller et al., 2018). Most interestingly, trajectories in subjective stress ratings varied depending on order of randomization: Patients who underwent the intervention condition first experienced a stress reduction from baseline to end, whereas patients who underwent the control condition first experienced increases in stress during this time frame. The stress ratings converged at follow-up but not at the end of the intervention. This is a relevant finding, as it shows that merely wearing a PA GPS watch without providing feedback on the actual activity might induce negative effects that even overshadow the beneficial effects of the feedback provided during participation in the intervention group. Even the therapeutic sessions that focused on activity and mood in both the intervention and the control group were unable to compensate for these demotivating effects of the control condition. At the same time, the patients who were first in the control group felt not only more stressed but were also physically more active. This might be interpreted as they might have felt pressured to be more active as the watch assessed their activity without providing feedback. This would be a dangerous consequence as PA should not be associated with stress and pressure in patients with LLD. The use of technology and a decrease in well-being are common observations. In their study, Wolf et al. (2021) observed that smartphone use among older adults was highly interindividual and intraindividual over time. In addition, smartphone use could also lead to a decline in well-being, although many participants used their smartphones for messaging with their social contacts, and social contacts were rated as important by the participants. These results illustrate that the negative effects of new technologies are also conceivable, and that the advantages and disadvantages of their use must be weighed individually.

Generally speaking, the PA level assessed by the GPS watch did not significantly differ between the control and intervention groups. In this respect, a recent systematic review (Gal et al., 2018) demonstrated moderate effects of wearable devices on PA for intervention periods ranging from 4 weeks to 12 months. However, we assessed an effect of randomization order for steps, time spent sitting, and time spent lying down. Notably, patients exhibited relatively high levels of PA regardless of group assignment ($\geq \sim$10,000 steps per day, Table 3). We cannot rule out that enrollment in the study per se caused a significant increase in PA. In addition, the fact that the control group also began wearing a new activity watch may have encouraged them to increase PA, resulting in no significant differences between the groups. In this context, "awareness of physical activity and curiosity" (Kononova et al., 2019) may have led to a marked increase in daily PA in both groups compared to prestudy levels. This assumption applies especially to the motivational effects of the intervention, as the mean levels of PA did not objectively differ, whereas different trajectories in the outcome measures emerged when controlling for order of randomization.

Studies in this research area often lack follow-up data (Gal et al., 2018). Although we conducted a follow-up 6 weeks

after the intervention, which revealed promising results, it is not apparent whether participants acquired their own wearable and/or maintained the marked level of daily PA ($\geq \sim$10,000 steps per day, Table 3). Because all patients had already been discharged from the hospital at the time of the follow-up, we were only able to analyze data from seven participants at this time point. Among them, we found a tendency for a sustained improvement in lifestyle (high scores on the items "Positive changes in daily routine," "Still try to implement what I have learned," "Always try to be active to improve my mood," and "Because of the study treatment I experience more joy in movement"). Kononova et al. (2019) emphasized that long-term use of wearables and long-term adherence to PA are highly dependent on social support. This is especially important in LLD patients, who are often isolated and benefit from sustained social life and support (Bruce, 2002; Haller, 2022; Herbolsheimer et al., 2018; Ussher et al., 2007).

Limitations and Outlook

While our exploratory findings demonstrate that the program was generally feasible and well accepted by patients, future studies clearly need to reassess our findings not only in larger samples but also by considering the lessons learned from this study. Therefore, we formulate three recommendations concerning the study design for future next steps: (1) We recommend re-assessing the controlled crossover design. On the one hand, the crossover design randomizing patients to either receive the intervention or the control condition first might have demotivating effects that overshadow potential benefits in patients initially in the control condition. On the other hand, because patients with LLD are particularly burdened by symptoms such as lower mood, reduction of energy, and decreased levels of PA, they might be additionally burdened by first being randomized into a control condition. In general, passive control conditions might be difficult in this particular patient population; active control conditions might be more suitable. In this regard, including a control group of age-matched persons without depression in a randomized controlled design should be considered. (2) We recommend a longer study duration of 8–10 weeks. This longer timeframe may also improve endurance performance and thus have an additional positive effect on depressive symptoms as defined by current guidelines (Rethorst & Trivedi, 2013). It should be noted, however, that such a study would be difficult to conduct in an inpatient setting; thus, an outpatient setting may be the first choice here. (3) We suggest collecting PA data prior to the intervention, thus enabling changes to be detected solely by wearing a novel GPS watch. Future studies need to assess the potential of wearables technology in patients with late-life depression in larger-scale studies that allow examining underlying mechanisms and the specific effects of the behavioral activation in differentiation to the effects of an inpatient stay in general. Furthermore, it will be important to differentiate the effects of our intervention on physical activity from the effects on depressive symptoms and to understand how these effects are driven by the use of the wearables per se. By using mediation analyses, one can assess the mechanisms underlying these complex interactions to ultimately stimulate therapeutic recommendations. This allows individually tailoring the use of wearables to the needs of patients with LLD and gaining knowledge on predictors for the successful and meaningful use of these technologies in behavioral activation interventions.

Conclusions

We conducted a PA intervention with innovative wearable technology in patients with late-life depression. The intervention was feasible and well accepted, leading to differential effects on PA and stress measures based on randomization order. In particular, patients in the control group first who did not initially receive activity feedback were more active and more stressed in the short term compared to patients who received activity feedback first. These differences in trajectories underline the need to critically re-assess the effect of continuous feedback on activity levels, although the differences disappeared at follow-up (but not at the end of the intervention). An examination of the mechanisms underlying the effect of PA in patients with late-life depression utilizing psychobiological stress markers was challenging and resulted in a high number of missing values. In patients with late-life depression, physiological stress markers that require less active patient engagement might be more suitable than repeatedly collecting saliva samples over time. Future studies should demonstrate whether wearable devices are suitable to increase PA in patients with late-life depression in the long term, and whether they can contribute to improving depressive symptoms as part of a multimodal therapy concept.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at https://doi.org/10.1024/1662-9647/a000312

ESM 1. Pairwise comparisons of paper-and-pencil questionnaires comparing T0-T1, T1-T2, T2-T3

ESM 2. Descriptive data on psychobiological stress markers assessed by ambulatory assessment. Error bars: ±1 standard error.

References

Al-Maroof, R. S., Alhumaid, K., Alhamad, A. Q., Aburayya, A., & Salloum, S. (2021). User acceptance of smart watch for medical purposes: An empirical study. *Future Internet, 13*(5), Article 127. https://doi.org/10.3390/fi13050127

Alexopoulos, G. S. (2019). Mechanisms and treatment of late-life depression. *Translational Psychiatry, 9*(1), 1–16. https://doi.org/10.1038/s41398-019-0514-6

Andreas, S., Schulz, H., Volkert, J., Dehoust, M., Sehner, S., Suling, A., Ausín, B., Canuto, A., Crawford, M., & Da Ronch, C. (2017). Prevalence of mental disorders in elderly people: The European MentDis_ICF65 + Study. *British Journal of Psychiatry, 210*(2), 125–131. https://doi.org/10.1192/bjp.bp.115.180463

Andreescu, C., Ajilore, O., Aizenstein, H. J., Albert, K., Butters, M. A., Landman, B. A., Karim, H. T., Krafty, R., & Taylor, W. D. (2019). Disruption of neural homeostasis as a model of relapse and recurrence in late-life depression. *American Journal of Geriatric Psychiatry, 27*(12), 1316–1330. https://doi.org/10.1016/j.jagp.2019.07.016

Blake, H., Mo, P., Malik, S., & Thomas, S. (2009). How effective are physical activity interventions for alleviating depressive symptoms in older people? A systematic review. *Clinical Rehabilitation, 23*(10), 873–887. https://doi.org/10.1177/0269215509337449

Blumenthal, J. A., Babyak, M. A., Doraiswamy, P. M., Watkins, L., Hoffman, B. M., Barbour, K. A., Herman, S., Craighead, W. E., Brosse, A. L., & Waugh, R. (2007). Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosomatic Medicine, 69*(7), 587–596. https://doi.org/10.1097/PSY.0b013e318148c19a

Blumenthal, J. A., Babyak, M. A., Moore, K. A., Craighead, W. E., Herman, S., Khatri, P., Waugh, R., Napolitano, M. A., Forman, L. M., & Appelbaum, M. (1999). Effects of exercise training on older patients with major depression. *Archives of Internal Medicine, 159*(19), 2349–2356. https://doi.org/10.1001/archinte.159.19.2349

Boettger, S., Wetzig, F., Puta, C., Donath, L., Müller, H.-J., Gabriel, H. H., & Bär, K.-J. (2009). Physical fitness and heart rate recovery are decreased in major depressive disorder. *Psychosomatic Medicine, 71*(5), 519–523. https://doi.org/10.1097/PSY.0b013e3181a55303

Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computer Education, 88*, 343–353. https://doi.org/10.1016/j.compedu.2015.07.013

Bridle, C., Spanjers, K., Patel, S., Atherton, N. M., & Lamb, S. E. (2012). Effect of exercise on depression severity in older people: Systematic review and meta-analysis of randomised controlled trials. *British Journal of Psychiatry, 201*(3), 180–185. https://doi.org/10.1192/bjp.bp.111.095174

Brown, L., Karmakar, C., Gray, R., Jindal, R., Lim, T., & Bryant, C. (2018). Heart rate variability alterations in late life depression: A meta-analysis. *Journal of Affective Disorders, 235*, 456–466. https://doi.org/10.1016/j.jad.2018.04.071

Bruce, M. L. (2002). Psychosocial risk factors for depressive disorders in late life. *Biological Psychiatry, 52*(3), 175–184. https://doi.org/10.1016/S0006-3223(02)01410-5

Catalan-Matamoros, D., Gomez-Conesa, A., Stubbs, B., & Vancampfort, D. (2016). Exercise improves depressive symptoms in older adults: An umbrella review of systematic reviews and meta-analyses. *Psychiatry Research, 244*, 202–209. https://doi.org/10.1016/j.psychres.2016.07.028

Chum, J., Kim, M. S., Zielinski, L., Bhatt, M., Chung, D., Yeung, S., Litke, K., McCabe, K., Whattam, J., & Garrick, L. (2017). Acceptability of the Fitbit in behavioural activation therapy for depression: A qualitative study. *Evidence Based Mental Health, 20*(4), 128–133. https://doi.org/10.1136/eb-2017-102763

Cohen, S. (1988). Perceived stress in a probability sample of the United States. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health* (pp. 31–67). Sage Publications, Inc.

Craft, L. L., & Landers, D. M. (1998). The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis. *Journal of Sport and Exercise Psychology, 20*(4), 339–357.

DGPPN, KBV, AkdÄ, BApK, DEGAM, & DGPs. (2015). *S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression–Langfassung* (2nd ed., Version 5). Springer.

Enright, P. L. (2003). The Six-Minute Walk Test. *Respiratory Care, 48*(8), 783–785. http://rc.rcjournal.com/content/48/8/783.abstract

Gal, R., May, A. M., van Overmeeren, E. J., Simons, M., & Monninkhof, E. M. (2018). The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: A systematic review and meta-analysis. *Sports Medicine Open, 4*(1), Article 42. https://doi.org/10.1186/s40798-018-0157-9

Gelenberg, A. J., Freeman, M., Markowitz, J., Rosenbaum, J., Thase, M., Trivedi, M., & Van Rhoads, R. (2010). American Psychiatric Association practice guidelines for the treatment of patients with major depressive disorder. *American Journal of Psychiatry, 167*(Suppl 10), 9–118.

Haller, N. (2022). Web-based exercise therapy for the treatment of depression. *Sports Psychiatry*. https://doi.org/10.1024/2674-0052/a000016

Haller, N., Lorenz, S., Pfirrmann, D., Koch, C., Lieb, K., Dettweiler, U., Simon, P., & Jung, P. (2018). Individualized web-based exercise for the treatment of depression: Randomized controlled trial. *JMIR Mental Health, 5*(4), Article e10698. https://doi.org/10.2196/10698

Haller, N., & Simon, P. (2017). Sport und Bewegung zur Therapie und Prävention. In A. Fellgiebel & M. Hautzinger (Eds.), *Altersdepression* (pp. 187–196). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-53697-1_17

Hautzinger, M. (2000). *Depression im Alter (Materialien für die klinische Praxis)*. Beltz PVU.

Herbolsheimer, F., Ungar, N., & Peter, R. (2018). Why is social isolation among older adults associated with depressive symptoms? The mediating role of out-of-home physical activity. *International Journal of Behavioral Medicine, 25*(6), 649–657. https://doi.org/10.1007/s12529-018-9752-x

Horackova, K., Kopecek, M., Machů, V., Kagstrom, A., Aarsland, D., Motlova, L. B., & Cermakova, P. (2019). Prevalence of late-life depression and gap in mental health service use across European regions. *European Psychiatry, 57*, 19–25. https://doi.org/10.1016/j.eurpsy.2018.12.002

Kessler, E.-M., & Blachetta, C. (2020). Age cues in patients' descriptions influence treatment attitudes. *Aging and Mental Health, 24*(1), 193–196. https://doi.org/10.1080/13607863.2018.1515889

Kononova, A., Li, L., Kamp, K., Bowen, M., Rikard, R. V., Cotten, S., & Peng, W. (2019). The use of wearable activity trackers among older adults: Focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. *JMIR mhealth uhealth, 7*(4), Article e9832. https://doi.org/10.2196/mhealth.9832

Laird, K. T., Krause, B., Funes, C., & Lavretsky, H. (2019). Psychobiological factors of resilience and depression in late life. *Translational Psychiatry, 9*(1), Article 88. https://doi.org/10.1038/s41398-019-0424-7

Lyons, E. J., Lewis, Z. H., Mayrsohn, B. G., & Rowland, J. L. (2014). Behavior change techniques implemented in electronic lifestyle activity monitors: A systematic content analysis. *Journal of Medical Internet Research, 16*(8), Article e3469. https://doi.org/10.2196/jmir.3469

Lyons, E. J., Swartz, M. C., Lewis, Z. H., Martinez, E., & Jennings, K. (2017). Feasibility and acceptability of a wearable technology physical activity intervention with telephone counseling for mid-aged and older adults: A randomized controlled pilot trial. *JMIR mhealth uhealth, 5*(3), Article e6967. https://doi.org/10.2196/mhealth.6967

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review from 1986 to 2013. *Universal Access in the Information Society, 14*(1), 81–95. https://doi.org/10.1007/s10209-014-0348-1

Naslund, J. A., Aschbrenner, K. A., Scherer, E. A., McHugo, G. J., Marsch, L. A., & Bartels, S. J. (2016). Wearable devices and mobile technologies for supporting behavioral weight loss among people with serious mental illness. *Psychiatry Research, 244*, 139–144. https://doi.org/10.1016/j.psychres.2016.06.056

National Collaborating Centre for Mental Health (UK). (2010). *Depression: The treatment and management of depression in adults* (Updated edition). British Psychological Society.

Reitan, R. M., & Wolfson, D. (1995). Category test and trail making test as measures of frontal lobe functions. *Clinical Neuropsychology, 9*(1), 50–56. https://doi.org/10.1080/13854049508402057

Rethorst, C. D., & Trivedi, M. H. (2013). Evidence-based recommendations for the prescription of exercise for major depressive disorder. *Journal of Psychiatric Practice, 19*(3), 204–212. https://doi.org/10.1097/01.pra.0000430504.16952.3e

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583. https://doi.org/10.1016/s0006-3223(02)01866-8

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health*.

Shindel, C., Holland, J. M., & Gallagher-Thompson, D. (2020). The link between activities of daily living and cortisol in late-life depression. *Clinical Gerontology, 43*(4), 430–440. https://doi.org/10.1080/07317115.2018.1561581

Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric Research, 77*, 42–51. https://doi.org/10.1016/j.jpsychires.2016.02.023

Schulz, P., Schlotz, W., & Becker, P. (2004). *TICS. Trierer Inventar zum chronischen Stress. Manual*. Hogrefe.

Schwarzer, R., & Jerusalem, M. (2003). SWE. Skala zur Allgemeinen Selbstwirksamkeitserwartung [Verfahrensdokumentation, Autorenbeschreibung und Fragebogen]. In Leibniz-Institut für Psychologie (ZPID) (Hrsg.), Open Test Archive. ZPID. https://doi.org/10.23668/psycharchives.4515

Strath, S. J., & Rowley, T. W. (2018). Wearables for promoting physical activity. *Clinical Chemistry, 64*(1), 53–63. https://doi.org/10.1373/clinchem.2017.272369

Tiemeier, H. (2003). Review: Biological risk factors for late life depression. *European Journal of Epidemiology, 18*(8), 745–750. https://doi.org/10.1023/A:1025388203548

Tombaugh, T. N. (2004). Trail Making Test A and B: normative data stratified by age and education. *Archives of clinical neuropsychology : the official journal of the National Academy of Neuropsychologists, 19*(2), 203–214. https://doi.org/10.1016/S0887-6177(03)00039-8

Toni, G., Murri, M. B., Piepoli, M., Zanetidou, S., Cabassi, A., Squatrito, S., Bagnoli, L., Piras, A., Mussi, C., & Senaldi, R. (2016). Physical exercise for late-life depression: Effects on heart rate variability. *American Journal of Geriatric Psychiatry, 24*(11), 989–997. https://doi.org/10.1016/j.jagp.2016.08.005

Ussher, M., Stanbury, L., Cheeseman, V., & Faulkner, G. (2007). Physical activity preferences and perceived barriers to activity among persons with severe mental illness in the United Kingdom. *Psychiatric Services, 58*(3), 405–408. https://doi.org/10.1176/ps.2007.58.3.405

Vorrink, S. N., Kort, H. S., Troosters, T., Zanen, P., & Lammers, J.-W. J. (2016). Efficacy of an mHealth intervention to stimulate physical activity in COPD patients after pulmonary rehabilitation. *European Respiratory Journal, 48*(4), 1019–1029. https://doi.org/10.1183/13993003.00083-2016

Wolf, F., Naumann, J., & Oswald, F. (2021). Digital social interactions in later life: Effects of instant messaging on situational mood of older smartphone users. In Q. Gao & J. Zhou (Eds.), *Human aspects of IT for the aged population. Technology design and acceptance. HCII 2021. Lecture notes in computer science, vol 12786*. Springer. https://doi.org/10.1007/978-3-030-78108-8_33

History
Received October 21, 2022
Accepted March 13, 2023
Published online May 12, 2023

Acknowledgments
We thank Therese Steudter for her skillful and diligent work analyzing the saliva samples. We further thank Ricarda Baake for her assistance in study organization, data collection, and data management.

Conflict of Interest
The authors declare no conflict of interest.

Editorial Note
The acting editor was Frieder R. Lang.

Funding
Alexandra Wuttke gratefully acknowledges funding by the Internal University Research Funding of the University Hospital Center of Mainz, Germany.
Open access publication enabled by Johannes Gutenberg University Mainz, Germany.

ORCID
Alexandra Wuttke
https://orcid.org/0000-0003-0655-4621
Kristina Endres
https://orcid.org/0000-0002-1099-8287
Perikles Simon
https://orcid.org/0000-0002-7996-4034
Nils Haller
https://orcid.org/0000-0002-7066-7615

Dr. Nils Haller
Department of Sports Medicine, Disease Prevention and Rehabilitation
Johannes Gutenberg University Mainz
Albert-Schweitzer-Straße 22
55128 Mainz
Germany
nhaller@uni-mainz.de