Till Utesch[1] · **Ludwig Piesch[2]** · **Lena Busch[2]** · **Bernd Strauss[2]** · **Katharina Geukes[3]**

[1] Institute for Educational Sciences, University of Münster, Münster, Germany

[2] Institute of Sport Sciences, University of Münster, Münster, Germany

[3] Department of Psychology, University of Münster, Münster, Germany

Self-tracking of daily physical activity using a fitness tracker and the effect of the 10,000 steps goal

A 6-week randomized controlled parallel group trial

Introduction

As noncommunicable diseases have become the predominant cause of global mortality (Lakerveld et al., 2020) and obesity has reached epidemic proportions worldwide (Krzysztoszek, Laudanska-Krzeminska, & Bronikowski, 2019), prioritized concerns of health agencies (Heath et al., 2012) are the reduction of sedentary behavior and the promotion of physical activity (PA). Yet, recent estimates show that substantial numbers of adults (27.5%) and adolescents (81%; Bull et al., 2020) fail to meet the *Global Recommendations on Physical Activity for Health* for aerobic exercise of the World Health Organization (WHO, 2010). Recognizing this pandemic of physical inactivity (Kohl et al., 2012), this failure points at the demand of successful, efficient, and globally applicable behavior-change interventions targeting the elevation of PA participation levels.

With recent advancements of digital instruments and applications and with the rapid growth of mobile device ownership and internet accessibility, electronical and mobile technologies find their way into health-related environments and research (Vandelanotte et al., 2016). Out of this trend, a new generation of consumer-based wearable devices has emerged in the shape of fitness tracker (FT) wristbands, often used with a companion smartphone application (app). FTs operate as electronic activity monitors (e.g., regarding step count, traveled distance, estimates of energy expenditure; Lynch, Bird, Lythgo, & Selva-Raj, 2020), providing consumers with real-time feedback on their activity (Brickwood, Watson, O'Brien, & Williams, 2019). As such, FTs seemingly offer an attractive, easily and widely applicable, as well as cost-effective alternative to conservative PA behavior change interventions (e.g., Lyons, Lewis, Mayrsohn, & Rowland, 2014; Sullivan & Lachman, 2017). Consistent with a growing FT market and considerable rises in expected sales figures (Loomba & Khairnar, 2017), increasing numbers of health agencies recognize the potential benefits of activity tracking (AT) devices for PA behavior and their utility as means for behavior change in PA interventions. Therefore, the purpose of this ambulatory assessment study is to empirically evaluate the effects of activity self-tracking, alone and in combination with a daily step goal, on daily PA behavior in a randomized-controlled parallel group trial.

Self-tracking and PA

The worldwide number of wearables capable of sensing and recording activity data has been estimated to reach over 1.1. billion by 2022 (Statista, 2021), promising a great potential of wearables in reaching out to individuals to promote PA levels. The underlying premise for this assumption lies in the expected positive relationship between self-tracking of PA parameters and PA-related outcomes, as for example described by the Social Cognitive Theory (SCT) of self-regulation (Bandura, 1991). According to this theory, human behavior is not solely regulated by external sources, but to a large extent by self-reflective and self-reactive capabilities (i.e., based on self-insight) that allow people to exert control over their thoughts, emotions, motivation, and actions. Thus, self-monitoring of behavior (here: the use of ATs) provides these self-insights (here: *I move too little*), which, in turn, provides the basis to infer that current behavior should be changed (here: *I will move more*) and to initiate a process of corrective behavior change into a desired direction (e.g., *I will walk home instead of taking the bus*; Bandura, 1991; Kersten-van Dijk, Westerink, Beute, & IJsselsteijn, 2017; Shull, Jirattigalachote, Hunt, Cutkosky, & Delp, 2014; Stiglbauer, Weber, & Batinic, 2019).

With respect to self-monitoring technology, this is referred to as the *self-improvement hypothesis of personal informatics*.

Despite being relatively new to the field of PA research, this assumption has been empirically tested in interventions studies (e.g., Cadmus-Bertram, Marcus, Patterson, Parker, & Morey, 2015; Giddens, Leidner, & Gonzalez, 2017; Wang et al., 2015) as well as described in reviews and meta-analyses (Bravata et al., 2007; Brickwood et al., 2019; Gal, May, van Overmeeren, Simons, & Monninkhof, 2018; Jee, 2017; Lynch et al., 2020; Romeo et al., 2019). Overall, evidence regarding this hypothesis is inconsistent. While some of the evidence supports the benefit of FTs for PA-related outcomes (e.g., Bravata et al., 2007; Brickwood et al., 2019), others question the surplus for health outcomes (e.g., blood pressure, weight; Finkelstein et al., 2016; Lynch et al., 2020) or even doubt their efficacy as a tool for elevating PA (e.g., McDermott et al., 2018; Melton, Buman, Vogel, Harris, & Bigham, 2016). These inconsistencies can be traced back to methodological, statistical, and clinical heterogeneity (i.e., differences in participants, interventions or outcomes; Lynch et al., 2020) in the studies, highlighting that an empirically based consensus on the utility of FTs for sustained PA behavior change needs yet to be found (Schoeppe et al., 2016; Sullivan & Lachman, 2017).

Goal setting and PA

Beyond mere activity self-tracking, goal setting is a prevalently applied technique in PA interventions (Sullivan & Lachman, 2017) and commonly implemented in the use of FTs with apps (Lyons et al., 2014). As such, users can predefine a certain PA quantity, for example, a desired number of steps per day, and monitor their progress towards this goal via a connected app or on the wearable itself. A frequently recommended PA goal, for example, is 10,000 steps per day for healthy adults as reasoned by Tudor-Locke et al. (2011) and used in PA promotion programs (e.g., 10,000 Steps Australia program; Duncan, Brown, Mummery, & Vandelanotte, 2018).

According to Gal et al. (2018), goal setting is an efficient method to increase PA levels and one of the most important behavior change techniques within combined wearables and smartphone apps. In line with the SCT of self-regulation (Bandura, 1991), self-insight facilitates behavior change. Goal setting is thought to aid this behavior change process because it provides individuals with the opportunity of *evaluating* the provided self-insight information based on a standard. Specifically, goal setting allows comparisons of a normative value (e.g., step goal) with a current real value (e.g., actual steps) and results of these comparisons (i.e., discrepancies) may initiate further behavioral adjustments (here: more steps; Bandura, 1991; Stiglbauer et al., 2019; Sullivan & Lachman, 2017).

With respect to goal setting in FT use, two approaches need to be distinguished: (1) personally set goals, that is, by the consumers themselves and (2) externally set goals are, that is, based on general recommendations or random company decisions in apps. Unsurprisingly, the effectiveness of personally set goals on behavior varies as a function of persons' ability to define goals that are conducive (i.e., challenging, realistic) to their behavior and suitable for eliciting changes in behavior (Bandura, 1991; Sullivan & Lachman, 2017). The effectiveness of externally set goals, however, is not yet conclusively investigated. For instance, while Bravata et al. (2007) identified generalized step goals such a 10,000 steps per day to be a relevant predictor of increased PA, they also explicitly call for randomized controlled trials comparing FT use with against FT use without a step goal. Furthermore, as users frequently just stick to default settings or define goals following general recommendations for minimum amounts of PA (Sullivan & Lachman, 2017), a systematic evaluation of the effectiveness of externally set goals on PA behavior is severely needed.

The present study

Taking these deliberations into account, this study aims to examine how activity-related self-tracking, alone and in combination with an externally assigned step goal, relates to PA levels (and trajectories) in individuals' daily lives. In doing so, we aspire to draw reliable conclusions on the usefulness of wearable AT devices and associated features for PA behavior change and to provide recommendations for future PA interventions. For this purpose, we conducted a randomized-controlled, parallel-group trial with two experimental groups and one additional non-randomized control group (C group). In all three groups, PA was assessed over the course of 6 weeks (42 days) via a daily PA questionnaire. In the experimental groups, participants were additionally equipped with the commercially available, wrist-worn FT Fitbit Flex 2 (Fitbit©, San Francisco, CA, USA) and a connected smartphone application to track their PA (i.e., daily step count). One of these experimental groups was additionally provided with a step goal of 10,000 steps per day that was displayed via the connected smartphone application (SG), whereas the other experimental group was not (NSG). Specifically, we expected increased average PA levels in the experimental groups compared to the C group based on the PA questionnaire data (hypothesis 1) and increased average PA levels in the SG group compared to the NSG group based on the daily step count (hypothesis 2). For both comparisons, we additionally explore potential differences in activity trajectories (i.e., explorative research questions: potential PA decline after getting used to the Fitbits).

Methods

Participants

The design and samples in this study were already used in two studies of a larger research program (German Clinical Trials Register, grant no. DRKS00014835;

Busch, Utesch, Bürkner, & Strauss, 2020a; Busch, Utesch, & Strauss, 2020b)[1].

Based on a recommended minimum of 20 subjects per group for the planned multilevel analysis (Kreft & de Leeuw, 1998) and guidelines from previous studies in this field (Jee, 2017; Schoeppe et al., 2016), a sample size with 50 subjects per group was planned in the study design. Inclusion criteria of this study were (a) to include young adults between 18 and 40 years of age (representative for the targeted population being at increased risk for unhealthy weight gain in this stage and a pronounced PA decline; Laska, Pelletier, Larson, & Story, 2012; Nelson, Story, Larson, Neumark-Sztainer, & Lytle, 2008) who (b) exercise on average less than 4 hours per week (representative for the targeted population needing to increase PA levels), (c) have not used a fitness app or comparable AT devices for more than 2 weeks during the past 6 months (to ascribe effects to the initial exposure to FTs), (d) have an internet enabled and Bluetooth compatible smartphone (required to display the fitness application as well as access to a USB port to charge the tracking device), (e) are free from injuries or diseases (to ensure valid outcome measure assessment), (f) do not plan to travel for more than 1 week during the intervention period (to ensure assessment in participants' usual context), and (g) do not work on nightshifts on a regular basis (to ensure that recording 1 day from midnight to midnight accurately reflects participants' natural daily routine). Overall, 152 participants met these inclusion criteria and were blinded to the study design. In all, 52 persons (i.e., the C group) could not be randomized because they were recruited for a study

not mentioning Fitbits in order to make them blind regarding any knowledge about the existence of the two experimental groups that use Fitbits. A total of 100 persons were randomly and equally assigned the experimental groups using Fitbits. Because data of 2 individuals had to be excluded due to recording errors, a final sample of 150 participants ($M_{age} = 24.66$, $SD_{age} = 4.75$; $n_{women} = 117$, $n_{men} = 33$) resulted, with 49 participants ($M_{age} = 25.51$, $SD_{age} = 4.57$) in the SG group, 50 participants ($M_{age} = 25.78$, $SD_{age} = 4.78$) in the NSG group and 51 participants ($M_{age} = 22.75$, $SD_{age} = 4.36$) in the C group.

Procedures

The study design has been extensively explained in detail elsewhere (Busch et al., 2020a, b).

In accordance to similar fitness app-based intervention studies (Schoeppe et al., 2016), we considered the duration of 42 days to be reasonably long (a) to ensure committed participation, (b) to responsibly manage participants' burden, and (c) to prevent potential effects from being solely due to a "honeymoon" or novelty effect that can be observed when consumers access new health technologies (Whelan et al., 2019). On the first and 42nd day of the study, each participant was invited into the laboratory. Within the initial session, participants of all groups were introduced to the study's protocol and provided written informed consent. Participants of the NSG and the SG groups, moreover, received the FT Fitbit Flex 2, which was set up, connected to their smartphone, and explained to the participants in the lab. Accordingly, participants of these groups were able to monitor steps, consumed calories, covered distance, and active minutes as displayed via the connected smartphone app. Only the SG group received an additional, externally assigned step goal of 10,000 steps per day (midnight to midnight). Our rationale for defining the external step goal at 10,000 steps per day based on the study by Tudor-Locke et al. (2011) that indicates that 10,000 steps are a reasonable and attainable target for the target population of healthy

Abstract

Activity self-tracking technologies have become a popular tool for monitoring physical activity, used to assist individuals in increasing physical activity. Previous findings, however, are inconclusive on whether standalone activity self-tracking, alone or in combination with a step goal, is efficient in encouraging physical activity behavior. This study aimed at investigating the effects of mere activity self-tracking and activity self-tracking with a daily 10,000 steps goal on physical activity levels in a randomized-controlled parallel group trial ($N = 150$; mean age = 24.66, standard deviation = 4.75), using a publicly available fitness tracker and associated smartphone application. Two experimental groups, one of which without and the other with step goal, recorded daily physical activity (daily step count) via a fitness tracker wearable and associated smartphone application and via daily self-reports of activity over the course of 6 weeks. An unequipped control group recorded physical activity only via daily self-reports of activity. Both experimental groups were compared against each other based on step count data, and against the control group based on daily self-reports of activity. Results of mixed-effect multilevel analyses suggest that activity tracking using fitness trackers and externally assigned goals do not drive greater physical activity behavior. Rather, results indicate substantial individual differences that should be targeted in future studies to better understand individual behavioral processes boosting or hindering daily activity.

Keywords
FitBit · Activity behavior · Experiment · Ambulatory assessment · Mobile device

T. Utesch · L. Piesch · L. Busch · B. Strauss · K. Geukes

Self-tracking of daily physical activity using a fitness tracker and the effect of the 10,000 steps goal. A 6-week randomized controlled parallel group trial

---

[1] The current research question with respect to capturing life as it is lived (i.e., daily physical activity and steps) has not been answered yet in these two studies. These studies focused on the effects of using fitness applications on people's body awareness and well-being (Busch et al., 2020a) and their autonomy need satisfaction and motivation (Busch et al., 2020b). The outcome variables assessed in the current study have not been used in another study. All variables used in this larger research program funded by the DFG can be found under (https://osf.io/tebnx/).

adults, which reflects the population in this study, and based on the standardized preset goal in the Fitbit app that has also been used in other intervention studies (e.g., Goodyear, Kerner, & Quennerstedt, 2017). The progression of the achievement of the daily step goals was visualized by means of a circular bar on the app interface, progressing to a full circle at achievement of 10,000 steps a day (Busch et al., 2020a). Participants of these experimental groups were instructed not to change any of these settings and were asked to wear the tracker throughout the whole day whenever possible across the duration of the study. Over the course of the study, participants in all three groups received a daily text message at 9 pm with an invitation and a link to complete a daily PA questionnaire allowing the comparison of (self-reported) PA of both experimental groups with the C group. On day 42, participants' laboratory visits served to unblind and debrief them. Participants of the NSG and the SG groups additionally returned their devices and completed a questionnaire on their FT and the associated app use throughout the study period. As days 1 and 42 were only partly assessed, these data remained unconsidered, so that a total of 40 full days were the basis of analysis. Prior to the conduction of this study, all procedures were approved by the ethics committee of the University of Münster.

Measures

Physical activity

Two measures of PA were used in this study, a subjective questionnaire-based PA assessment and an objective assessment, that is, step count, as provided by the Fitbit Flex 2.

PA Questionnaire. To assess daily PA via a questionnaire, the Godin Leisure-Time Exercise Questionnaire (LTEQ; Shephard, 1997) was used. The LTEQ represents a reliable and valid instrument and has been widely applied to assess self-reported PA participation among adults (Gionet & Godin, 1989; Godin & Shephard, 1985; Jacobs, Ainsworth, Hartman, & Leon, 1993; van Poppel, Chinapaw, Mokkink, Van Mechelen, &

Terwee, 2010). In the original LTEQ, participants are asked to report how often *in a week* they engage in mild, moderate, and strenuous exercise for more than 15 min. In this study, the LTEQ was modified to assess activity *on a day.* Based on the subjects' reports and analogous to the original formula, a daily total LTEQ score was calculated multiplying the daily frequencies of mild, moderate and strenuous activities by three, five and nine metabolic equivalents of task (METs), respectively. Values above 200 were treated as outliers and excluded from analysis, as data inspection suggested that subjects seldomly reported activities in full minutes instead of 15-min units, resulting in implausible LTEQ scores stemming from data recording errors.

Step count. The Fitbit Flex 2 provides objective quantities of PA in terms of steps, distance, consumed calories, and active minutes, recorded in the coupled smartphone application (Fitbit Inc, 2018). Wrist-worn activity trackers from the Fitbit Flex model group and the Fitbit Flex 2 data in particular have demonstrated reasonable reliability (intraclass correlation (ICC) = 0.9) and validity (ICC = 0.77–0.85) for various PA measures (e.g., Diaz et al., 2015; Kooiman et al., 2015; Venetsanou et al., 2020). In the present study, the prevalently used (Sullivan & Lachman, 2017) and easily comprehendible step count data were used to reflect participants' daily PA levels. Values above 30,000 steps were treated as outliers and excluded from analysis following cutoff practice in other studies and programs from this field (e.g., Hohepa, Schofield, Kolt, Scragg, & Garrett, 2008; Silva, Meyer, & Jayawardana, 2020).

Statistical approach

All calculations were computed with R Studio (Version 1.1.463; R Core Team, 2018). Multilevel mixed-effects models (lme4 package; Bates, Sarkar, Bates, & Matrix, 2007) were computed to determine the effects of the intervention on PA levels (and trajectories) over the study period because daily PA assessments (level 1; within-person) were nested

in participants (level 2; between-person). All models were computed using maximum likelihood estimation.

We adhered to a bottom-up approach that is frequently applied in multilevel modeling research (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 1999), which means that we will apply likelihood ratio tests in order to evaluate improved model fit for less parsimonious models (i.e., adding predictor variables or random effects). The likelihood ratio test is based on comparing information criteria AIC and BIC between models while lower criteria indicate better model fit. This means that when comparing two adjacent models with different specifications, a negative Δ AIC/BIC indicates better fit to the data for the more complex model (i.e., added aspect as predictor, random effect). In this case, the added specification is considered to be meaningful. Further, a χ^2 test is conducted to compare adjacent models, especially if Δ AIC/BIC is inconclusive, indicating a better model fit for the less parsimonious model if the result is significant. Further, alpha (α) level was set to 0.05 for all tests. For both targeted outcomes (daily LTEQ scores of all groups for testing hypothesis 1 and daily step count scores of SG and NSG groups for testing hypothesis 2), the following step-by-step added models will be fitted to the data, respectively: (1) unconditional model with random intercepts, (2) random-intercept fixed-slopes model adding time as predictor, (3) random-intercept random-slopes model, (4) random-intercept random-slopes model adding group as additional predictor (i.e., effects of group on intercepts, Step Goal Group as reference group), and (5) adding group × time interaction. Results of models (1) to (3) are used to describe the data, results of model (4) are used to test the hypotheses (effects of group on intercept) and model (5) to answer the explorative research questions (effects of groups on slopes).

Results

Means and standard deviations of average daily PA scores per group are presented in ▶ **Table 1**. LTEQ scores are based on a to-

Table 1 Means and standard deviations of average daily PA scores (LTEQ scores and step counts)

Measures	C Group		NSG Group		SG Group	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
LTEQ scores	18.47	21.12	16.3	17.48	21.78	23.8
Step count scores	–	–	8974.71	4705.41	9186.95	4552.41
Average daily correlation	–	–	0.39	0.16	0.42	0.17

SD standard deviation, *C* control, *NSG* no step target, *SG* step target, *LTEQ* Leisure-Time Exercise Questionnaire

tal of 5030 units and step counts are based on a total of 3716 units. Visual inspection of group differences suggests that differences generally were small. Within NSG group, we calculated the average daily correlation of LTEQ scores and step count scores. With a significant positive correlation of $0.39 < r < 0.42$ for both groups, these correlations indicate that the measures considerably converge to conclude that both measures refer to the same construct but in parts to different aspects of it.

All details of the results can be found in the supplement as an R Markdown (html file with tabs). Further, the used R script, data and code are open and can be accessed here (osf.io/cgvw4, Utesch, Piesch, Busch, Strauss, & Geukes, 2021). Hence, results are summarized here in text form.

LTEQ

Unconditional model
Fitted to daily LTEQ scores, the first model revealed an observed interclass correlation (ICC) of 0.225, indicating sizeable PA variation both within (77.5%; within-person variance: $SD = 18.55$, 95% confidence interval [CI]: 18.19, 18.93) and between (22.5%; between-person variance: $SD = 10$, 95% CI: 8.85, 11.38) participants supporting the multilevel approach for this data. The grand mean of LTEQ scores was 19.02 ($SE = 0.87$, $p < 0.001$) and significantly different from zero.

Random-intercept model with fixed slopes
Second, *time* (measured in days) was added to the model as level-1 predictor to investigate differences in average LTEQ scores (intercept) and the average slope over time. Model fit did not im-

prove, $\chi^2 (1) = 0.77$, $p = 0.38$ (Δ AIC = 1, Δ BIC = 7). Results of this model indicated that intercepts ($SD = 10$, 95% CI: 8.85, 11.38) varied significantly between participants. Fixed effect estimates indicate that the average LTEQ score is different from zero with a significant intercept of 19.43 ($SE = 0.98$, $p < 0.001$) and a nonsignificant slope of -0.02 ($SE = 0.02$, $p = 0.38$). Accordingly, the daily LTEQ score was not found to change meaningfully across the observation period across all participants (fixed effect).

Random-intercept model with random slopes
Third, random effects of the slope coefficient (i.e., individual slope differences) are introduced to the model. Adding random slopes significantly improved model fit, $\chi^2 (2) = 96.77$, $p < 0.001$ (Δ AIC = –92, Δ BIC = –79). Results of this model indicated that intercepts ($SD = 12.58$, 95% CI: 10.9, 14.55) and slopes ($SD = 0.36$, 95% CI: 0.29, 0.43) varied significantly between participants. The slope–intercept correlation of $\rho 01 = -0.6$ reflects that those participants who had a greater LTEQ score initially tended to experience a steeper decline in LTEQ scores over time, than those starting with lower levels.

Random-intercept model with random slopes and group as predictor
Fourth, adding *group* (SG vs NSG vs C group; fixed effect) as level-2 predictor did improve model fit, $\chi^2 (2) = 6.99$, $p < 0.05$ (Δ AIC = –3, Δ BIC = 10). Adding *group* did not significantly predict LTEQ intercepts between the SG and C group ($b = -3.22$, $p = 0.12$), but did significantly predict LTEQ intercepts between the SG and NSG group ($b = -5.56$, $p < 0.01$). Further, NSG and

C group did not significantly differ ($b = 2.34$, $p = 0.5$). Yet, time effects did not change ($b = -0.02$, $p = 0.62$). As such, hypothesis 1 is not supported by the questionnaire data, but hypothesis 2 is supported by questionnaire data.

Random-intercept model with random slopes and group and group:time interaction as predictor
Finally, testing for an interaction between *time* and *group* did not improve model fit, $\chi^2 (2) = 0.05$, $p = 0.97$ (Δ AIC = 3, Δ BIC = 17), and revealed no interaction effect (i.e., SG and NSG [$b = -0.02$, $p = 0.84$] and the NSG and C group [$b = -0.02$, $p = 0.84$]).

Step count

Unconditional model
Fitted to daily step count scores, the first model indicated an observed ICC of 0.232, i.e., sizeable variation in daily step counts within (76.8%; within-person variance: $SD = 4058.52$, 95% CI: 3966.77, 4153.87) and between (23.2%; between-person variance: $SD = 2232.66$, 95% CI: 1924.72, 2616.99) participants and again supporting the multilevel approach for this data. The grand mean of step count scores was 9097.04 ($SE = 235.83$, $p < 0.001$) and significantly different from zero.

Random-intercept model with fixed slopes
Second, when *time* was added as level-1 predictor, model fit did not improve, $\chi^2 (1) = 0.12$, $p = 0.73$ (Δ AIC = 2, Δ BIC = 8). Results indicated that intercepts ($SD = 2232.83$, 95% CI: 1924.84, 2617.22) varied significantly between participants. Fixed effect estimates indicate that the average step count score is different from zero with a significant intercept of 9055.16 ($SE = 265.13$, $p < 0.001$) and a nonsignificant slope of 2.02 ($SE = 5.85$, $p = 0.73$). As such, the daily step count score did not change and would—even if significant—on average increase only slightly by 2.02 steps each day of the study period.
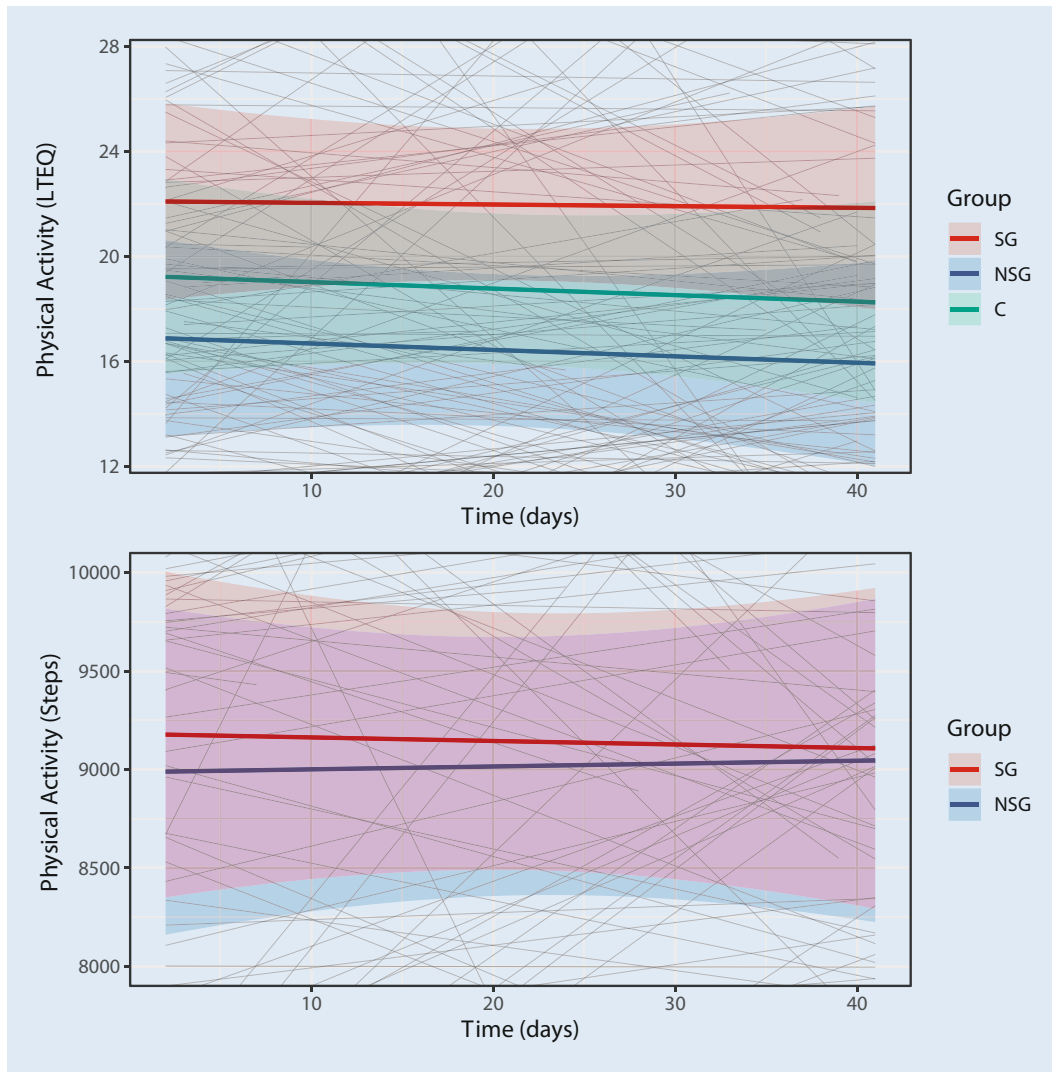
Fig. 1 ◄ Group differences for physical activity questionnaire based (*top*, LTEQ) and actual steps (*bottom*) during the six week period. Groups: *C* control, *SG* step target, *NSG* no step target

Random-intercept model with random slopes

Third, random effects of the slope coefficient (i.e., individual slope differences) are introduced to the model. Compared to the model with fixed slopes, model fit significantly improved, χ^2 (2) = 54.89, $p < 0.001$ (Δ AIC = −51, Δ BIC = −38). Results of this model indicated that intercepts (*SD* = 2757.05, 95% CI: 2319.76, 3290.65) and slopes (*SD* = 70.42, 95% CI: 54.72, 88.49) varied significantly between participants. The slope–intercept correlation of $\rho 01 = -0.59$ reflects that those participants, who walked more steps a day initially, tended to experience a steeper decline in steps, than those starting with fewer steps.

Random-intercept model with random slopes and group as predictor

Fourth, adding *group* (SG vs NSG group; fixed effect) as level-2 predictor did not improve model fit, χ^2 (1) = 0.07, $p = 0.79$ (Δ AIC = 2, Δ BIC = 8). Results indicated that group did not significantly predict step count score intercepts (b = −123.82, $p > 0.05$), and neither did group significantly predict different step count score trajectories (b = −0.17, $p = 0.79$). Hence, hypothesis 2 is not supported by the data.

Random-intercept model with random slopes and group and group:time interaction as predictor

Lastly, adding an interaction between *group* and *time* neither improved model fit, χ^2 (1) = 0.03, $p = 0.86$ (Δ AIC = 2, Δ

BIC = 8) nor disclosed a significant interaction effect (b = 3.26, $p = 0.86$; ◘ **Fig. 1**).

Discussion

The purpose of this ambulatory assessment study realized as randomized-controlled, parallel group trial was to examine whether the use of activity self-tracking with an accompanying smartphone application, alone and in combination with a step goal, promotes individuals' daily physical activity (PA). Across a 6-week period, participants' PA levels were recorded via a questionnaire (experimental groups and control group) and a fitness tracker (FT) with an app (experimental groups), allowing the comparisons of (a) FT-based PA levels (and trajectories) of a group using FTs against a group using FTs with an additional daily step

target of 10,000 steps, and (b) questionnaire-based PA levels (and trajectories) of these groups against a control group. Findings indicated that neither had standalone activity tracking (AT) nor had the additional step goal a significant effect on objective PA levels and trajectories. Notably, the additional step goal only had an effect on subjective PA levels indicated by questionnaire.

Based on the Social Cognitive Theory (SCT) of self-regulation (Bandura, 1991), it was expected that individuals' self-reflective and self-reactive capabilities (i.e., based on self-insight provided by the FT) would assist them to exert control over their thoughts, emotions, motivation, and actions, that is, over their daily PA. Specifically, activity self-tracking was thought to provide self-insights, which, in turn, provide the basis to infer that current activity should be increased, and to initiate a process of corrective behavior change into the desired direction of more PA (Bandura, 1991; Kersten-van Dijk et al., 2017; Shull et al., 2014; Stiglbauer et al., 2019). In addition, goal setting was thought to aid this behavior change process by providing individuals with the opportunity to evaluate the provided PA self-insight information against a normative standard (Bandura, 1991; Stiglbauer et al., 2019; Sullivan & Lachman, 2017). These study's findings could suggest that the "information" obtained from the mere awareness of daily PA being assessed, as also present in the control group, may be stimulating enough to induce such self-reflective processes, but not enough to boost individuals' efforts to move. However, in conjunction with goal setting, AT was potent enough to elicit increases in self-reported PA (i.e., LTEQ scores), suggesting that self-monitoring using FTs with the additional opportunity of evaluating current PA by means of a step goal elevates perceived PA levels. Yet, these findings did not persist when tested based on an objective measure of PA (i.e., step count). It is probable that participants overestimated their PA levels out of a desire to attain and report a PA level that approximates the one which has been assigned to them via the step goal. Thus, the step goal may have induced intentional misinterpreta-

tion and biased self-perception of PA. This is partly also referred to by Bandura (1991) stating that self-monitoring of behavior that bears on self-esteem (e.g., when conceding that the desired goal is not met) elicits affective reactions which can distort self-perceptions of a behavior during later recollections of it.

Importantly, these conclusions need to be drawn on a group level, but findings—within and across groups—suggest that there are substantial individual differences, both in PA levels and trajectories over the 6-week period. As such, one core conclusion is that individuals differ in their responsiveness vs. reluctance to such PA-related interventions. While some individuals might respond to such interventions with the desired behavior change and increased levels of daily PA, others might just keep up with their routine, missing the aim of PA increase or even lowering their PA levels over time. Thus, one crucial task for future research will be to identify boosting and boundary conditions for the individual success of such and comparable PA interventions. In this regard, additional incentives such as monetary rewards (Finkelstein et al., 2016), game design elements (Patel et al., 2017) or SMS (short message service)-based prompting (Wang et al., 2015) have already shown potential to increase PA engagement in FT-based interventions. Moreover, a focus on individual difference variables predicting differences in responsiveness vs. reluctance might be promising to customize interventions to either specific groups of individuals or even the specific individual. With respect to externally assigned step goals, a goal that is static and not tailored to an individual may fail to account for interindividual differences in response rates and PA levels, thereby restraining PA-enhancing effects. In line with Bandura (1991), externally assigned goals may be less meaningful to the user and lack personal significance, decreasing the probability of goal attainment. Adjustable goals which adapt to individual response rates and PA levels and take variations in context (e.g., weather, work) and prior goal achievement (e.g., was the goal of the previous day met or not?) into consideration may represent a promising alternative. This

is supported by previous research (e.g., Korinek et al., 2018). Comparing constant vs. adjustable goals in randomized controlled trials may constitute an object of future studies.

Based on the present findings, FT users are advised to not solely rely on activity monitoring as means for increasing PA behavior. Likewise, adhering to default step goal settings like 10,000 steps per day to boost PA levels is not supported by the data and thus not recommended as best strategy. While activity monitoring may be conducive to PA awareness and evoke self-reflective processes, the necessity to consult additional incentives and intervention components besides AT to yield sizable changes in PA behavior can be a valuable implication for health agencies promoting PA or individuals seeking to increase current PA levels.

Limitations and future directions

Despite this study's strengths, that is, being theory-based, adhering to design recommendations previously requested (e.g., Jee, 2017; Schoeppe et al., 2016) such as using a longitudinal, control-compared, randomized trial with larger sample sizes with a large number of individual observations, and using multilevel modeling, it does not come without limitations. First, participants were WEIRD (western, educated, industrialized, rich and democratic) and mostly female, as such, the generalizability is limited and needs to be tested to other populations in future studies. Second, the study period was limited to 6 weeks, so that effects beyond this time frame could not be investigated. Third, subjective reports of PA, as based on a daily version of the LTEQ, may not reflect a perfectly accurate but biased representation of PA. Daily self-reports are necessarily subject to the ability of accurately memorizing PA and reporting over a day and these reports typically become less precise when a questionnaire of PA is completed on several occasions (Shephard, 2003). There was, however, a substantial and significant correlation with step count as more objective measure, indicating criterion validity. Fourth, although the step count measure for daily PA represents a more

objective estimate of PA, it still comes with disadvantages. On the one hand, it does not take the intensity of an activity into account. People may, for instance, meet the 10,000 steps a day recommendation by Tudor-Locke et al. (2011), yet, at the same time, this may not meet the current PA guidelines for its intensity (Sullivan & Lachman, 2017). On the other hand, it does not take all instances of PA into account. Because, in some sports (e.g., soccer, basketball), it is not allowed to wear wristbands during training or matches, actual PA might have been underestimated for some participants involved in such sports. Furthermore, the fact that participants were able to monitor other outcomes besides step counts may have affected participants' behavior (e.g., participants may not see the necessity to become active when consumed calories are low). Lastly, it needs to be mentioned that the control group was aware of the PA assessment purpose of the study and could not be randomized, so it cannot be ruled out that they increased their daily PA, compared to outside the study, just because of their participation and according awareness.

Future studies are encouraged to test more diverse samples, to increase time frames, and to assess multiple measures of PA, involving intensity information, to increase the knowledge on their validities, respectively. Moreover, approaches should take individual differences into account, enabling customized intervention programs with greater success rates than one-fits-all solutions. Lastly, as informative intervention studies necessarily assess hierarchical, that is, longitudinal, data structure with time points nested in participants, it is well advised to use multilevel modeling as statistical approach to answer the research questions at hand.

Conclusion

This study evaluated the effects of activity self-tracking, alone and with combination of a daily step goal of 10,000 steps, on PA levels and trajectories against those of a control group. Findings indicated that, irrespective of the group, daily PA levels and trajectories were similar suggesting that both forms of self-tracking interventions were unsuccessful. Findings also indicated, however, that there are substantial individual differences in daily PA levels and trajectories that might help understand individual boosting and boundary conditions in being responsive or reluctant to such interventions and ultimately increase PA and health-related behaviors in general.

Corresponding address

Till Utesch
Institute for Educational Sciences, University of Münster
Bispinghof 5/6, 48143 Münster, Germany
till.utesch@uni-muenster.de

© WWU-Lukas Walbaum

Declarations

References

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational behavior and human decision processes*, *50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-L.

Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, *2*(1), 74.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., & Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. *Jama*, *298*(19), 2296–2304. https://doi.org/10.1001/jama.298.19.2296.

Brickwood, K. J., Watson, G., O'Brien, J., & Williams, A. D. (2019). Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. *JMIR mHealth and uHealth*, *7*(4), e11819. https://doi.org/10.2196/11819.

Bull, F. C., Al-Ansari, S. S., Biddle, S., Borodulin, K., Buman, M. P., Cardon, G., & Willumsen, J. F. (2020). World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *British journal of sports medicine*, *54*(24), 1451–1462. https://doi.org/10.1136/bjsports-2020-102955.

Busch, L., Utesch, T., Bürkner, P. C., & Strauss, B. (2020a). The influence of fitness-app usage on psychological well-being and body awareness—a daily diary randomized trial. *Journal of Sport and Exercise Psychology*, *42*(3), 249–260. https://doi.org/10.1123/jsep.2019-0315.

Busch, L., Utesch, T., & Strauss, B. (2020b). Normalised step targets in fitness apps affect users' autonomy need satisfaction, motivation and physical activity—a six-week RCT. *International Journal of Sport and Exercise Psychology*. https://doi.org/10.1080/1612197X.2020.1854820.

Cadmus-Bertram, L. A., Marcus, B. H., Patterson, R. E., Parker, B. A., & Morey, B. L. (2015). Randomized trial of a Fitbit-based physical activity intervention for women. *American journal of preventive medicine*, *49*(3), 414–418. https://doi.org/10.1016/j.amepre.2015.01.020.

Diaz, K. M., Krupka, D. J., Chang, M. J., Peacock, J., Ma, Y., Goldsmith, J., & Davidson, K. W. (2015). Fitbit®: An accurate and reliable device for wireless physical activity tracking. *International journal of cardiology*, *185*, 138. https://doi.org/10.1016/j.ijcard.2015.03.038.

Duncan, M. J., Brown, W. J., Mummery, W. K., & Vandelanotte, C. (2018). *10,000 Steps Australia: a community-wide eHealth physical activity promotion programme*. https://doi.org/10.1136/bjsports-2017-097625.

Finkelstein, E. A., Haaland, B. A., Bilger, M., Sahasranaman, A., Sloan, R. A., Nang, E. E. K., & Evenson, K. R. (2016). Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. *The lancet Diabetes & endocrinology*, *4*(12), 983–995. https://doi.org/10.1016/S2213-8587(16)30284-4.

Fitbit Inc (2018). Fitbit flex 2: user manual. https://help.fitbit.com/manuals/manual_flex_2_en_US.pdf. Accessed 30.07.2021.

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), 1–15. https://doi.org/10.1186/s40798-018-0157-9.

Giddens, L., Leidner, D., & Gonzalez, E. (2017). The role of Fitbits in corporate wellness programs: does step count matter? *Proceedings of the 50th Hawaii International Conference on System Sciences 2017*. 3627–3635. https://doi.org/10.24251/hicss.2017.438.

Gionet, N. J., & Godin, G. (1989). Self-reported exercise behavior of employees: a validity study. *Journal of Occupational Medicine.: Official Publication of the Industrial Medical Association*, *31*(12),

969–973. https://doi.org/10.1097/00043764-198912000-00007.

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Sciences*, *10*(3), 141–146.

Goodyear, V. A., Kerner, C., & Quennerstedt, M. (2017). Young people's uses of wearable healthy lifestyle technologies; surveillance, self-surveillance and resistance. *Sport, Education and Society*, *24*(3), 212–225. https://doi.org/10.1080/13573322.2017.1375907.

Heath, G. W., Parra, D. C., Sarmiento, O. L., Andersen, L. B., Owen, N., Goenka, S., & Lancet Physical Activity Series Working Group (2012). Evidence-based intervention in physical activity: lessons from around the world. *The lancet*, *380*(9838), 272–281. https://doi.org/10.1016/S0140-6736(12)60816-2.

Hohepa, M., Schofield, G., Kolt, G. S., Scragg, R., & Garrett, N. (2008). Pedometer-determined physical activity levels of adolescents: differences by age, sex, time of week, and transportation mode to school. *Journal of physical activity and health*, *5*(s1), S140–S152. https://doi.org/10.1123/jpah.5.s1.s140.

Jacobs Jr, D. R., Ainsworth, B. E., Hartman, T. J., & Leon, A. S. (1993). A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Medicine and science in sports and exercise*, *25*(1), 81–91. https://doi.org/10.1249/00005768-199301000-00012.

Jee, H. (2017). Review of researches on smartphone applications for physical activity promotion in healthy adults. *Journal of exercise rehabilitation*, *13*(1), 3. https://doi.org/10.12965/jer.1732928.464.

Kersten-van Dijk, E. T., Westerink, J. H., Beute, F., & IJsselsteijn, W. A. (2017). Personal informatics, self-insight, and behavior change: a critical review of current literature. *Human–Computer Interaction*, *32*(5–6), 268–296. https://doi.org/10.1080/07370024.2016.1276456.

Kohl 3rd, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., & Lancet Physical Activity Series Working Group (2012). The pandemic of physical inactivity: global action for public health. *The lancet*, *380*(9838), 294–305. https://doi.org/10.1016/S0140-6736(12)60898-8.

Kooiman, T. J., Dontje, M. L., Sprenger, S. R., Krijnen, W. P., Van der Schans, C. P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, *7*(1), 1–11. https://doi.org/10.1097/CIN.0000000000000443.

Korinek, E. V., Phatak, S. S., Martin, C. A., Freigoun, M. T., Rivera, D. E., Adams, M. A., & Hekler, E. B. (2018). Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. *Journal of behavioral medicine*, *41*(1), 74–86. https://doi.org/10.1007/s10865-017-9878-3.

Kreft, I. G. G., & de Leeuw, J. (1998). *Introducing multilevel modeling*. SAGE.

Krzysztoszek, J., Laudanska-Krzeminska, I., & Bronikowski, M. (2019). Assessment of epidemiological obesity among adults in EU countries. *Annals of Agricultural and Environmental Medicine*. https://doi.org/10.26444/aaem/97226.

Lakerveld, J., Woods, C., Hebestreit, A., Brenner, H., Flechtner-Mors, M., Harrington, J. M., & Ahrens, W. (2020). Advancing the evidence base for public policies impacting on dietary behaviour, physical activity and sedentary behaviour in Europe: The Policy Evaluation Network promoting a multidisciplinary approach. *Food Policy*, *96*, 101873. https://doi.org/10.1016/j.foodpol.2020.101873.

Laska, M. N., Pelletier, J. E., Larson, N. I., & Story, M. (2012). Interventions for weight gain prevention during the transition to young adulthood: a review of the literature. *Journal of Adolescent Health*, *50*(4), 324–333. https://doi.org/10.1016/j.jadohealth.2012.01.016.

Loomba, S., & Khairnar, A. (2017). Fitness trackers market by device type (fitness bands, smartwatch, and others), display type (monochrome and colored), sales channel (online and offline), and compatibility (iOS, Android, Windows, Tizen, and others). https://www.alliedmarketresearch.com/fitness-tracker-market Global Opportunity Analysis and Industry Forecast, 2023, 2018. Accessed 30.07.2021.

Lynch, C., Bird, S., Lythgo, N., & Selva-Raj, I. (2020). Changing the physical activity behavior of adults with fitness trackers: a systematic review and meta-analysis. *American Journal of Health Promotion*, *34*(4), 418–430. https://doi.org/10.1177/0890117119895204.

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), e192. https://doi.org/10.2196/jmir.3469.

McDermott, M. M., Spring, B., Berger, J. S., Treat-Jacobson, D., Conte, M. S., Creager, M. A., & Rejeski, W. J. (2018). Effect of a home-based exercise intervention of wearable technology and telephone coaching on walking performance in peripheral artery disease: the HONOR randomized clinical trial. *JAMA*, *319*(16), 1665–1676. https://doi.org/10.1001/jama.2018.3275.

Melton, B. F., Buman, M. P., Vogel, R. L., Harris, B. S., & Bigham, L. E. (2016). Wearable devices to improve physical activity and sleep: a randomized controlled trial of college-aged African American women. *Journal of Black Studies*, *47*(6), 610–625. https://doi.org/10.1177/F0021934716653349.

Nelson, M. C., Story, M., Larson, N. I., Neumark-Sztainer, D., & Lytle, L. A. (2008). Emerging adulthood and college-aged youth: an overlooked age for weight-related behavior change. *Obesity*, *16*(10), 2205. https://doi.org/10.1038/oby.2008.365.

Patel, M. S., Benjamin, E. J., Volpp, K. G., Fox, C. S., Small, D. S., Massaro, J. M., & Murabito, J. M. (2017). Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. *JAMA internal medicine*, *177*(11), 1586–1593. https://doi.org/10.1001/jamainternmed.2017.3458.

van Poppel, M. N., Chinapaw, M. J., Mokkink, L. B., Van Mechelen, W., & Terwee, C. B. (2010). Physical activity questionnaires for adults. *Sports medicine*, *40*(7), 565–600. https://doi.org/10.2165/11531930-000000000-00000.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods*. Vol. 1.

Romeo, A., Edney, S., Plotnikoff, R., Curtis, R., Ryan, J., Sanders, I., Crozier, A., & Maher, C. (2019). Can smartphone apps increase physical activity? Systematic review and meta-analysis. *Journal of Medical Internet Research*, *21*(3), e12053. https://doi.org/10.2196/12053.

Schoeppe, S., Alley, S., Van Lippevelde, W., Bray, N. A., Williams, S. L., Duncan, M. J., & Vandelanotte, C. (2016). Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *13*(1), 1–26. https://doi.org/10.1186/s12966-016-0454-y.

Shephard, R. (1997). Godin leisure-time exercise questionnaire. *Medicine and Science in Sports and Exercise*, *29*(6), S36–S38. https://doi.org/10.1097/00005768-199706001-00009.

Shephard, R. J. (2003). Limits to the measurement of habitual physical activity by questionnaires. *British journal of sports medicine*, *37*(3), 197–206. https://doi.org/10.1136/bjsm.37.3.197.

Shull, P. B., Jirattigalachote, W., Hunt, M. A., Cutkosky, M. R., & Delp, S. L. (2014). Quantified self and human movement: a review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. *Gait & posture*, *40*(1), 11–19. https://doi.org/10.1016/j.gaitpost.2014.03.189.

Silva, S. S. M., Meyer, D., & Jayawardana, M. (2020). Detecting possible persons of interest in a physical activity program using step entries: Including a web-based application for outlier detection and decision-making. *Biometrical Journal*, *62*(2), 414–427. https://doi.org/10.1002/bimj.201900008.

Snijders, T. A., & Bosker, R. J. (1999). *An introduction to basic and advanced multilevel modeling*. SAGE.

Statista (2021). Number of connected wearable devices worldwide by region from 2015 to 2022. Retrieved from. https://www.statista.com/statistics/490231/wearable-devices-worldwide-by-region/. Accessed 30.07.2021.

Stiglbauer, B., Weber, S., & Batinic, B. (2019). Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial. *Computers in Human Behavior*, *94*, 131–139. https://doi.org/10.1016/j.chb.2019.01.018.

Sullivan, A. N., & Lachman, M. E. (2017). Behavior change with fitness technology in sedentary adults: a review of the evidence for increasing physical activity. *Frontiers in public health*, *4*, 289. https://doi.org/10.3389/FPUBH.2016.00289.

Tudor-Locke, C., Craig, C. L., Brown, W. J., Clemes, S. A., De Cocker, K., Giles-Corti, B., & Blair, S. N. (2011). How many steps/day are enough? For adults. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 1–17. https://doi.org/10.1186/1479-5868-8-79.

Utesch, T., Piesch, L., Busch, L., Strauss, B., & Geukes, K. (2021). Self-tracking of daily physical activity using Fitbits and the effect of the 10,000 steps goal: A six-week randomized controlled parallel group trial. *October*., Retrieved from osf.io/cgvw4.

Vandelanotte, C., Müller, A. M., Short, C. E., Hingle, M., Nathan, N., Williams, S. L., & Maher, C. A. (2016). Past, present, and future of eHealth and mHealth research to improve physical activity and dietary behaviors. *Journal of nutrition education and behavior*, *48*(3), 219–228. https://doi.org/10.1016/j.jneb.2015.12.006.

Venetsanou, F., Emmanouilidou, K., Soutos, K., Sotiriou, S. A., Bastida, L., Moya, A., & Kambas, A. (2020).

Towards a functional approach to the assessment of daily life physical activity in children: are the PAQ-C and Fitbit flex-2 technically adequate? *International Journal of Environmental Research and Public Health*, *17*(22), 8503. https://doi.org/10.3390/ijerph17228503.

Wang, J. B., Cadmus-Bertram, L. A., Natarajan, L., White, M. M., Madanat, H., Nichols, J. F., & Pierce, J. P. (2015). Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. *Telemedicine and e-Health*, *21*(10), 782–792. https://doi.org/10.1089/tmj.2014.0176.

Whelan, M. E., Orme, M. W., Kingsnorth, A. P., Sherar, L. B., Denton, F. L., & Esliger, D. W. (2019). Examining the use of glucose and physical activity self-monitoring technologies in individuals at moderate to high risk of developing type 2 diabetes: randomized trial. *JMIR mHealth and uHealth*, *7*(10), e14195. https://doi.org/10.2196/14195.

World Health Organization (2010). *Global recommendations on physical activity for health*. https://doi.org/10.1080/11026480410034349.