CHAPTER 4
THE DEVELOPMENT OF ACTIVE2GETHER
APP-BASED INTERVENTION COMBINING EVIDENCE-BASED BEHAVIOR CHANGE TECHNIQUES WITH A MODEL-BASED REASONING SYSTEM TO PROMOTE PHYSICAL ACTIVITY AMONG YOUNG ADULTS (ACTIVE2GETHER): DESCRIPTIVE STUDY OF THE DEVELOPMENT AND CONTENT

Anouk Middelweerd
Saskia J te Velde
Julia S Mollee
Michel CA Klein
Johannes Brug

APP-BASED INTERVENTION COMBINING EVIDENCE-BASED BEHAVIOR CHANGE TECHNIQUES WITH A MODEL-BASED REASONING SYSTEM TO PROMOTE PHYSICAL ACTIVITY AMONG YOUNG ADULTS (ACTIVE2GETHER): A DESCRIPTIVE STUDY OF THE DEVELOPMENT AND CONTENT

Anouk Middelweerd
Saskia J te Velde
Julia S Mollee
Michel CA Klein
Johannes Brug

Chapter 4

ABSTRACT

Background
The Active2Gether intervention is an app-based intervention designed to help and encourage young adults to become and remain physically active by means of personalized, real-time activity tracking and context-specific feedback.

Objectives
The objective of our study was to describe the development and content of the Active2Gether intervention for physical activity promotion.

Methods
A systematic and stepwise approach was used to develop the Active2Gether intervention. This included formulating objectives and a theoretical framework, selecting behavior change techniques (BCTs), specifying the tailoring, pilot testing, and describing an evaluation protocol.

Results
The development of the Active2Gether intervention comprised seven steps: analyzing the (health) problem, developing a program framework, writing (tailored) messages, developing tailoring assessments, developing the Active2Gether intervention, pilot testing, and testing and evaluating the intervention. The primary objective of the intervention was to increase the total time spent in moderate-vigorous physical activity for those who do not meet the Dutch guideline, maintain physical activity levels of those who meet the guideline, or further increase physical activity levels if they so indicated. The theoretical framework is informed by the social cognitive theory, and insights from other theories and evidence were added for specific topics. Development of the intervention content and communication channel resulted in the development of an app that provides highly tailored coaching messages that are framed in an autonomy-supportive style. These coaching messages include BCTs aiming to address relevant behavioral determinants (e.g. self-efficacy and outcome expectations) and are partly context specific. A model-based reasoning engine has been developed to tailor the intervention with respect to the type of support provided by the app, send relevant and context-specific messages to the user, and tailor the graphs displayed in the app. For the input of the tailoring, different instruments and sensors are used, such as an activity monitor (Fitbit One), Web-based and mobile questionnaires, and the location services on the user’s mobile phone.

Conclusions
The systematic and stepwise approach resulted in an intervention that is based on theory and input from end users. The use of a model-based reasoning system to provide context-specific coaching messages goes beyond many existing eHealth and mHealth interventions.
INTRODUCTION

Insufficient physical activity is a risk factor for avoidable burden of disease. About 25% of the adult population worldwide and around 50% in many western countries such as the US and the Netherlands do not meet the recommended guidelines for physical activity. Moreover, engagement in moderate-vigorous physical activity decreases with age, in particular when transitioning from adolescence into (young) adulthood.

In general, health promotion interventions informed by established health behavior theory have been found to be associated with higher effect sizes than interventions not based on theory. Research examining the determinants of physical activity mainly focuses on social cognitive and social ecological factors. Social cognitive theories and models, such as the health belief model, the theory of planned behavior, and the social cognitive model, have been developed to explain health behaviors and guide health behavior research and behavior change. Although these models mainly focus on intrapersonal and interpersonal factors, social ecological models more explicitly recognize that behavior may also be strongly influenced by contextual factors, such as the sociocultural and physical environments people live in; for example, Sallis et al proposed a framework recognizing that individuals are physically active within different domains (e.g. recreation, transport, household, and occupation), where different factors on multiple levels influence their overall physical activity behavior. Thus, interventions that aim to increase levels of physical activity should not only target intra- and interpersonal factors but also take their physical and social environments into account.

Besides interventions being informed by theory, interventions are more likely to be effective when established behavior change techniques (BCTs) are incorporated. More specifically, interventions that included a self-monitoring feature in combination with features such as prompting intention formation, specific goal setting, providing feedback on performance, or reviewing behavioral goals were significantly more effective at promoting physical activity and healthy eating than interventions that did not include these BCTs.

Systematic reviews further showed that Information and communications technology (ICT)-supported, individually tailored interventions are superior to generic interventions in promoting physical activity and user engagement and appreciation. Moreover, Krebs et al demonstrated that dynamic tailoring (i.e. iteratively assessing and providing feedback) was associated with larger effect sizes than static tailoring (i.e. all feedback is based on one baseline assessment). Additionally, Rabbi et al reported promising results when using machine-learning techniques to automatically create contextualized and personalized feedback to increase levels of physical activity. Modern technology,
such as smartphones, smartphone apps, and activity trackers, offer new possibilities in health promotion, especially in young adults, of whom the majority owns a smartphone \(^{150, 151}\). Furthermore, the rapid growth of the popularity and variety of health and fitness apps and activity trackers suggest that young adults will appreciate and adopt an app-based physical activity intervention.

Several content analyses have been conducted to identify if and how constructs of behavior change theories and BCTs are incorporated in physical activity promotion apps. Generally, the apps analyzed were lacking applications of behavior change theories and the use of evidence-based BCTs \(^{36, 37, 70-72}\). Moreover, apps mostly provide generic advice or tips about physical activity; gamification, punishment, and context-aware feedback are rare among physical activity apps \(^{152}\). Only a few apps incorporate some form of adaption to the user \(^{152}\). Lastly, existing apps fail to meet the guidelines for physical activity \(^{153, 154}\). Despite the fact that health and fitness apps are popular among smartphone users \(^{155, 156}\), recent research indicates that most presently available apps lack the necessary empirical basis to make a meaningful difference in physical activity promotion \(^{26}\). Thus, those apps are less likely to be effective, and room for improvement exists when using an app to promote physical activity. A recently published systematic review examined studies that used apps in interventions to influence health behavior, including physical activity \(^{157}\). The majority of those studies that targeted adults reported significant short-term intervention effects on levels of physical activity \(^{157}\). Furthermore, the majority of the interventions that reported significant changes in behaviors and health-related outcomes included BCTs, such as goal setting, self-monitoring, and feedback on the performance \(^{157}\).

In summary, innovative ICT-supported mobile technology-based approaches that are evidence based and include dynamic tailoring using intelligent data interpretation techniques may help to effectively support achievement and maintenance of behavior change in the physical activity domain. However, both the empirical basis and dynamic tailoring are lacking in current apps. Thus, physical activity apps that incorporate constructs of behavior change theories and BCTs and provide dynamically tailored feedback are needed. Therefore, we developed the Active2Gether intervention that combines mobile (app-based) technology with dynamically tailored feedback and aims to go beyond existing (mobile) physical activity interventions. The Active2Gether intervention is an app-based intervention designed to help and encourage young adults to become and remain physically active by focusing on the domains of active transport, stair climbing, and sports participation. To do so, participants of the Active2Gether intervention will be categorized into one of the 3 awareness categories (education, coaching, and feedback). Participants in the education category will receive educational messages on the benefits of physical activity, whereas participants in the feedback category will receive motivational messages to maintain their active lifestyle. Participants who are in the coaching category will be coached on sports participation, taking the stairs, or active transport. Every week, the
participants will be asked to choose their coaching domain and to set a weekly goal. Participants will receive a message with a suggestion for a coaching domain and a weekly goal based on their previous behavior, but the final decision will be up to the user. The participants will receive a Fitbit One activity tracker that can be synced to the Active2Gether app and will allow the participants to monitor their physical activity behavior through the Active2Gether app. Additionally, participants will receive (daily) coaching messages addressing relevant behavioral determinants. The content of the messages will be tailored to the user’s behavioral determinants, occupational status, and weather. Lastly, the intervention offers the opportunity to monitor and compare the behaviors with those of other Active2Gether participants because the app will display the activity data of the participant, including a graph displaying the activity data of 6 other participants, preferably friends. The graph with the activity data of others will rank the participants based on their step activity and the user preferences for social comparison (i.e. upward or downward comparison). Taking this preference into account does influence the effectiveness of social comparison as a behavior change technique. The aim of this paper was to describe the systematic development and content of this Active2Gether physical activity-promotion intervention. The methods section provides a brief overview of the stepwise approach that was used to develop the intervention and a brief description of the target population and the methodology used to develop the intervention. The results section will provide more detailed information on the results of the systematic development and content of the intervention.

**METHODS**

**Target Population**

The Active2Gether intervention focuses on healthy and highly educated young adults aged 18-30 years who have a suitable smartphone running on Android version 4.0 or higher.

**Intervention Development**

We used a 7-step systematic approach to develop and evaluate the intervention (Table 4.1). To ensure that the app was informed by relevant health behavior and health behavior change theory and evidence, the development was guided by the program-planning model developed by Kreuter et al. Some steps were adapted because it felt more logical to the research team, and the order of some steps were changed; for example, creating tailoring algorithms, automating the tailoring process, and developing the communication channel are described in the same step. The 7 steps are further described in the Results sections.
Table 4.1 - Description of the stepwise process for the development of Active2Gether.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Step description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Analyzing the (health) problem</td>
<td>Describing a theoretical framework on how to promote MVPA&lt;sup&gt;a&lt;/sup&gt; Selecting behavior change techniques based on theory and evidence to address determinants of behavior, based on existing studies and reviews&lt;sup&gt;55, 160&lt;/sup&gt; Assessing existing apps (what is available?)&lt;sup&gt;70&lt;/sup&gt; Exploring preferences of end users&lt;sup&gt;76, 161&lt;/sup&gt;</td>
</tr>
<tr>
<td>Step 2: Developing a Program Framework</td>
<td>Identifying relevant physical activity behaviors to increase MVPA. Defining the main and subobjectives of the intervention Describing framework components</td>
</tr>
<tr>
<td>Step 3: Writing (tailored) messages (the order of this step was changed: Step 5)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Writing tailored messages</td>
</tr>
<tr>
<td>Step 4: Developing tailoring assessments (the order of this step was changed: Step 3)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Selecting and developing measurements to assess levels of physical activity, behavioral determinants, locations, and connected friends</td>
</tr>
<tr>
<td>Step 5: Developing the Active2Gether intervention (steps were merged)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Designing tailoring algorithms for the reasoning system Channel of communication: building a Web-based app and system to combine and interpret data and send messages</td>
</tr>
<tr>
<td>Step 6: Pilot testing</td>
<td>Pilot-testing the intervention to detect errors and impracticalities in order to improve the intervention prior to its implementation</td>
</tr>
<tr>
<td>Step 7: Testing and evaluating the intervention</td>
<td>The intervention will be used by a larger group of participants and then analyzed and evaluated with respect to effect, process, and impact</td>
</tr>
</tbody>
</table>

<sup>a</sup>MVPA: moderate-vigorous physical activity.

<sup>b</sup>According to the program-planning model by Kreuter et al<sup>159</sup>, the tailored messages should be written in step 5, whereas the tailoring assessments should be developed in step 3.

<sup>c</sup>Creating tailoring algorithms, automating the tailoring process, and developing the communication channel are described in the same step, whereas according to the program-planning model, these are steps 6 and 7, respectively.

Step 1: Analyzing the (Health) Problem

Identifying Determinants of Change and Reviewing Applicable Theories and Models

Because theory-based interventions are associated with higher effect sizes than interventions not based on theory<sup>5, 28</sup>, defining the theoretical framework for the intervention is necessary. To do so, prominent health behavior theories and scientific literature were reviewed.

Social cognitive theory (SCT) was adopted as a basis for the theoretical framework as it is one of the most prominent behavior change theories used to inform interventions targeting health behavior change<sup>21-23</sup>, and a recent meta-analysis reported that SCT concepts may explain 31% of variance in physical activity<sup>22</sup>. SCT addresses both individual and social factors and recognizes the reciprocal relation between individuals and their context or environment. For these reasons, SCT thus guided and informed the intervention’s theoretical framework; insights from other theories and evidences were added for specific topics. Figure 4.1 shows the structural pathways of Bandura’s SCT<sup>24</sup>, and Figure 4.2 shows the specific theoretical framework<sup>24</sup> that is used for the Active2Gether intervention.
The development of Active2Gether

Table 4.1 - Description of the stepwise process for the development of Active2Gether.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1: Analyzing the (health) problem</td>
<td>Describing a theoretical framework on how to promote MVPA</td>
</tr>
<tr>
<td></td>
<td>Selecting behavior change techniques based on theory and evidence to address determinants of behavior, based on existing studies and reviews.</td>
</tr>
<tr>
<td></td>
<td>Assessing existing apps (what is available?).</td>
</tr>
<tr>
<td></td>
<td>Exploring preferences of end users.</td>
</tr>
<tr>
<td>Step 2: Developing a Program</td>
<td>Identifying relevant physical activity behaviors to increase MVPA.</td>
</tr>
<tr>
<td></td>
<td>Defining the main and subobjectives of the intervention.</td>
</tr>
<tr>
<td></td>
<td>Describing framework components.</td>
</tr>
<tr>
<td>Step 3: Writing (tailored) messages</td>
<td>Writing tailored messages.</td>
</tr>
<tr>
<td>Step 4: Developing tailoring assessments</td>
<td>Selecting and developing measurements to assess levels of physical activity, behavioral determinants, locations, and connected friends.</td>
</tr>
<tr>
<td>Step 5: Developing the Active2Gether</td>
<td>Designing tailoring algorithms for the reasoning system.</td>
</tr>
<tr>
<td></td>
<td>Channel of communication: building a Web-based app and system to combine and interpret data and send messages.</td>
</tr>
<tr>
<td>Step 6: Pilot testing</td>
<td>Pilot-testing the intervention to detect errors and impracticalities in order to improve the intervention prior to its implementation.</td>
</tr>
<tr>
<td>Step 7: Testing and evaluating the</td>
<td>The intervention will be used by a larger group of participants and then analyzed and evaluated with respect to effect, process, and impact.</td>
</tr>
</tbody>
</table>

**A** MVPA: moderate-vigorous physical activity.

**B** According to the program-planning model by Kreuter et al., the tailored messages should be written in step 5, whereas the tailoring assessments should be developed in step 3.

**C** Creating tailoring algorithms, automating the tailoring process, and developing the communication channel are described in the same step, whereas according to the program-planning model, these are steps 6 and 7, respectively.

### Step 1: Analyzing the (Health) Problem

#### Identifying Determinants of Change and Reviewing Applicable Theories and Models

Because theory-based interventions are associated with higher effect sizes than interventions not based on theory, defining the theoretical framework for the intervention is necessary. To do so, prominent health behavior theories and scientific literature were reviewed. Social cognitive theory (SCT) was adopted as a basis for the theoretical framework as it is one of the most prominent behavior change theories used to inform interventions targeting health behavior change, and a recent meta-analysis reported that SCT concepts may explain 31% of variance in physical activity. SCT addresses both individual and social factors and recognizes the reciprocal relation between individuals and their context or environment. For these reasons, SCT thus guided and informed the intervention’s theoretical framework; insights from other theories and evidences were added for specific topics.

![Figure 4.1 - The structural pathways of Bandura’s social cognitive theory](image1)

**Figure 4.1** - The structural pathways of Bandura’s social cognitive theory

![Figure 4.2 - The specific theoretical framework that is used for the Active2Gether intervention](image2)

**Figure 4.2** – The specific theoretical framework that is used for the Active2Gether intervention

*Note.* The bold lines and boxes represent the elements that are based on the Social Cognitive Theory, and the dotted lines and oval boxes represent behavioral determinants added to the theoretical framework.

### Selection of Behavior Change Techniques

We first identified evidence-based and relevant BCTs and linked these with the behavioral determinants of the theoretical framework by means of a review of the relevant literature, based on an existing taxonomy of BCTs ([Appendix 4.1](#)). To explore which BCTs were used in already existing physical activity promotion apps, a systematic content analysis of such apps available in iTunes and Google Play was conducted. Additionally, focus group discussions with the target population were conducted. The methods and results of these focus groups have been published in more detail elsewhere. Finally, a Web-based cross-sectional survey was conducted among 179 young adults to
assess their ratings with respect to the importance of specific BCTs applied in apps and their preferences for personalized tailoring\textsuperscript{161}.

**Step 2: Developing a Program Framework**

*Defining the Intervention’s Primary and Secondary Objectives*

The foundation of the intervention is the definition of the program’s outcomes and objectives\textsuperscript{6, 89}. Therefore, specifying who and what will change as a result of the intervention is necessary\textsuperscript{6, 89}. Intervention objectives were based on the Dutch guidelines for physical activity for adults, which included the following: 30 minutes of moderate physical activity for at least 5 days a week or 20 minutes of vigorous physical activity for 3 days a week\textsuperscript{78}.

*Describing Framework Components*

Based on steps 1 and 2 and the research team’s expertise, a general framework was developed. The aim was to develop a highly tailored intervention that contains a self-monitoring tool, goal setting, social comparison, and motivational and context-specific messages.

Tailoring and personalization of the intervention content is realized in the following 6 ways: determining the personally appropriate type of support (i.e. education, coaching, or feedback), selecting the personally relevant and preferred domain of physical activity for user coaching (i.e. sports participation, stair use, or active transport), suggesting a weekly goal, selecting the personally appropriate behavioral determinants for coaching, sending only relevant coaching messages and filtering out nonrelevant messages, and tailoring and personalization of the app content.

**Step 3: Writing (Tailored) Messages**

Next, we translated the BCTs into actual tailored feedback messages and advice. For each BCT per behavioral determinant (Appendix 4.1), a set of messages was created that was tailored to the three coaching domains (i.e. sports participation, active transport, and stair climbing). Consequently, a message library was created that contained feedback and advice messages tailored to all possible levels of the relevant behavioral determinants, as recognized in the underlying theoretical framework.

Creating the message library was an iterative process of brainstorming, writing a set of messages (AM), and providing feedback and suggestions (JM and StV). To test whether the tone of voice and content appealed to the target population, a subset of messages was pilot-tested among 7 people of the target population.

**Step 4: Developing Tailoring Assessments**

To tailor the messages to the individual users, assessment methods were selected.
Assessment of Activity
First, after considering functionalities, validity, and costs of a range of available activity trackers, the user’s activity was monitored using Fitbit One, which includes monitoring of steps and stairs climbed. Fitbit One was chosen because of its functionalities and small size. The activity monitor communicates with the Fitbit app and website that display the collected data for example by showing a color-coded chart indicating the proximity to the step goal, which is set to a default of 10,000 steps per day.

Fitbit allows developers and researchers to access Fitbit data and thus integrate the Fitbit data into health behavior interventions such as Active2Gether. To access Fitbit data, Fitbit offers an application programming interface (API). Fitbit One was validated using smaller time intervals (i.e. minutes, hours, and days) relevant for real-time feedback and instant behavioral insights to its users. Healthy young adults (N=34) wore the ActiGraph GT3x+ and a Fitbit One for one week. Detailed information on the methodology can be found elsewhere.

Assessment of Behavioral Determinants
Literature was reviewed for relevant, existing, and validated questionnaires to assess behavioral determinants. Behavioral determinants are assessed by means of a questionnaire with both its long and short versions, which were selected based on validations of such questionnaires. The long version is part of an “intake” questionnaire before the actual intervention and as a point of departure for the tailored intervention, whereas the short version is used repeatedly throughout the intervention period to dynamically tailor the intervention content to the user.

Assessment of Location Data
A questionnaire was designed for the purpose of assessing information on significant places. In addition, the Active2Gether app was built in a way that enabled the collection of the user’s location data.

Assessment of Connected Friends
To increase the users’ engagement, we assessed whether users’ friends were also participating in the Active2Gether intervention. Because Facebook is very popular among Dutch young adults—93% of Dutch adults aged 18-24 use Facebook.—Facebook was used to find connected friends that were also participating in the study.
**Step 5: Developing the Active2Gether Intervention**

**Creating Tailored Algorithms**

To realize such tailored coaching, we developed a system that combines detailed behavior monitoring with intelligent data interpretation and model-based predictions. Thus, by combining data from the different sources, the system enables personalization of the coaching strategies to try to achieve the most positive effect on behavior change. Detailed information on the system and the development of the system can be found elsewhere and is not described in the Results section [165].

**Designing and Developing the Communication Channel**

We decided that the communication channel of the Active2Gether intervention should be a smartphone app. The app shows the website in a format that is viewable on smaller screens. Thus, the intervention content was accessible through the app or the website. The research team developed the design template of the smartphone app. Information on the development of the app can be found elsewhere [165].

**Step 6: Pilot Testing**

To detect possible bugs in the system and to assess user friendliness and appreciation, the app was pilot-tested in two steps. First, the Active2Gether team (AM, JM, AMR, StV, and MK) used the initial version of the Active2Gether app. Bugs, nuisances, etc were monitored, listed, and fixed accordingly when and where possible. Second, 7 people from the target population (5 women, 21-28 years old, all highly educated, or studying at the bachelor’s or master’s level) were recruited to use the adjusted version of the app, monitor bugs and nuisances, provide feedback in person, and answer a questionnaire regarding use, user friendliness, and appreciation. The app was further adjusted based on that information.

**Step 7: Testing and Evaluating the Intervention**

The intervention, the Active2Gether app, will be evaluated for its efficacy to change weekly levels of moderate–vigorous in young adults and for the usability of the app.

**RESULTS**

**Step 1: Analyzing the (Health) Problem**

**Identifying Determinants of Change and Reviewing Applicable Theories and Models**

As a result of Step 1, a theoretical framework was built based on the relevant scientific literature (please see further details below). The theoretical framework was subsequently used to develop the
content of the intervention and predict the physical activity behavior of the users so that the intervention content could be tailored to each individual user.

Self-efficacy, a key construct within SCT (and in other health behavior theories)\(^\text{24, 145}\), was adopted as a key construct in Active2Gether. Self-efficacy is defined as someone’s beliefs in his or her own capabilities to perform certain actions needed to achieve a desired outcome. Self-efficacy affects physical activity both directly and indirectly, as seen in Figures 1 and 2. Self-efficacy may influence outcome expectations—one’s beliefs about the positive and negative consequences of one’s behavior, such as participating in physical activities\(^\text{24, 145}\). In other words, people who are more efficacious about being physically active will also be more likely to expect the favorable outcomes of participating in physical activities.\(^\text{24}\) Moreover, self-efficacy may also influence how people perceive potential obstacles and impediments\(^\text{24}\) and may also influence intentions to engage in physical activity behaviors\(^\text{166}\). Goal setting was adopted as a second important basis for change, where goals can be either proximal (i.e. shorter-term intentions to act) or distal (i.e. longer-term goals to achieve something)\(^\text{24, 167}\). Distal goals are goals set for the longer term and they set the course for personal change\(^\text{167}\). According to Bandura\(^\text{24}\), distal or long-term goals can initiate behavior change but are not sufficient to change behavior directly, as seen in Figure 4.1. Goal setting is dependent on the levels of self-efficacy and perceived barriers and opportunities. In line with this notion, a meta-analysis inspired by the action-control framework indicated that 48% of the participants who intended to be physically active failed to do so. Therefore, forming intentions is often not sufficient to realize behavior change; self-regulatory and action-control techniques are needed to support behavioral enactment\(^\text{168}\). A further meta-analysis on effective techniques in healthy eating and physical activity interventions concluded that interventions that offered self-monitoring and addressed self-regulation were more successful in increasing physical activity than interventions not including those techniques\(^\text{56}\). SCT posits that when individuals adapt and revise their behavior, they may adjust their beliefs and goals regarding this behavior\(^\text{24}\). In our theoretical framework, we therefore included “satisfaction,” which is defined as an evaluation of the physical activity behavior.

In line with SCT, we also recognized that the social environment influences behavior through social norms and that performing certain behaviors can evoke social reactions, both positive and negative\(^\text{24}\). In the Active2Gether intervention, we address not only intrapersonal (e.g. lack of motivation and tiredness) and social barriers (e.g. lack of support) but also contextual impediments (e.g. lack of time, weather and travel distance), as seen in Figure 4.2. Lastly, it was decided that users will be categorized based on their awareness of their personal physical activity levels before they will be coached; people who are overly optimistic about their physical activity levels (i.e. who believe they engage in adequate
amounts of physical activity while their data show insufficient levels) will be much less likely to be motivated to increase their physical activity levels.

Selection of Behavior Change Techniques

Content analysis showed that the apps available to date generally lack sufficient incorporation of evidence-based BCTs. BCTs that were applied most often were providing feedback on performance, prompting self-monitoring of behavior, prompting specific goal setting, and planning social support or social change. Additionally, focus group discussions with the target population indicated that participants preferred self-monitoring, goal setting, and a ranking feature but were not willing to share their accomplishments on social media for social comparison and initiating social support. The focus groups further suggested that the Active2Gether app should be highly personalized, have an easy-to-use design and format, include a coaching feature that provides tailored feedback to self-set goals, enable competition with friends by ranking or earning rewards, and include the option to personally customize the application. Finally, a Web-based cross-sectional survey among 179 young adults to assess their ratings with respect to the importance of specific BCTs applied in apps and their preferences for personalized tailoring confirmed the need for a personal coaching feature and showed that BCTs addressing goal setting, goal reviewing, feedback, and self-monitoring were rated as important to be incorporated in an app, whereas social support and social comparison were considered less important. The combined results of the literature review, focus group discussions, and survey guided the selection of BCTs to be included in Active2Gether (Appendix 4.1).

Step 2: Developing a Program Framework

Defining the Intervention’s Primary and Secondary Objectives

Step 2 resulted in the decision to make the following the primary objective of the Active2Gether intervention: increase total time spent in moderate–vigorous for those who do not meet the Dutch guideline, maintain physical activity levels of those who meet the guideline, or further increase physical activity levels if they so indicated. The secondary aims were defined as follows: to increase the underlying specific categories of moderate–vigorous (i.e. minutes of weekly sports participation, weekly numbers of stairs climbed, and weekly minutes of active transport) and to enhance the underlying determinants of the physical activity behaviors.

Describing Framework Components

The framework contained information on the levels of tailoring and an outline of the steps taken to deliver tailored messages. Detailed information on the framework components can be found in Appendix 4.2.
Step 3: Writing (Tailored) Messages

In line with Self-Determination Theory, the messages were written in an autonomy-supportive style. Messages were also written in a way that supports relatedness and individualization (e.g. by addressing the users personally by their names). By respecting their autonomy and making them feel related to the Active2Gether intervention, we aimed to increase the user’s willingness to follow up on the coaching messages. Moreover, the messages were written in a positive gain-framed style, that is, a style that describes the potential gains (e.g. in health, fitness, and relaxation) when participating in physical activity rather than focusing on loss (i.e. ill health, lack of fitness, and stress) when not engaging in physical activity. The majority of the messages were tailored to determinants in the theoretical framework, the weather, and occupational status.

A pilot test of a subset of messages among 7 female bachelor’s and master’s students indicated that the messages were friendly, motivational, and empathic; some were perceived as autocratic, whereas some were not. Some minor changes were made to the messages.

Step 4: Developing Tailoring Assessments

Further decisions were made on how to measure the characteristics for tailoring messages.

Assessment of Physical Activity

Our test of the validity of the Fitbit One indicated that Fitbit can be considered a valid device to assess step activity for real-time minute-by-minute self-monitoring, although an overestimation of 677 steps per day by Fitbit was seen compared with the ActiGraph. However, the validation study indicated that Fitbit is less suitable for providing instant real-time feedback and daily feedback on physical activity intensity levels (i.e. minutes of moderate, vigorous, or moderate–vigorous) because it substantially and systematically overestimates the time spent per intensity level per hour. For that reason, Fitbit is only used to assess step activity.

Participants need to give permission once for the application to access their activity data. These then can be collected regularly, and a summarized version of the data is stored in the Active2Gether database. These data are utilized in the following several ways: for presenting the activity level (i.e. number of steps and number of stairs climbed) to the user, for determining the type of coaching, and for tailoring coaching messages.

Assessment of Behavioral Determinants

We decided to assess behavioral determinants by means of a questionnaire with both its long and short versions, which were selected based on the validations of such questionnaires. The long version is based on existing questionnaires that have previously been validated (i.e. Neighborhood Quality of
Chapter 4

Life Survey and Self-efficacy scales) or questions used in previous studies and were translated and adapted where necessary\textsuperscript{147,148,172}. In the short questionnaire, we decided to use single item questions to assess each of the behavioral determinants that are part of the framework and the system. In the short version of the questionnaire, all determinants are specified for each coaching domain (i.e. sports participation, stairs use, and active transport). These items were not pretested as such but were based on the long questionnaire. Appendix 4.3 provides an overview of the questions asked in the long and short versions of the questionnaire, including the answer options.

Assessment of Location Data

We also included questions about the participants’ significant places (e.g. home address, parental home, sports location, university, work location) in the intake questionnaire. These questions focus on travel options from their home to significant locations, thus information about the active and nonactive transportation options. Additionally, information about the number of stairs available at each location and the maximum number of stairs the participant is willing to climb in one go is assessed as well.

The user’s location (GPS coordinates) is collected using Google’s location services that can be linked with the Active2Gether app. The location data are used to determine whether the user visited his or her significant locations (e.g. home, study or work place, and sports club) and to derive information about transport and travels that have been made. In addition, information about the characteristics of locations is used for personalized coaching messages to the user. For instance, if a person is being coached on using the stairs more often at their place of work or study, it is only useful to suggest this when the option to climb the stairs is indeed present at the worksite or university.

Assessment of Connected Friends

Information regarding the participants’ friends is collected using the Facebook API. Users are asked to provide access to their Facebook ID and their connections by logging into Facebook once and giving permission for this. It is important to note that Facebook does not provide personal information about someone’s Facebook connections but only a list of Facebook IDs of their connections. This information can be used to see whether any Active2Gether users are connected on Facebook. If two participants of the current intervention are connected on Facebook, they see a ranking within the app that shows achievements of both users. In this way, the users only share their achievements with a closed group and not with “everybody,” according to the preferences stated in the focus group discussions.
Step 5: Developing the Active2Gether Intervention

Designing and Developing the Communication Channel

The Active2Gether app shows a nonpersonalized, generic avatar with a welcome message that mentions the user’s current weekly goal. The app displays the current number of daily steps and stairs climbed. In addition, the app shows the following 4 graphs: a bar chart with the step progress toward 70,000 steps per week, a ranking with 6 other Active2Gether users—where possible Facebook friends—based on the step activity over the last seven days, the activity data for each weekday for the current coaching domain (i.e. minutes of sport activity, numbers of stairs climbed, or minutes of active transport), and the step activity for each weekday. The third and fourth graphs display the user’s own data and the average data assessed within Active2Gether. Moreover, these graphs can be adjusted according to the user’s preferences, that is, they can show data for the last week, last month, or from the first use.

Tailored messages and short questions are sent via push messages through the app. After the user reads the messages, they are displayed at the bottom of the app. Only the 5 messages sent most recently are displayed in the app. Appendix 4.4 shows a screenshot of the app.

Step 6: Pilot Testing

The app was adjusted based on the feedback of the 7 participants who pilot-tested the app; for example, the timing of the different steps in the tailoring process (i.e. determining the type of feedback, the coaching domain, the weekly goal, and the most promising behavioral determinants) did not originally account for exceptional cases in which a user takes very long to complete a step, causing the next step to be skipped. In the adjusted version, multiple checks and safety mechanisms were implemented to make sure that the tailoring process could still be finished correctly in such conditions. Also, automated messages to remind users to charge their Fitbit and to synchronize their data were added to the system because of the observation that participants in the pilot study sometimes did not notice when it was necessary to do so.

Step 7: Testing and Evaluating the Intervention

After developing the intervention, an evaluation study was conducted for which data have been collected between March 2016 and September 2016 and data cleaning and initial analyses are now being conducted. A three-arm quasi-experimental trial—with two active control groups—with a baseline and two follow-up assessments at 6 and 12 weeks was conducted to examine the effectiveness of the Active2Gether intervention. This trial is registered in the Dutch trial registry, No. NTR5630. A detailed description of the study protocol can be found in Appendix 4.5.
DISCUSSION

This study describes the development and content of Active2Gether, an app-based intervention, which was developed using a systematic and stepwise approach. The aim of the Active2Gether intervention is to empower young adults to become and remain physically active by providing them with app-based tailored coaching and feedback. Active2Gether makes use of an activity tracker and personalized, context-specific feedback. It focuses on 3 physical activity domains, builds on established behavior theory, and applies evidence-based BCTs and a model-based reasoning system to provide individually tailored coaching messages based on current scores on the behavioral determinants.

The development and content creation of Active2Gether was a stepwise process. The program-planning model proposed by Kreuter et al. was mainly used to guide the development and content of the Active2Gether intervention. It states that the health problem needs to be analyzed before developing an intervention, the intervention needs to be based on theory and scientific evidence, and the developmental process is a loop of development, evaluation, and adjustment of the intervention. Program-planning models provide detailed guidance to develop an intervention, which also takes time. Because the possibilities of modern technology in interventions are rapidly evolving, possibilities and preferences that were assessed at the beginning of a lengthy development process may be outdated at the time of implementation or evaluation. The development and content of Active2Gether were guided by relevant health behavior theories and scientific evidence, aiming to develop an intervention that provides a highly tailored feedback. Consequently, less attention was paid to app design and aesthetics that might have resulted in a less appealing app compared with commercial apps. Furthermore, the app is only available for Android devices running on version 4.0 or higher and is therefore not available for older Android devices and smartphones running on other operating systems. Active2Gether incorporates a number of conditions to secure high levels of engagement. First, our approach, integrating a model-based reasoning system, allows us to provide the user with a dynamically tailored intervention that adjusts to the changes in the user. Second, by applying multiple levels of tailoring in the app and the content of the messages (i.e. type of support, coaching domain, coaching messages, and weekly goals), the app is likely to be regarded as personally relevant and increase feelings of relatedness. Third, by comparing the physical activity of the user with that of other Active2Gether users (if possible with their Facebook friends), we expect to further increase personal relevance and relatedness. Lastly, by giving the user the option to select from 3 physical activity domains and set their own goals with guidance and suggestions based on their own input, we expect higher levels of autonomy, resulting in higher motivation to follow up on the coaching messages. However, to implement these different levels of tailoring, detailed user information is needed repeatedly; thus, frequent user input is needed, which increases user burden.
To date, mobile phones and personal digital assistants have been used to monitor physical activity with either smartphone apps or external devices, deliver feedback, provide information, and offer a support system to the participants. Active2Gether makes use of an external device, Fitbit One, to monitor physical activity and provide feedback through the app based on the user’s behavior. However, Active2Gether goes beyond existing interventions by combining data from multiple sources to send context-specific messages. Furthermore, the majority of the published interventions focuses on step activity, whereas Active2Gether focuses on sports activity, stair walking, and active transport as well. Therefore, the app may be more appealing to participants who do not like to participate in sports, especially because the user can adapt to his or her coaching domain every week. However, Active2Gether does not yet incorporate geofencing (i.e. sending location-triggered messages), which would further improve the possibilities for context specificity and real-time feedback and advice by, for example, sending a reminder to climb the stairs at work when users are close to their work location.

So far, the majority of the app-based interventions to promote physical activity showed positive short-term effects. In line with other app-based physical activity interventions, Active2Gether makes use of self-monitoring, goal setting, and providing feedback. However, Active2Gether provides dynamically tailored feedback using artificial intelligence-based techniques and including conditional factors (i.e. weather), whereas other interventions use logic statements and decision rules to specify which messages should be sent to the user; for example, Active2Gether uniquely assesses behavioral determinants every week to provide tailored advice and feedback on the current behavior, whereas most studies mostly provide feedback on the current behavior only. Current app-based interventions to promote physical activity focus on step activity or overall moderate–vigorous, whereas Active2Gether focuses on sports activities, active transport, and stair walking as well. The majority of papers on app-based interventions reported significant effects, and a study that combined machine learning techniques to send personalized messages that were contextualized to the user’s environment and previous behavior showed promising results with regard to the efficacy of the intervention. Because the Active2Gether intervention went beyond the majority of those apps and included BCTs proven to be effective, we expected to find significant intervention effects compared with the 2 (active) control groups.

Active2Gether is ambitious and innovative and incorporates certain risks, for example, the intervention highly relies on input from the activity monitor and location sensor and thus on the user to turn on and synchronize the tracker with the server. Furthermore, it relies on responses from the users on repeated questionnaires. If they do not provide input at all or if they do not provide true and honest answers, the coaching messages that are informed by this information may become irrelevant and nontailored. Moreover, if a participant is not a Facebook user or has no appropriate contacts, the personalization
could be limited. Finally, if technical problems are encountered, this may result in errors in synchronization and sending messages late or not at all. To limit the burden for the participants and minimize their input to reduce potential technical problems, future research could make use of smartphone sensors to assess the participant’s behavior.

The overall effectiveness of Active2Gether thus needs to be, and is being, evaluated in a quasi-experimental trial with a 12-week follow-up. However, because app-based interventions offer the possibility to deliver just-in-time interventions that are relevant for the user’s situation for that particular moment, a study is needed to examine the possible effectiveness of specific real-time feedback and advice moments. An ecological momentary assessment in such a quasi-experimental trial setting may help to assess potential specific effects throughout the intervention period. An evaluation of the efficacy of the intervention and the usability can help to further adapt and improve the intervention for future research. Furthermore, data collected during the trial can provide insights on how to further personalize content to the users. The quasi-experimental trial also includes monitoring of app use and a process evaluation of app use and appreciation that will provide information on larger scale dissemination, implementation, and changes required to improve conditions for wider use of the app.

Because the intervention has been developed with an early consideration for the preferences of the target population, it is more likely to meet the expectations of the target population. Consequently, the intervention is more likely to be adopted by the target population. However, the intervention might be prone to technical errors, and a significant input from the user is needed to provide tailored feedback. This might be a burden for the participants, leading to a lower adoption rate. We conducted a small pilot study to test the Active2Gether app and to detect bugs and technical errors; ideally, the pilot study would have been conducted with a larger sample. The current version of the Active2Gether intervention has been developed for young adults with higher education owning a smartphone running on Android version 4.0 or higher. The content needs to be adjusted before offering the intervention to other target populations.

This paper describes the systematic development of an intervention that is based on theory and input from end users. The use of a model-based reasoning system to provide context-specific coaching messages goes beyond many existing eHealth and mHealth interventions.
## APPENDIXES

### Appendix 4.1 – Overview of the behavior change techniques (BCTs) that were selected to target the behavioral determinants of the theoretical framework and how they were applied within the intervention

<table>
<thead>
<tr>
<th>Determinant</th>
<th>BCTs</th>
<th>Methods</th>
<th>Parameters for use</th>
<th>How applied</th>
<th>Theory BCT</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome expectations</td>
<td>Provide general information on consequences of behavior in general</td>
<td>Belief selection</td>
<td>Requires investigation of the current attitudinal, normative and efficacy beliefs of the individual before choosing the beliefs on which to intervene.</td>
<td>Messages in general and tailored to aspects of the intake questionnaire</td>
<td>Information-motivation-behavioral skills model</td>
<td>55, 181</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>Action planning/ time management</td>
<td>Guided practice</td>
<td>Subskill demonstration, instruction and enactment with individual feedback</td>
<td>Messages prompting the planning of physical activity, e.g. the suggestion to mark time and day in the calendar</td>
<td>Goal setting theory</td>
<td>108, 109, 137, 181</td>
</tr>
<tr>
<td>Social comparison</td>
<td>Modeling</td>
<td></td>
<td>Attention, remembrance, skills, reinforcement; credible source, method and channel</td>
<td>Graph tailored to preference social comparison (up-/downward)</td>
<td>Social comparison theory</td>
<td>109, 181</td>
</tr>
<tr>
<td>Persuasion / Problem solving</td>
<td>Verbal persuasion</td>
<td></td>
<td>Credible source.</td>
<td>Persuasive messages on how to overcome barriers</td>
<td>Social cognitive theory</td>
<td>109, 181</td>
</tr>
<tr>
<td>Prompt self-monitoring</td>
<td>Self-monitoring of behavior</td>
<td></td>
<td>The monitoring must be of the specific behavior (that is, not of a physiological state or health outcome). The data must be interpreted and used. The reward must be reinforcing to the individual.</td>
<td>Messages that prompt to look at the monitoring graphs and display of graph</td>
<td>Self-regulation, Social cognitive theory, Control theory</td>
<td>108, 181, 182</td>
</tr>
<tr>
<td>Plan social support</td>
<td>Mobilizing social networks</td>
<td></td>
<td>Presence of a network that can potentially support health behavior</td>
<td>Messages with suggestions to tell friends and ask for support</td>
<td>Social support theories</td>
<td>108, 109, 181</td>
</tr>
<tr>
<td>Imaginary reward</td>
<td>Provide contingent rewards/reinforcement</td>
<td></td>
<td>The reward needs to be tailored to the individual, to follow the behavior in time, and to be seen as a consequence of the behavior.</td>
<td>Messages that tell the user to be proud if they did well</td>
<td>Self-regulation, Social cognitive theory, Learning theories</td>
<td>181, 182</td>
</tr>
</tbody>
</table>

*table continues*
<table>
<thead>
<tr>
<th>Intenions</th>
<th>Motivational messages/verbal persuasion about capacity</th>
<th>Feedback</th>
<th>Feedback needs to be individual, follow the behavior in time, and be specific.</th>
<th>Messages that tell the user how much he/she has already achieved and display of graph</th>
<th>Self-regulation, Social Cognitive Theory</th>
<th>108, 109, 181</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivational messages/verbal persuasion about capacity</td>
<td>Modeling</td>
<td>Modeling</td>
<td>Attention, remembrance, skills, reinforcement; credible source, method and channel</td>
<td>Messages stating how well others are doing and display of graph</td>
<td>Social cognitive theory, Theory of planned behavior</td>
<td>55, 181</td>
</tr>
<tr>
<td>Provide instruction</td>
<td>Model setting</td>
<td>Feedback</td>
<td>Feedback needs to be individual, follow the behavior in time, and be specific.</td>
<td>Messages telling the user how well he/she is doing and to keep up the good work or telling the user some advantage of being physically active</td>
<td>Social cognitive theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Progress towards goal/Discrepancy between current behavior and goal</td>
<td>Goal setting</td>
<td>Goal setting</td>
<td>Commitment to the goal; goals that are difficult but available within the individual’s skill level</td>
<td>Messages prompting the user to set a goal and providing a suggestion</td>
<td>Self-regulation, Social cognitive theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Impediments</td>
<td>Prompt barrier identification / Problem solving</td>
<td>Planning coping responses</td>
<td>Identification of high-risk situations and practice of coping response.</td>
<td>Message that provides information on how to deal with a specific barrier</td>
<td>Social cognitive theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Social norm (descriptive and inductive)</td>
<td>Social comparisons</td>
<td>Graph tailored to preference social comparisons (up-/downward</td>
<td>Upward comparison may help setting better goals; downward comparison may help feeling more self-efficacious.</td>
<td>Messages that provide information on how to deal with a specific barrier</td>
<td>Social comparison theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Information about others’ approval</td>
<td>Information about others’ approval</td>
<td>Information about others’ approval</td>
<td>Positive expectations available in social environment</td>
<td>Messages that provide information on how to deal with a specific barrier</td>
<td>Social cognitive theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Self-regulation</td>
<td>Self-monitoring</td>
<td>Feedback</td>
<td>Feedback needs to be individual, follow the behavior in time, and be specific.</td>
<td>Messages that prompt to look at the monitoring graphs and display of graph</td>
<td>Self-regulation, Social cognitive theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Goal setting</td>
<td>Goal setting</td>
<td>Goal setting</td>
<td>Commitment to the goal; goals that are difficult but available within the individual’s skill level</td>
<td>Messages that prompt the user to set a weekly goal</td>
<td>Control theory</td>
<td>55, 181</td>
</tr>
<tr>
<td>Intentions</td>
<td>Progress towards goal/Discrepancy between current behavior and goal</td>
<td>Feedback</td>
<td>Messages that tell the user how much he/she has already achieved and display of graph</td>
<td>Control theory 55, 181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------</td>
<td>---------</td>
<td>-----------------------------------------------------------------</td>
<td>----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Cognitive Theory</td>
<td>Motivational Feedback</td>
<td>Self-regulation, Discrepancy between current behavior and goal</td>
<td>Messages prompting the user to evaluate how he/she is feeling about failing or achieving the self-set goal</td>
<td>Control theory, Integrated theory of health behavior change 55, 181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td>Goal setting</td>
<td>Self-evaluation, Monitoring outcome(s) of behavior</td>
<td>Messages that tell the user to evaluate how he/she is feeling about failing or achieving the self-set goal</td>
<td>Control theory, Integrated theory of health behavior change 55, 181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imaginary reward</td>
<td>Imaginary reward</td>
<td>Imaginary reward</td>
<td>Messages that tell the user to be proud if they did well</td>
<td>Social cognitive theory, Self-determination theory 55, 62, 181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>Satisfaction</td>
<td>Satisfaction</td>
<td>Messages prompting the user to evaluate how he/she is feeling about failing or achieving the self-set goal</td>
<td>Control theory, Integrated theory of health behavior change 55, 181</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-term goals</td>
<td>Long-term goals</td>
<td>Consciousness raising</td>
<td>Messages providing general information on consequences of behavior in general</td>
<td>Information-motivation-behavioral skills model 55, 181</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4.2 - Components and flow chart of the tailored intervention

Intake and Assessment week

Determining type of support
Type of support is based on the user’s actual activity level and the perceived activity levels

Education
Feedback
Coaching

Selecting coaching domain
Based on the user’s activity level and opportunities for increasing physical activity in the three domains based on the context from the intake. The final decision is up to the user.

Active Transport
Stair walking
Sports participation

Suggesting weekly goal
Individual domain specific goals tailored to actual activity levels of the user. The final decision is up to the user.

Selecting behavioral determinants
Advanced artificial intelligence based techniques were used. The system makes use of a computational model of behavior change, based on the theoretical framework, that is used to predict behaviors (e.g., sports, stair use, active transport) for each participant. The relevant behavioral determinants are assessed with the short questionnaire on a weekly basis based on the single items described in Multimedia Appendix 4.3.

Compiling and sending tailored messages
Users receive feedback and advice messages from the message library that are tailored to their personal activity and behavioral personal determinants and contextual factors. During the day, the system checks three times whether a relevant message can be sent to the user.
## Appendix 4.3 - Overview of the questions used for the short and long version of the questionnaire

<table>
<thead>
<tr>
<th>Behavior determinant</th>
<th>Example question long version</th>
<th>Number of items long version</th>
<th>Question short version</th>
<th>Answer possibilities</th>
</tr>
</thead>
</table>
| **Outcome Expectations** [NQLS a, 183] | If I participate in regular physical activity or sports, then:  
- I will improve my health  
- I will feel more attractive  
- I will lose weight  
- I will improve my physical fitness  
- I will feel relaxation  
- I will feel less tension and stress | 6 | If I participate in regular sports, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel.  
If I regularly take the stairs, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel.  
If I regularly bike or walk, then it will have a positive effect, for example on my healthy, appearance, weight or how I will feel. | 1 – No reason at all  
2 – A slightly important reason  
3 – A quite important reason  
4 – A very important reason |
| **Self-efficacy** [Self-efficacy scales for exercise 117, NQLS a 183] | How confident are you that you could do PA, in each of the following situations? I’m confident that I could:  
- Do PA even when I’m tired  
- Do PA even when I’m in a bad mood  
- Do PA even when I feel I don’t have time  
- Do PA even when I am on holiday  
- Do PA even when it is raining | 12 | How confident are you that you will do sports in the next week even when you’re tired, busy or when it’s bad weather?  
How confident are you that you will take the stairs in the next week even when you’re tired, you’re in a hurry or you’re with others?  
How confident are you that you will cycle or walk to work/the university in the next week, even when you’re tired, you’re busy or when it’s bad weather? | 1 – Not at all confident  
2 – Slightly confident  
3 – Moderate Confident  
4 – Very Confident  
5 – Extremely confident |
| **Perceived barriers for sport** [NQLS a 183] | How often do the following barriers prevent you from doing sports activities?  
- Bad weather  
- Lack of time | 12 | How often do barriers prevent you from participating in sports or exercise activities?  
Think for example of lack of time, lack of | 1 – Never  
2 – Rarely  
3 – Sometimes  
4 – Often  
5 – Very often |

*table continues*
Lack of interest in exercise
– Other priorities
– Lack of skills or knowledge
– Lack of equipment
– Lack of facilities or space
– Lack of physical fitness
– Lack of energy
– Lack of money
– Lack of company
– Self-conscious about my looks when I exercise.

energy, costs, lack of company.

Perceived barriers for active transport

How often do the following barriers prevent you from traveling by bike or by walking instead of traveling by car or public transport?
– Bad weather
– Lack of time
– Lack of physical fitness
– Lack of energy
– Too many pieces of luggage
– Travel distance is too far away
– No suitable bike

How often do barriers prevent you from cycling or walking to work/the university instead of traveling by public transport or car? Think for example of lack of time, lack of physical fitness, lack of energy or too many pieces of luggage.

1 – Never
2 – Rarely
3 – Sometimes
4 – Often
5 – Very often

Perceived barriers for stairs climbing

How often do the following barriers from climbing the stairs?
– Lack of physical fitness
– Lack of energy
– Too many pieces of luggage
– Too many flights of stairs

How often do barriers prevent you from climbing the stairs? For example barriers as being in a hurry, lack of physical fitness, lack of energy or carrying too many pieces of luggage?

1 – Never
2 – Rarely
3 – Sometimes
4 – Often
5 – Very often

Social norm descriptive

– My friends think that I should be sufficient physically active.
– My fellow students think that I should be sufficient physically active.
– My brother(s) and/or sister(s) think that I should be sufficient physically active.

For the people around me (friends, fellow students, family), it’s important that I sufficiently participate in sports.

For the people around me (friends, fellow students, family), it’s important that I use the stairs instead of the elevator.

1 – I strongly disagree
2 – I somewhat disagree
3 – Neutral
4 – I somewhat agree
5 – I strongly agree

Satisfaction

How satisfied are you about how physically active you are?

How satisfied are you about how often you did sports in the last week?

How satisfied are you about how often you...
students, family), it’s important that I regularly bike or walk to work/the university.

<table>
<thead>
<tr>
<th>Social norm injunctive</th>
<th>– How often do your friends/roommates/brothers or sisters/parents participate in physical activities</th>
<th>4</th>
<th>Not applicable (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention</td>
<td>Do you intend do sports (more often) within the next week/month/6 months even if you think you’re already sufficiently active?</td>
<td>3</td>
<td>1 – Never (^c)</td>
</tr>
<tr>
<td></td>
<td>I intend to do sports (more often) within the next week.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I intend to climb the stairs (more often) within the next week.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>I intend to bike or walk to work/the university (more often) within the next week.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulation</td>
<td>I keep track of how active I am. I check whether I met my goals. In the last three months I:</td>
<td>7</td>
<td>1 – Most definitely will not</td>
</tr>
<tr>
<td></td>
<td>– Set aside time for my daily physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>– Walked or biked instead of drove or travelled by public transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>– I exercised or did physical activities with someone else</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>– I wrote it down in my calendar to do sports/physical activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>– I planned do sports/exercise even when the weather was bad</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>How satisfied are you about how physically active you are?</td>
<td>1</td>
<td>0 – Very unhappy</td>
</tr>
<tr>
<td></td>
<td>How satisfied are you about how often you did sports in the last week?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>How satisfied are you about how often you</td>
<td>10 – Very happy</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Each item is rated on a five point scale: 1 – Never, 2 – Rarely, 3 – Sometimes, 4 – Often, 5 – Very often.

\(^b\) Not applicable.

\(^c\) Scale ranges from 1 (Never) to 5 (Very often).
climbed the stairs last week?

How satisfied are you about how often you biked or walked to work/the university last week?

<table>
<thead>
<tr>
<th>Long-term goals</th>
<th>Not applicable</th>
<th>0</th>
<th>How motivated are you to be more physically active?</th>
<th>0 – Very unmotivated</th>
<th>10 – Very motivated</th>
</tr>
</thead>
</table>

\(^a\) NQLS= Neighborhood Quality of Life Study  
\(^b\) Questions assessing social norm injunctive were not included in the short questionnaire.  
\(^c\) Long-term goals were not included in the long questionnaire

Appendix 4.3 - Screenshot of the Active2Gether app

Note. This picture shows a screenshot of the Active2Gether app with a chart indicating the proximity to the 70,000 steps per week goal, a graph with an overview of the weekly steps, the ranking of steps compared to others, and an example of a message.
Appendix 4.4 - Study protocol

STUDY DESIGN

A three-arm quasi-experimental trial with a baseline and two follow-up assessments at 6 and 12 weeks is being conducted to examine the effectiveness of the Active2Gether intervention. This trial is registered in the Dutch trial registry, No. NTR5630.

The three study arms are two versions of Active2Gether and an active control group. The first version of Active2Gether is the full version with personalized coaching messages for the three coaching domains, weekly goals, self-monitoring and social comparison, as described in Step 1-3. The second version of Active2Gether is similar, but does not include the coaching part, i.e., selecting a coaching domain, setting a weekly goal, receiving personalized messages to increase levels of PA. The second version of Active2Gether does still allow the users to monitor their step activity and to compare themselves with other participants participating in the trial. The control condition will receive the Fitbit One activity tracker and the Fitbit app, i.e. an existing app-based intervention with much lower levels of tailored feedback and advice. The comparison between the two Active2Gether versions will provide information regarding the effects, use and appreciation of the coaching part of Active2Gether, the comparison with the Fitbit control arm will provide information about the effects, use and appreciation as compared to an existing ‘usual care’ PA monitoring and promotion device.

HYPOTHESES

The null hypothesis is that there is no difference in mean daily minutes of moderate-vigorous physical activity (MVPA) between the three intervention groups, i.e. Active2Gether Full, Active2Gether Light and Fitbit, at 12-week follow up. The alternative hypothesis is that the Active2Gether Full condition will show significantly higher levels of mean daily minutes of MVPA in comparison to the Active2Gether Light and Fitbit condition at 12-week follow up.

RECRUITMENT AND PARTICIPANTS

Participants (N = 160-200) that are representative for the target population, i.e. young adults between 18 and 30 years, are recruited by flyers, posters, social media and personal contacts, and snowball strategies in the Amsterdam and Utrecht regions.

Participants are then asked to sign up for the trial through the Active2Gether website. For stratification purposes during the randomization process, the participants are asked to provide information about sex, age and type of smartphone they use (i.e., iOS or Android).

Stratified randomization is applied based on the type of smartphone and gender. As the A2G app only runs on Android, iPhone owners will automatically be assigned to the Fitbit condition. Android users
will be assigned to the Fitbit condition when possible. The aim is to divide men and women with an Android phone equally over the two A2G conditions. This was done by using a 1:1 ratio applied to the order of registering.

Participants are eligible for this study when they meet the following criteria: (a) aged 18-30 years; (b) being (apparently) healthy; (c) Dutch speaking and (d) signed the informed consent form, (e) have a suitable smartphone running on iOS or Android. The study has been approved by the Medical Ethical Committee of the VU Medical Center (2015.363).

**MEASURES**

**Physical Activity**

PA is assessed with two types of tri-axial accelerometers. The ActiGraph wGT3x-BT is used to evaluate the efficacy – thus mean daily minutes of MVPA – of the intervention and will be assessed at baseline and 12 weeks follow-up for one week. Furthermore, all participants will receive a Fitbit One that is used to continuously monitor their PA behavior in minute-by-minute intervals and allows the participants to monitor their PA behavior for 12 weeks.

**Behavioral determinants**

To evaluate changes in behavior determinants, participants are asked to fill in a questionnaire assessing all behavioral determinants that are in the theoretical framework at baseline, 6 and 12 weeks follow-up. The questionnaire is based on existing questionnaires that have previously been validated.

**App usage and appreciation**

Information regarding the use and appreciation of the app - the two versions of the Active2Gether app and the Fitbit app - is collected at 12 weeks follow-up. During the trial, additional qualitative research is conducted to explore the users’ experiences to improve the Active2Gether intervention. A subgroup of participants that use the Active2Gether app including the coaching part are asked for an interview capturing their experiences with the app. Lastly, for both Active2Gether apps information about the number of app-log-ins are collected as well.

**App use and dropouts**

Engagement with the intervention is assessed using the number of coaching messages read – only for the A2G-Full condition - and Fitbit usage (for all participants). As all participants are asked to wear the Fitbit during the intervention, the number of valid days the Fitbit was worn during the 12 weeks – thus 84 days – can be used as an indicator for attrition.
The development of Active2Gether

User experience

User appreciation is assessed with 20 items using a 7-point Likert Scale. Examples of statements in this assessment tool are: (1) satisfaction: “the app meets my expectations”, (2) user friendliness: “I can easily find the information I’m looking for”, (3) perceived effectiveness: “the app motivates me to achieve my goals”, and (4) professionality: “the app looks professional”.

Evaluation of questions and messages (A2G)

Evaluation of the number of questions (Active2Gether-Full and Light) and messages (Active2Gether-Full) sent to the respondent by the app is assessed on 5-point Likert scales asking, ‘the number of questions/messages sent by the app are...’. Additionally, participants in the Active2Gether Full condition receive eight statements on: a) tone (i.e., how friendly is the message written), b) authority (i.e., the message is an obligation to do something), c) personal relevancy, d) trustworthiness, e) motivational strength, f) empathy.

Positive and negative aspects

Participants could list up to three positive and three negative features of the app in a free text.

ANALYSES

The trial data will be analyzed for differences in effects of the two versions of Active2Gether as compared to the Fitbit app. The primary outcomes will be total MVPA. Secondary outcomes will be changes in determinants of PA, i.e. self-efficacy, outcome expectations, intention, self-regulation, social norm, long-term goals, and satisfaction.

The efficacy of the intervention will be assessed in multiple steps. First, linear regression analyses will be used to compare MVPA at follow-up, adjusted for baseline values, between the different conditions. Second, secondary outcome measures will also be compared across the three conditions. Thus the effectiveness of the intervention to change weekly minutes spent in sports activities, number of stairs climbed, and minutes of active transport will be evaluated as well. Third, the hypothesized underlying processes will be evaluated using mediation analysis. With mediation analysis we will examine whether the intervention successfully changes the addressed behavioral determinants and whether that in turn resulted in changes in MVPA.

Usability and levels of engagement will be evaluated, by analyzing the duration and frequency of involvement with the Active2Gether app, and indicators of non-usage attrition.
A Validation Study of the Fitbit One in Daily Life Using Different Time Intervals

Anouk Middelweerd
Hidde P Van Der Ploeg
Aart Van Halteren
Jos WR Twisk
Johannes Brug
Saskia J Te Velde

ABSTRACT

Purpose
Accelerometer-based wearables can provide the user with real-time feedback through the device’s interface and the mobile platforms. Few studies have focused on the minute-by-minute validity of wearables, which is essential for high-quality real-time feedback. This study aims to assess the validity of the Fitbit One compared with the ActiGraph GT3x+ for assessing physical activity (i.e. steps, time spent in moderate, vigorous, and moderate–vigorous physical activity) in young adults using traditional time intervals (i.e. days) and smaller time intervals (i.e. minutes and hours). 

Methods
Healthy young adults (N = 34) wore the ActiGraph GT3x+ and a Fitbit One for 1 wk. Three aggregation levels were used: minute, hour, and day. Mixed models analyses, intraclass correlation coefficients, Bland–Altman analyses, and absolute error percentage for steps per day were conducted to analyse the validity for steps and minutes spent in moderate, vigorous, and moderate–vigorous physical activity.

Results
As compared with ActiGraph (mean = 9 steps per minute, 509 steps per hour and 7636 steps per day), the Fitbit One systematically overestimated physical activity for all aggregation levels: on average 0.82 steps per minute, 45 steps per hour, and 677 steps per day. Strong and significant associations were found between ActiGraph and Fitbit results for steps taken (B = 0.72–0.89). Weaker but statistically significant associations were found for minutes spent in moderate, vigorous, and moderate–vigorous physical activity for all time intervals (B = 0.39–0.57).

Conclusions
Although the Fitbit One overestimates the step activity compared with the ActiGraph, it can be considered a valid device to assess step activity, including for real-time minute-by-minute self-monitoring. However, agreement and correlation between ActiGraph and Fitbit One regarding time spent in moderate, vigorous, and moderate–vigorous physical activity were lower.
INTRODUCTION

Commercial wearable technologies that continuously monitor physical activity may be helpful tools for monitoring and self-monitoring and for providing feedback on physical activity behaviors. For example, the Fitbit One is an accelerometer-based wearable activity tracking device that is easy to use and has a user-friendly interface. The activity monitor wirelessly uploads the activity data to the user’s account, which is accessed through a smartphone application or website \(^{184,185}\). It assesses daily activities, allows the user to monitor his/her progress, and provides real-time feedback (e.g. current amount of steps taken, stairs taken and a growing or shrinking flower depicts the activity level) to encourage the user to be more physically active \(^{184}\). In addition, the Fitbit One and its web interface include behavior change techniques typically used in physical activity interventions \(^{40}\). The Fitbit One—like other similar wearables—may offer new opportunities in interventions that target physical activity promotion.

Several studies examined the validity of different types of Fitbit activity trackers \(^{186-189}\); however, as far as we are aware, only eight studies examined the validity of the Fitbit One. So far, five studies used a controlled laboratory setting to examine the validity of the Fitbit One with respect to steps taken \((n = 4)\) and energy expenditure \((n = 1)\) \(^{188,190-192}\). Overall, in laboratory settings, studies reported acceptable percentages of relative errors \((1.3–10.5\%\) for number of steps taken with speeds of \(0.7–1.78\ m \cdot s^{-1}\) on a treadmill \(^{188,189,192}\). Moreover, the Fitbit underestimated the energy expenditure compared with indirect calorimetry. So far, only three studies examined the validity of the Fitbit One in daily life \(^{187,193,194}\). High Pearson’s \(r\) and intraclass correlations were reported for steps taken per day compared with the ActiGraph \(^{193}\). However, the Fitbit One was less valid in assessing daily minutes spent in moderate–vigorous physical activity \(^{187,193}\). These findings suggest that the Fitbit One is a valid instrument to assess daily step activity but less valid to assess daily minutes spent in moderate–vigorous physical activity.

Moreover, the Fitbit One can be used as a stand-alone intervention or as an intervention tool to monitor daily life physical activity and to provide real-time feedback. It can be used with its own applications or in combination with a third-party smartphone application. The Fitbit website provides access to minute-by-minute activity data through the application programming interface (API). A recent study in overweight, postmenopausal women showed promising results in favor of the intervention group that received the Fitbit One as an intervention tool \(^{195}\). Because real-time feedback is and should be based on data collected throughout the day and thus on data collected over smaller time intervals than the whole day, it is important to know the validity of the Fitbit for levels of physical activity based on smaller time intervals than the whole day (i.e. total daily activity). This is important because even if studies have indicated that the Fitbit One is valid in...
assessing daily activity, this could be because measurement errors—over- and underestimations—can balance out throughout the day which would result in good daily validity but not necessarily good minute or hourly validity. Therefore, it is important to know the validity of the Fitbit using smaller time intervals (i.e. per minute and hour) and, thus, if the Fitbit One can be used for real-time feedback. This focus on the smaller time intervals is important as a poor hourly or minute-by-minute validity might be off-putting to users that experience the wearable is not matching their activities in real time, and consequently, it might be less suitable for physical activity interventions.

One previous study looked at 3-h periods and demonstrated a good concurrent validity of the Fitbit Ultra—the forerunner of the Fitbit One—to assess physical activity in daily life in patients with chronic obstructive pulmonary disease\textsuperscript{196}. Another study used smaller time intervals, i.e. minutes, hours, and days, to assess the interdevice reliability, but the study only had one participant\textsuperscript{186}. Therefore, to date, little is known about the validity of the Fitbit One in detecting physical activity using smaller time intervals relevant for real-time feedback and instant behavioral insights to its users. Hence, the primary aim of the current study is to assess the construct validity of the Fitbit One for steps and minutes spent in moderate physical activity, vigorous physical activity, and moderate–vigorous physical activity for minute and hourly intervals against the ActiGraph. The secondary aim was to assess the more traditional construct validity of the Fitbit One for steps, minutes spent in moderate physical activity, vigorous physical activity, and moderate–vigorous physical activity per day against the ActiGraph.

**METHODS**

**Participants**

A sample of 34 participants (23 females and 11 males) agreed to participate in the current study. Participants were actively recruited in Amsterdam by flyers distributed on one of the main university campuses, and by e-mail, and direct person-to-person communication at this university. Participants were eligible when they met the following criteria: (a) age 19–30 years, (b) owned a smartphone, (c) fluent in Dutch or English, and (d) signed the informed consent form.

**Procedure**

Participants were asked to visit the study center to pick up the instructions and devices. After providing informed consent, the participants completed an online questionnaire collecting demographic information such as age, gender, self-reported height, and weight. The participants received log-in information for the Fitbit website and instructions on how to use the devices; for
example, how to synchronize the Fitbit, and how to wear the devices. Participants were instructed to wear both activity monitors on the right hip using an elastic belt for seven consecutive days during waking hours. In addition, they were instructed to remove the activity monitors during water activities and sleeping. After the assessment week, participants returned the activity monitors and received a 10-euro voucher as incentive for participating.

The study protocol was approved by the Medical Ethical Committee of the VU University Medical Centre Amsterdam.

Measurements

The ActiGraph GT3X+ (ActiGraph Inc., Pensacola, FL) was used as the reference method (construct validity) to compare the Fitbit with. The ActiGraph is recognized as a reasonable, valid tool to assess physical activity objectively in adults and has been used in numerous studies\(^{197, 198}\). The ActiGraph GT3X+ is a triaxial accelerometer that is able to convert accelerations to step counts. The monitors were set to collect raw data at 100 Hz. After data were aggregated into 1-min intervals, Troiano’s definitions—using the vertical counts—were used to calculate time spent in sedentary (<100 counts per minute), light (100–2019 counts per minute), moderate (2020–5998 counts per minute), vigorous (5999≤ counts per minute), and moderate to vigorous (moderate–vigorous physical activity, 2020≤ counts per minute) physical activities\(^{199}\).

The Fitbit One (Fitbit Inc., San Francisco, CA) is a light-weight triaxial accelerometer with a build in altitude monitor. The Fitbit collects data in 1-min intervals. Data were retrieved using the open API and saved as xml file. Fitbit uses an algorithm to detect motion patterns and to convert these accelerations to step counts\(^{200}\). Fitbit calculates the intensity of the activity by means of an algorithm only known within the Fitbit company. On the basis of the intensity of the activity, Fitbit then estimates the corresponding MET values to classify the intensity in sedentary, lightly active, fairly active, and very active\(^{201}\). The monitor can be worn in the front pocket, on a belt or a bra; however, for the present study, participants were asked to wear the Fitbit on the right hip with a waist belt, next to the ActiGraph.

Participants were asked to record their wear time through a daily e-mail with a link to a questionnaire recalling the previous day. This e-mail was sent at the beginning of each day—starting at the second day.

Data handling

First, ActiGraph data were checked for nonwear time using Actilife 6.0 (ActiGraph Inc.). Troiano’s definitions were used to identify nonwear time; for example, periods with consecutive strings of zero’s
for at least 60 min\textsuperscript{199}. In addition, up to two interruption intervals of <100 counts that appeared in the middle of long strings of zero-count intervals were filtered out\textsuperscript{202, 203}. Intervals with 920,000 counts per minute were considered to be spurious\textsuperscript{202, 204}.

Second, Fitbit data were matched to the periods of wear time of the ActiGraph data. All Fitbit data with zero steps per hour were compared with the participant’s reports and if necessary excluded from the analyses. Some participants reported low battery of the Fitbit, and those time intervals were deleted. For both monitors, at least 4 days of minimal 10 hour of wear time was required to for a participant to be included in the analyses.

**Data analyses**

In preparation for the analyses, three levels of aggregation were distinguished (i.e. minutes, hours, and days). For each aggregation level and each type or intensity of activity (i.e. steps, minutes spent in moderate, vigorous, and moderate–vigorous activities), the validity was examined in five steps.

First, the absolute error percentage \(\left(\frac{|\text{Fitbit} - \text{Actigraph}|}{\text{Actigraph}} \times 100\right)\) was calculated to provide an indicator of the overall difference between the Fitbit and the ActiGraph. Because the absolute error percentage cannot be calculated for ActiGraph values of zero or close to zero, the error percentage was calculated for steps per day only.

Second, because longitudinal data observations within one subject over time are correlated, systematic differences were obtained by means of linear mixed model analyses with a four-level structure (minute measurements were clustered within hours, within days, and within weeks). The dependent variable was the continuous activity data assessed by either the Fitbit or the ActiGraph, and the independent dichotomous variable was the device (ActiGraph = 0, Fitbit = 1). The obtained regression coefficient represents the systematic difference between the Fitbit and the ActiGraph adjusted for the nested design.

Third, Bland–Altman plots, including the limits of agreement (LoA), were used to provide a visual representation of the systematic differences and to assess potential non-systematic differences between the ActiGraph and the Fitbit. As the activity data in the current study comprise a mixture of between and within individual information on the differences between the two measurements, we applied the method proposed by Bland and Altman to estimate the LoA for repeated measures by adjusting for the nonindependence of the observations. Because of the fact that Fitbit and ActiGraph collect data in 1-min epochs, the variable for time spent in moderate, vigorous, or moderate–vigorous physical activity at the minute level was dichotomous, either that minute was at that intensity or not.
Therefore, Bland–Altman analyses were not conducted for minutes spent in moderate, vigorous, and moderate–vigorous physical activity for the aggregation level minute.

Fourth, intraclass correlation coefficients (ICC) for continuous data were used as parameters for criterion validity of the Fitbit compared with the ActiGraph. The ICC values (two-way random effects model with an absolute agreement definition) were calculated for all aggregation levels for step activity and time spent per intensity level. For the present study, an ICC Q0.75 was defined as “excellent,” an ICC between 0.4 and 0.75 was defined as “fair to good” and an ICC G0.4 as “poor” 202, 205.

Lastly, the association between ActiGraph data (dependent variable) and Fitbit data (independent variable) was also obtained with linear mixed model analyses with the same four-level structure as described previously. Regression coefficients were interpreted as “excellent” (B = 0.8–1.0), “substantial” (B = 0.6–0.8), “moderate” (B = 0.4–0.6), “fair” (B = 0.2–0.4), and as “poor” (B = 0.0–0.2) correlations 203.

All analyses were checked for outliers (≥3 SD of the residuals) and when necessary sensitivity analyses was conducted without outliers. Results of the sensitivity analyses can be found in Appendix 4.5. Statistical analyses were performed in R Studio (R Studio, Boston) using lme, lmer, glm, and ggplot2 packages.

RESULTS

Descriptive characteristics. Aggregating the activity data into the three aggregation levels resulted in data sets consisting of 138,224 observations for data aggregated into minute intervals, 2518 observations for data aggregated into hour intervals, and 168 observations for data aggregated into day intervals.

Table 4.2 shows descriptive characteristics of the study sample (N = 34); the majority was female and of normal weight. Three participants were excluded from the analyses because of the lack of eligible data (at least 4 d of 10 h), and one participant did not return the measurement devices. The remaining 30 participants wore the devices for 13.7 (SD = 1.9) hours per day on average.

For steps per day, the absolute mean error percentages was 11.4 percent (median = 7.7 percent). As mentioned previously, this analysis was only applicable for steps per day because of ActiGraph values of zero or close to zero for the other measurements.
Mixed model analyses showed a systematic overestimation of the Fitbit for all activity data. The analyses for aggregation level minute showed a significant overestimation of 0.82 (95% confidence interval [CI] = 0.64–1.00, mean ActiGraph = 9.28) steps per minute (Table 4.3). The analyses for the hour level aggregation showed a significant overestimation of 45.16 (95% CI = 3.12–87.20, mean ActiGraph = 509.46) steps, 2.86 (95% CI = 2.55–3.16, mean ActiGraph = 2.30) minutes of moderate physical activity, 1.03 (95% CI = 0.86–1.19, ActiGraph = 0.29) minutes of vigorous physical activity, and 3.88 (95% CI = 3.49–4.27, ActiGraph = 2.59) minutes of moderate–vigorous physical activity per hour.

The analyses for day level aggregation showed a significant overestimation of 42.82 (95% CI = 36.29–49.34, ActiGraph = 34.49) minutes of moderate physical activity, 15.36 (95% CI = 12.29–18.44, ActiGraph = 4.29) minutes of vigorous physical activity, and 58.18 (95% CI = 50.1–66.25, ActiGraph = 38.78) minutes of moderate–vigorous physical activity per day.

### Table 4.2 - Descriptive information (mean ± SD) for the participants and those in- and excluded from the analyses

<table>
<thead>
<tr>
<th></th>
<th>All participants</th>
<th>Included from analyses</th>
<th>Excluded from analyses¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>34</td>
<td>30</td>
<td>4</td>
</tr>
<tr>
<td>Gender [N= females]</td>
<td>23</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Age [years]</td>
<td>23.9 ± 3.9</td>
<td>23.9 ± 3.9</td>
<td>23.8 ± 5.2</td>
</tr>
<tr>
<td>Self-reported height [cm]</td>
<td>174.8 ± 9.9</td>
<td>173.8 ± 9.3</td>
<td>182.5 ± 12.0</td>
</tr>
<tr>
<td>Self-reported weight [kg]</td>
<td>66.5 ± 8.9</td>
<td>65.2 ± 7.5</td>
<td>76.3 ± 14.1</td>
</tr>
<tr>
<td>Self-reported BMI [kg/m²]</td>
<td>21.7 ± 2.2</td>
<td>21.6 ± 2.2</td>
<td>22.7 ± 2.0</td>
</tr>
</tbody>
</table>

¹ Excluded from analyses, are participants with ineligible data (N=3) or who dropped out (N=1)
### Table 4.3 - Mean values of the ActiGraph data and results of the analyses that were used to assess the agreement, association and correlation of the Fitbit compared with the ActiGraph for the four types of activity data (e.g. steps, moderate, vigorous, moderate-vigorous activity) and four time intervals (i.e. minute \(N=138,224\), hour \(N=2,518\), day \(N=168\))

<table>
<thead>
<tr>
<th>Type of activity data</th>
<th>Time interval</th>
<th>ActiGraph [mean ± SD]</th>
<th>Fitbit [mean ± SD]</th>
<th>Bland-Altman Analyses [mean difference (limits of agreement)]</th>
<th>Systematic differences with mixed models analyses [B (95% CI)]</th>
<th>Association with mixed models analyses [B (95% CI)]</th>
<th>Intra class correlation (95%CI)(^1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steps</td>
<td>Minute</td>
<td>9.28 ± 23.42</td>
<td>10.10 ± 25.97</td>
<td>0.82 (-30.72; 32.36)</td>
<td>0.82 (0.64; 1.00)(^4)</td>
<td>0.72 (0.72; 0.72)(^4)</td>
<td>0.80 (0.79; 0.80)(^1)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>509.46 ± 744.77</td>
<td>554.62 ± 814.99</td>
<td>45.16 (-320.89; 411.21)</td>
<td>45.16 (3.12; 87.20)(^3)</td>
<td>0.89 (0.89; 0.90)(^3)</td>
<td>0.97 (0.97; 0.98)(^3)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>7635.79 ± 4553.39</td>
<td>8312.65± 5135.68</td>
<td>676.86 (-1584.09; 2937.82)</td>
<td>676.86 (-152.78; 1506.50)(^2)</td>
<td>0.86 (0.84; 0.89)(^2)</td>
<td>0.96 (0.91; 0.98)(^2)</td>
</tr>
<tr>
<td>Moderate activity</td>
<td>Minute</td>
<td>0.04 ± 0.20</td>
<td>0.09 ± 0.29</td>
<td>2.86 (-6.73; 12.44)</td>
<td>2.86 (2.55; 3.16)(^3)</td>
<td>0.51 (0.49; 0.54)(^3)</td>
<td>0.57 (0.33; 0.77)(^2)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>2.30 ± 4.86</td>
<td>5.16 ± 6.33</td>
<td>1.31 (-3.82; 5.87)</td>
<td>1.31 (-0.86; 1.19)(^4)</td>
<td>0.50 (0.48; 0.51)(^4)</td>
<td>0.66 (0.54; 0.74)(^4)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>34.49 ± 26.94</td>
<td>77.30 ± 42.79</td>
<td>42.82 (-24.38; 110.01)</td>
<td>42.82 (36.29; 49.34)(^2)</td>
<td>0.39 (0.31; 0.47)(^2)</td>
<td>0.33 (-0.09; 0.61)(^2)</td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>Minute</td>
<td>0.01 ± 0.07</td>
<td>0.02 ± 0.15</td>
<td>1.03 (-3.82; 5.87)</td>
<td>1.03 (0.86; 1.19)(^3)</td>
<td>0.50 (0.48; 0.51)(^3)</td>
<td>0.66 (0.54; 0.74)(^3)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>0.29 ± 2.40</td>
<td>1.31 ± 3.66</td>
<td>15.36 (-11.66; 42.38)</td>
<td>15.36 (12.29; 18.44)(^2)</td>
<td>0.44 (0.38; 0.50)(^2)</td>
<td>0.46 (-0.05; 0.72)(^3)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>4.29 ± 11.55</td>
<td>19.65± 20.00</td>
<td>15.36 (12.29; 18.44)</td>
<td>15.36 (12.29; 18.44)(^2)</td>
<td>0.44 (0.38; 0.50)(^2)</td>
<td>0.46 (-0.05; 0.72)(^3)</td>
</tr>
<tr>
<td>Moderate- Vigorous activity</td>
<td>Minute</td>
<td>0.05 ± 0.21</td>
<td>0.12± 0.32</td>
<td>3.88 (5.88; 13.64)</td>
<td>3.88 (3.49; 4.27)(^3)</td>
<td>0.57 (0.56; 0.59)(^3)</td>
<td>0.67 (0.28; 0.83)(^2)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>2.59 ± 5.72</td>
<td>6.47 ± 8.42</td>
<td>58.18 (-12.66; 129.02)</td>
<td>58.18 (50.11; 66.25)(^2)</td>
<td>0.46 (0.41; 0.52)(^2)</td>
<td>0.37 (-0.10; 0.69)(^1)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>38.78 ± 31.96</td>
<td>96.96± 54.68</td>
<td>58.18 (-12.66; 129.02)</td>
<td>58.18 (50.11; 66.25)(^2)</td>
<td>0.46 (0.41; 0.52)(^2)</td>
<td>0.37 (-0.10; 0.69)(^1)</td>
</tr>
</tbody>
</table>

Note. \(N=\) the number of observation for the aggregated data sets per time interval, \(B=\) regression coefficient, 95% CI= 95% confidence interval \((B ± 1.96 * standard error)\), Bland-Altman Analyses show the mean difference between the Fitbit One and ActiGraph (Fitbit One – ActiGraph) and limits of agreement (mean difference ± 2* standard deviation)

\(^1\)Intra-class correlation coefficient values (ICC) denote the inter-device reliability of the Fitbit and ActiGraph and was used for continuous variables

\(^2\)adjustments for clustering within individuals for the mixed models analyses

\(^3\)adjustments for clustering within individuals and days for the mixed models analyses

\(^4\)adjustments for clustering within individuals, days and hours for the mixed models analyses

Note. N= the number of observation for the aggregated data sets per time interval, B= regression coefficient, 95% CI= 95% confidence interval (B ± 1.96 * standard error), Bland-Altman Analyses show the mean difference between the Fitbit One and ActiGraph (Fitbit One – ActiGraph) and limits of agreement (mean difference ± 2* standard deviation)
The Bland–Altman plots provide a visual representation of the systematic differences and the plots are shown in Figures 4.3–4.5. All Bland–Altman plots (Figures 4.3–4.5) show wide ranges for the LoA. For the assessment of steps per day, per hour, and per minute, the differences between the two measurements (y-axis) seem to be similar for different mean values of the two measurements (x-axis) (Figure 4.3). However, a smaller range in differences in steps is seen for ≥100 mean steps per minute. By contrast, the physical activity intensities tend to show a larger range in differences between the two methods (y-axis) with larger mean values, especially for minutes in moderate–vigorous activity per day (Figure 4.5-C), and in minutes in moderate activity per hour (Figure 4.4-A) and per day (Figure 4.5-A). Finally, there seems to be a positive bias for the comparison of the two methods regarding time spent in moderate–vigorous physical activity per hour (Figure 4.4-C) as well as per day (Figure 4.5-C). This is reflected by the more positive different scores (y-axis) for the larger mean values (x-axis). Table 4.3 shows the results of all Bland–Altman analyses, mixed model analyses, and ICCs.
The development of Active2Gether

The Bland–Altman plots provide a visual representation of the systematic differences and the plots are shown in Figures 4.3–4.5. All Bland–Altman plots (Figures 4.3–4.5) show wide ranges for the LoA. For the assessment of steps per day, per hour, and per minute, the differences between the two measurements (y-axis) seem to be similar for different mean values of the two measurements (x-axis) (Figure 4.3). However, a smaller range in differences in steps is seen for ≥100 mean steps per minute. By contrast, the physical activity intensities tend to show a larger range in differences between the two methods (y-axis) with larger mean values, especially for minutes in moderate–vigorous activity per day (Figure 4.5-C), and in minutes in moderate activity per hour (Figure 4.4-A) and per day (Figure 4.5-A). Finally, there seems to be a positive bias for the comparison of the two methods regarding time spent in moderate–vigorous physical activity per hour (Figure 4.4-C) as well as per day (Figure 4.5-C). This is reflected by the more positive different scores (y-axis) for the larger mean values (x-axis).

Table 4.3 shows the results of all Bland–Altman analyses, mixed model analyses, and ICCs.

Figure 4.3 - Bland-Altman plot for steps per day, hour, and minute for measurements with the ActiGraph and Fitbit, with adjusted means and limits of agreements

Note. Figure 4.2-A shows the Bland-Altman plot for steps per minute, Figure 4.2-B shows the Bland-Altman plot for steps per hour, and Figure 4.2-C shows the Bland-Altman plot for steps per day. The x-axis shows the average number of steps and the y-axis shows the differences in steps between the two methods. The mean of all steps per day, hour and minute per individual are plotted in red.
Figure 4.4 - Bland-Altman plot for minutes moderate, vigorous physical activity (PA) and moderate-vigorous (MVPA) per hour, for measurements with the ActiGraph and Fitbit, with adjusted means and limits of agreements

Note. Figure 4.3-A shows the Bland-Altman plot for minutes of moderate physical activity per hour, Figure 4.3-B shows the Bland-Altman plot for minutes of vigorous physical activity per hour, and Figure 4.3-C shows the Bland-Altman plot for minutes of moderate-vigorous physical activity per hour. The x-axis shows the average of the two measurements and the y-axis shows the differences between the two methods. The mean of all minutes of moderate, vigorous and moderate-vigorous physical activity per hour per individual are plotted in red.

Analyses for the minute level aggregation showed an excellent ICC for steps per minute (ICC = 0.80). Analyses for the hour level aggregation showed an excellent ICC for steps per hour (ICC = 0.97), but ICC values for time spent on moderate, vigorous activity, and moderate-vigorous physical activity were much lower (ICC = 0.57–0.67). Analyses for the day level aggregation showed an excellent ICC for steps...
Figure 4.4 - Bland-Altman plot for minutes moderate, vigorous physical activity (PA) and moderate-vigorous (MVPA) per hour, for measurements with the ActiGraph and Fitbit, with adjusted means and limits of agreements

Note. Figure 4.3-A shows the Bland-Altman plot for minutes of moderate physical activity per hour, Figure 4.3-B shows the Bland-Altman plot for minutes of vigorous physical activity per hour, and Figure 4.3-C shows the Bland-Altman plot for minutes of moderate-vigorous physical activity per hour. The x-axis shows the average of the two measurements and the y-axis shows the differences between the two methods. The mean of all minutes of moderate, vigorous and moderate-vigorous physical activity per hour per individual are plotted in red.

Analyses for the minute level aggregation showed an excellent ICC for steps per minute (ICC = 0.80). Analyses for the hour level aggregation showed an excellent ICC for steps per hour (ICC = 0.97), but ICC values for time spent on moderate, vigorous activity, and moderate–vigorous physical activity were much lower (ICC = 0.57–0.67). Analyses for the day level aggregation showed an excellent ICC for steps per day.
per day (ICC = 0.96), but ICC values for time spent on moderate, vigorous activity, and moderate–
vigorous physical activity were much lower (ICC = 0.33–0.46).

Mixed model analysis for the aggregation level minute were conducted for steps only and showed a
substantial association (B = 0.72, 95% CI = 0.79–0.80). Analyses for aggregation level hour showed an
excellent association for steps (B = 0.89, 95% CI = 0.89–0.90) but moderate associations for time spent
per intensity level (B = 0.51–0.57). Analyses for the aggregation level day showed an excellent
association for steps (B = 0.86, 95% CI = 0.84–0.89), moderate association for time spent in vigorous
and moderate– vigorous physical activity (B = 0.46–0.51), and a poor association for time spent in
moderate physical activity (B = 0.38, 95% CI = 0.31–0.46).

The sensitivity analyses without outliers showed smaller systematic errors for steps per day, hour, and
minute compared with analyses with outliers.

DISCUSSION

This study compared Fitbit One measurements with ActiGraph GT3X+ in terms of assessing physical
activity in different time intervals to assess the construct validity of the Fitbit One for providing real-
time physical activity feedback. The results indicate that the Fitbit One systematically overestimates
physical activity, shows excellent associations with ActiGraph measures for step counts, but only fair
to good associations for time spent in moderate, vigorous physical activity, and combined moderate–
vigorous physical activity. These measurement properties of the Fitbit One were similar in the small
time intervals (i.e. minutes and hours) and the more aggregated time intervals (i.e. days), which
suggests that the Fitbit One is well suited for providing real-time feedback on steps activity for self-
monitoring.

The primary aim of this study was to assess the construct validity of the Fitbit One using smaller time
intervals compared with earlier studies. This focus on the smaller time intervals is important because
even if the Fitbit One was shown to be valid in assessing daily activity, this could be because
measurement errors, i.e. over- and underestimations, can balance out throughout the day, which
would result in good daily validity but not necessarily good minute or hourly validity. Wearables, such
as the Fitbit One, are often used to provide the user with real-time feedback for self-monitoring
purposes, which is increasingly used in physical activity interventions. Hence, minute-by-minute
validity is important as poor validity might be off-putting to users that experience the wearable is not
matching their activities in real time. Therefore, it is important to also know the validity of the Fitbit
using smaller time intervals and, thus, if the Fitbit One can be used for real-time feedback throughout
the day. The results showed a substantial correlation and significant association for steps per minute
and an excellent correlation and significant association for steps per hour of the Fitbit One and
measurement errors, i.e. over- and underestimations, can balance out throughout the day, which
intervals compared with earlier studies. This focus on the smaller time intervals is important because
The primary aim of this study was to assess the construct validity of the Fitbit One using smaller time
monitoring.

suggests that the Fitbit One is well suited for providin g real-time feedback on steps activity for self -
the day. The results showed a substantial correlation and significant association for steps per minute
using smaller time intervals and, thus, if the Fitbit One can be used for real-time feedback throughout
matching their activities in real time. Therefore, it is important to also know the validity of the Fitbit
validity is important as poor validity might be off-putting to users that experience the wearable is not
purposes, which is increasingly used in physical activity interventions. Hence, minute -by-minute
would result in good daily validity but not necessarily good minute or hourly validity. Wearables, such

This study compared F itbit One measurements with ActiGraph GT3X+ in terms of assessing physical
DISCUSSION

moderate physical activity (B = 0.38, 95% CI = 0.31–0.46).

These measurement properties of the Fitbit One were similar in the small

vigorous physical activity. These measurement properties of the Fitbit One were similar in the small

mixed model analysis for the aggregation level  minute were conducted for steps only and showed a

vigorous physical activity were much lower (ICC = 0.33–0.46).

The secondary aim of this study was to assess the construct validity of the Fitbit One for daily physical
activity. Similar results were found as for the analyses with smaller time intervals. An excellent
correlation and significant association were seen for daily step activity; however, time per intensity
level varied substantially among the Fitbit and ActiGraph. These results are in line with Ferguson et al.
who reported a high Pearson correlation for step activity (r = 0.99) and minutes of moderate–
vigorous activity (r = 0.91) and an excellent interdevice agreement for step activity (ICC = 0.95) of the
Fitbit One and ActiGraph, but a lower value for minutes of moderate–vigorous activity (ICC = 0.46). The
Fitbit substantially overestimates daily step activity and daily minutes spent in moderate, vigorous, and
moderate–vigorous physical activity activities compared with the ActiGraph, which is in line with
previous research. However, Gomersall et al. reported that the Fitbit One underestimated
daily minutes of daily minutes of moderate–vigorous physical activity with 19.2 minutes per day. On
the basis of these findings—when being compared with the ActiGraph, the Fitbit is not suitable to
assess intensity levels of physical activity in daily life. The substantial differences between the two
measurements could have arisen from differences in the algorithms that converted the accelerometers
activity counts into intensity levels of physical activity. Unfortunately, Fitbit’s algorithm is not publicly
available, and therefore it remains unclear how the Fitbit calculates the intensity levels of physical
activity. However, it should be noted that in a validation study of Lee et al.—in a laboratory setting—
lower mean absolute error percentages were found between energy expenditure assessed by indirect
calorimetry and estimated energy expenditure by the Fitbit One than energy expenditure estimated
by the ActiGraph. These results indicate, that the Fitbit One might be more suitable for assessing
energy expenditure than the ActiGraph in a laboratory setting.

Although the Fitbit systematically overestimates step activity compared with the ActiGraph, according
to Tudor- Locke et al., the median of the absolute error percentage is within the range of acceptable
error. In general, an error percentage of 10 percent is considered to be acceptable for assessing step activity. Thus, the acceptable systematic differences between and the high correlations of the Fitbit and the ActiGraph strengthen the hypothesis that the Fitbit is suitable for assessing daily step activity and that it could replace more expensive accelerometers in physical activity interventions. Nevertheless, an overestimation of 677 steps per day by the Fitbit One should be taken into account. Furthermore, it should be noted that although accelerometers are commonly used for the assessment of physical activity in daily life, they have some limitations. Accelerometers are limited in their ability to capture upper body movement, and cycling\textsuperscript{198}. Further, accelerometers are not waterproof and thus they are not able to record water activities\textsuperscript{198}. Furthermore, accelerometers also underestimate the energy expenditure for activities while carrying heavy loads because the acceleration patterns remain unchanged, despite the loads carried. Thus, accelerometers such as the ActiGraph and Fitbit One might underestimate total daily physical activity and energy expended\textsuperscript{198}. A strength of this study is the novel approach by assessing the validity of the Fitbit for different time intervals. Although several studies have assessed the validity of the Fitbit One, our study did not only assess the validity of the Fitbit for total daily physical activity but focused on smaller time intervals to determine the validity of the Fitbit for providing real-time feedback for self-monitoring as well. Furthermore, this study used linear mixed models analyses—and thus adjusted for clustering—to determine the correlations and systematic differences for the measurements of the Fitbit and ActiGraph in addition to the traditional analyses (e.g. Bland–Altman analyses and ICC), which do not adjust for the dependency of the data. In addition, the protocol was designed to be comparable with previous validation studies\textsuperscript{207-209}. Lastly, this study used the API to obtain the Fitbit data, instead of manually copying the displayed data on the website or app that is prone to errors. This study also has some limitations. First, the ActiGraph GT3x+ was used as a reference measurement, and the limitations of this accelerometer, which are mentioned previously, should be taken into account. Second, we used Troiano’s cut points to categorize time spent in different intensities what may have influenced the results: the cut points are based on regression models and do not adjust for the individual’s characteristics and Fitbit’s algorithm to categorize time spent at different intensities are unknown and most likely differ from Troiano’s cut points. Troiano’s cut points (or any other ActiGraph cut points) are certainly not a gold standard for determining physical activity and have their own shortcomings. This is also illustrated by the results of Lee et al.\textsuperscript{191} discussed previously that indicated the Fitbit One might be more suitable for the assessment of energy expenditure than the ActiGraph when using the traditional linear regression models. Third, because of dichotomous data for the intensity levels per minute, it was not possible to assess the validity for minutes of moderate, vigorous, and moderate-vigorous physical activity per minute. Lastly, the current study assessed the validity of
the Fitbit while worn on the hip. Therefore, the current study cannot generalize the findings to situations in which the Fitbit is worn on other sides (e.g. wrist or bra).

In conclusion, this is the first study examining the validity of the Fitbit One for the assessment of levels of physical activity in real-life using smaller time intervals, which is essential for high quality self-monitoring, which is increasingly used in physical activity intervention studies. Although the Fitbit One over-estimates the step activity somewhat, the Fitbit One can be considered a valid device to assess step activity even for real-time minute-by-minute self-monitoring. However, compared with the ActiGraph, the Fitbit One appears less suitable to assess time spent in moderate activity, vigorous activity, and moderate–vigorous physical activity. To date, these data are not shown to the user and validity should be improved before this should be done. Further research is needed to examine whether the Fitbit One can be used as a (scientific) measurement tool. Thus, on the basis of our findings, the Fitbit One appears as suitable to deliver real-time feedback based on step data for self-monitoring, even for small time intervals as minute and hour intervals. Furthermore, the Fitbit One also appears as a valid instrument to assess daily step activity.
APPENDIX

Appendix 4.5 - Results of the analyses that were used to assess the agreement, association and correlation of the Fitbit compared with the ActiGraph for the four types of activity data (e.g. steps, moderate, vigorous, moderate-vigorous activity) and four time intervals (i.e. minute (N=138,224), hour (N=2,518), day (N=168) ) without outliers

<table>
<thead>
<tr>
<th>Outcome measurement</th>
<th>Agreement analyses</th>
<th>Association and Correlation analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time interval</td>
<td>Bland-Altman Analyses [mean difference (limits of agreement)]</td>
</tr>
<tr>
<td>Type of activity data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Steps</td>
<td>Minute</td>
<td>1.02 (-21.37; 23.41)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>34.81 (-198.70; 268.32)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>622.74 (-1155.49; 2400.97)</td>
</tr>
<tr>
<td>Moderate activity</td>
<td>Minute</td>
<td>2.75 (-6.37; 11.87)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>42.82 (-22.49; 106.64)</td>
</tr>
<tr>
<td></td>
<td>Day</td>
<td>15.29 (-3.77; 5.74)</td>
</tr>
<tr>
<td>Vigorous activity</td>
<td>Minute</td>
<td>0.93 (-3.77; 5.74)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>15.29 (-11.74; 42.32)</td>
</tr>
<tr>
<td>Moderate-Vigorous activity</td>
<td>Minute</td>
<td>3.74 (-5.50; 12.97)</td>
</tr>
<tr>
<td></td>
<td>Hour</td>
<td>57.36 (-10.43; 125.15)</td>
</tr>
</tbody>
</table>

Note. N= the number of observation for the aggregated data sets per time interval. B= regression coefficient. 95% CI= 95% confidence interval

1 Intra-class correlation coefficient values (ICC) denote the inter-device reliability of the Fitbit and ActiGraph and was used for continuous variables

2 Adjustments for clustering within individuals for the mixed models analysis

3 Adjustments for clustering within individuals and days for the mixed models analyses

4 Adjustments for clustering within individuals, days and hours for the mixed models analyses