

# Human Cues in Self-help Lifestyle Interventions: an Experimental Field Study

Talia R Cohen Rodrigues, Thomas Reijnders, David R de Buissonjé, Prabhakaran Santhanam, Tobias Kowatsch, Veronica R Janssen, Roderik A Kraaijenhagen, Douwe E Atsma, Andrea W M Evers

Submitted to: Journal of Medical Internet Research  
on: April 29, 2021

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

## *Table of Contents*

---

<b>Original Manuscript</b> .....	<b>5</b>
<b>Supplementary Files</b> .....	<b>33</b>
Figures .....	34
Figure 1.....	35
Figure 2.....	36
Multimedia Appendixes .....	37
Multimedia Appendix 1.....	38
Multimedia Appendix 2.....	38

Preprint  
JMIR Publications

# Human Cues in Self-help Lifestyle Interventions: an Experimental Field Study

Talia R Cohen Rodrigues<sup>1</sup> MSc; Thomas Reijnders<sup>1,2</sup> PhD; David R de Buissonjé<sup>1</sup> MSc; Prabhakaran Santhanam<sup>3</sup> MSc; Tobias Kowatsch<sup>3,4</sup> Dr; Veronica R Janssen<sup>1,5</sup> Dr; Roderik A Kraaijenhagen<sup>6,7</sup> Dr; Douwe E Atsma<sup>5</sup> Prof Dr; Andrea W M Evers<sup>1,8,9</sup> Prof Dr

<sup>1</sup>Health, Medical, and Neuropsychology Unit Faculty of Social and Behavioural Sciences Leiden University Leiden NL

<sup>2</sup>Department of Human-Centered Design Faculty of Industrial Design Engineering Technical University of Delft Delft NL

<sup>3</sup>Centre for Digital Health Interventions Department of Management, Technology, and Economics ETH Zürich Zürich CH

<sup>4</sup>Centre for Digital Health Interventions Institute of Technology Management University of St.Gallen St. Gallen CH

<sup>5</sup>Department of Cardiology Leiden University Medical Center Leiden NL

<sup>6</sup>NDDO Institute for Prevention and Early Diagnostics (NIPED) Amsterdam NL

<sup>7</sup>Vital10 Amsterdam NL

<sup>8</sup>Department of Psychiatry Leiden University Medical Center Leiden NL

<sup>9</sup>Medical Delta Leiden University, Technical University of Delft, Erasmus University Rotterdam Leiden, Delft, Rotterdam NL

## Corresponding Author:

Talia R Cohen Rodrigues MSc  
Health, Medical, and Neuropsychology Unit  
Faculty of Social and Behavioural Sciences  
Leiden University  
Wassenaarseweg 52  
Leiden  
NL

## Abstract

**Background:** Self-help eHealth interventions are generally less effective than human-supported ones, as they suffer from a low level of adherence. Nevertheless, self-help interventions are useful in the prevention of non-communicable diseases, as they are easier and cheaper to widely implement. Adding humanness in the form of a text-based conversational agent (TCA) could provide a solution to non-adherence. In this study we investigate whether adding human cues to a TCA facilitates relationship-building with the agent, and makes interventions more attractive for people to adhere to. We will investigate the effects of two types of human cues, which are visual cues (eg, human avatar) and relational cues (eg, showing empathy).

**Objective:** We aim to investigate if adding human cues to a TCA can help increase adherence to a self-help eHealth lifestyle intervention and explore the role of working alliance as a possible mediator of this relationship.

**Methods:** Participants (N=121) followed a 3-week app-based physical activity intervention delivered by a TCA. Two types of human cues used by the TCA were manipulated, resulting in four experimental groups, which were (1) visual cues-group, (2) relational cues-group, (3) both visual and relational cues-group, and (4) no cues-group. Participants filled out the Working Alliance Inventory Short Revised form after the final day of the intervention. Adherence was measured as number of days participants responded to the messages of the TCA.

**Results:** One-way ANOVA revealed a significant difference for adherence between conditions. Against our expectations, the groups with visual cues showed lower adherence compared to those with relational only or no cues ( $t(117) = -3.415, P = .001$ ). No significant difference was found between the relational- and no cues-groups. Working alliance was not affected by cue-type, but showed to have a significant positive relationship with adherence ( $t(75) = 4.136, P < .001$ ).

**Conclusions:** We hypothesize that the negative effect of visual cues is due to a lack of transparency about the true nature of the coach. Visual resemblance of a human coach could have led to high expectations that could not be met by our digital coach. Furthermore, the inability of TCAs to use non-verbal communication could provide an explanation for the lack of effect of relational cues or the effect of cue-type on working alliance. We give suggestions for future studies to test these potential mechanisms. Clinical Trial: Pre-registration: OSF Registries, <https://osf.io/mgw2s>

(JMIR Preprints 29/04/2021:30057)

DOI: <https://doi.org/10.2196/preprints.30057>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.  
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#)

## Original Manuscript

Preprint  
JMIR Publications

**Human Cues in Self-help Lifestyle Interventions: an Experimental Field Study**

Talia R. Cohen Rodrigues<sup>1</sup>, Thomas Reijnders<sup>1,2</sup>, David R. de Buissonjé<sup>1</sup>, Prabhakaran Santhanam<sup>3</sup>, Tobias Kowatsch<sup>3,4</sup>, Veronica R. Janssen<sup>1,5</sup>, Roderik A. Kraaijenhagen<sup>6,7</sup>, Douwe E. Atsma<sup>5</sup>, Andrea W.M. Evers<sup>1,8,9</sup> on behalf of the BENEFIT consortium

<sup>1</sup>*Health, Medical, and Neuropsychology Unit, Leiden University, The Netherlands*

<sup>2</sup>*Department of Human-Centered Design, Faculty of Industrial Design Engineering, TU Delft, Delft, The Netherlands*

<sup>3</sup>*Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zürich, Zurich, Switzerland*

<sup>4</sup>*Centre for Digital Health Interventions, Institute of Technology Management, University of St.Gallen, St.Gallen, Switzerland*

<sup>5</sup>*Department of Cardiology, Leiden University Medical Center, The Netherlands*

<sup>6</sup>*NDDO Institute for Prevention and Early Diagnostics (NIPED), Amsterdam, The Netherlands*

<sup>7</sup>*Vital10, Amsterdam, the Netherlands*

<sup>8</sup>*Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands*

<sup>9</sup>*Medical Delta, Leiden University, Technical University of Delft, Erasmus University Rotterdam*

## ABSTRACT

**Background:** Self-help eHealth interventions are generally less effective than human-supported ones, as they suffer from a low level of adherence. Nevertheless, self-help interventions are useful in the prevention of non-communicable diseases, as they are easier and cheaper to widely implement. Adding humanness in the form of a text-based conversational agent (TCA) could provide a solution to non-adherence. In this study we investigate whether adding human cues to a TCA facilitates relationship-building with the agent, and makes interventions more attractive for people to adhere to. We will investigate the effects of two types of human cues, which are visual cues (eg, human avatar) and relational cues (eg, showing empathy).

**Objective:** We aim to investigate if adding human cues to a TCA can help increase adherence to a self-help eHealth lifestyle intervention and explore the role of working alliance as a possible mediator of this relationship.

**Methods:** Participants (N=121) followed a 3-week app-based physical activity intervention delivered by a TCA. Two types of human cues used by the TCA were manipulated, resulting in four experimental groups, which were (1) visual cues-group, (2) relational cues-group, (3) both visual and relational cues-group, and (4) no cues-group. Participants filled out the Working Alliance Inventory Short Revised form after the final day of the intervention. Adherence was measured as number of days participants responded to the messages of the TCA.

**Results:** One-way ANOVA revealed a significant difference for adherence between conditions. Against our expectations, the groups with visual cues showed lower adherence compared to those with relational only or no cues ( $t(117) = -3.415, P = .001$ ). No significant difference was found between the relational- and no cues-groups. Working alliance was not affected by cue-type, but showed to have a significant positive relationship with adherence ( $t(75) = 4.136, P < .001$ ).

**Conclusions:** We hypothesize that the negative effect of visual cues is due to a lack of transparency about the true nature of the coach. Visual resemblance of a human coach could have led to high expectations that could not be met by our digital coach. Furthermore, the inability of TCAs to use non-verbal communication could provide an explanation for the lack of effect of relational cues or the effect of cue-type on working alliance. We give suggestions for future studies to test these potential mechanisms.

**Pre-registration:** OSF Registries, <https://osf.io/mgw2s>

**Keywords:** eHealth; digital health; lifestyle change; physical activity; intervention; conversational

agent; chatbot; adherence; working alliance.





Non-communicable diseases (eg, cardiovascular diseases, type-2 diabetes) are the leading cause of death globally [1]. Engaging in a healthy lifestyle can help in the treatment and prevention of many of those diseases [2, 3]. This could be facilitated by eHealth, which are digital tools that can be used by a healthcare professional to provide remote support, or that can provide automated support [4]. Studies show that eHealth is effective in improving lifestyle behaviors, and in the prevention and treatment of non-communicable diseases [5, 6]. Automated self-help interventions are easier and cheaper to widely implement as they require no interference of healthcare professionals, who indicate to experience barriers in lifestyle support such as a lack of time or insufficient experience with lifestyle support [7-9]. However, meta-analyses show that human-supported digital lifestyle interventions are more effective than self-help ones [5, 10, 11]. Adherence, or the extent to which a person uses the eHealth intervention as intended, is a problematic within self-help interventions [12-15]. As intervention adherence is related to more positive health outcomes [16], finding ways to make people adhere to self-help interventions would be necessary to reach optimal effectiveness.

Although human-supported interventions generally outperform self-help ones, this does not necessarily imply the support of a healthcare professional. Meta-analyses revealed that contact with a nonprofessional or administrative support by a human being is enough to both ensure intervention effectiveness and prevent people from dropping out of the intervention [17-20]. It seems that some level of “humanness” rather than professional guidance is the key ingredient within human-supported interventions. The underlying reason why people would like a level of humanness into the intervention, could be the need of a personal relationship with their care provider [21]. This so-called working alliance can be defined as the level of agreement on goals that are set for treatment, on tasks that must be performed to reach this goal, and the relational bond between healthcare professional and patient [22, 23]. Working alliance with a human care provider is a predictor of intervention adherence and effectiveness both in regular face-to-face [24, 25] and in digital therapy or treatment [26, 27]. However, people are also able to form relationships with computers. People interact with computers as they would do with human beings, and apply similar social rules and heuristics [28, 29]. Studies show that people can also develop a working alliance within fully automated digital interventions, and that this leads to more positive treatment outcomes [30-34].

In self-help interventions, humanness can be added by the use of a so-called conversational agent (CA). These computer-based agents can mimic human-like conversational behavior (eg, respond to input, generate output, apply turn-taking) [35], and be used to provide automated support.

An embodied conversational agent (ECA) is visually present on screen and can provide non-verbal cues (eg, hand gestures), while a text-based conversational agent (TCA) is able to communicate with text only [36]. A TCA has the advantage of being easier to develop, being easier to apply in a mobile app, and therefore being more suitable for widespread implementation. Studies show that people show more relational behaviors or feel more social presence when they believe that their interaction partner is a human being rather than a computer [37, 38]. To enhance these perceptions while interacting with CAs, human cues could be applied, such as an avatar of a human being, a human tone-of-voice [39], or lower speed of feedback [40]. Furthermore, non-verbal communication can be replaced by adding emoticons [41]. Besides the appearance of the messages and CA, human cues could be applied to the content of its messages and behavior. It is possible to add human conversation rules in computer-generated conversations, such as humor, empathy, and small talk, which are often used by humans to establish a relationship [42, 43]. Studies with ECAs show that applying such human cues to the interaction increases the working alliance users experience with the ECA [42] and their intention to use the ECA [44].

Although studies about digital interventions with TCAs have been conducted before, only a small amount has focused on their application in lifestyle change interventions [45]. Additionally, the effects of human cues are predominantly tested with ECAs. Therefore there is little knowledge about how human cues affect people's relationship with TCAs, or their adherence to TCA-supported interventions. Furthermore, the majority of studies test the effects of either using human cues or not, while we are interested in the effects of two different types of human cues and how these interact. In this study we will focus on both the appearance and behavior, or in other words, the visual and relational cues, that can be integrated into a TCA to increase the level of humanness. We predict that these will improve the working alliance people experience with the TCA, and in turn their adherence to the intervention. To test this, we conducted a field-experiment in which people followed a 3-week app-based physical activity intervention with automated support from a TCA. We manipulated the cues used by the TCA, and measured the working alliance that people experience with the TCA, and the number of days they adhere to the intervention. This allows us to answer the following research questions:

RQ1: Is there an effect of human cue type (visual and relational cues) on adherence?

RQ2: Does working alliance mediate the effect of human cue type (visual and relational cues)

on

adherence?

## METHODS

### Study Design

The three-week field experiment was conducted in March and April 2020. We employed a 2 (visual cues: yes, no) x 2 (relational cues: yes, no) between-subjects design. There were four experimental groups, in which the TCA used (1) visual and relational cues, (2) visual cues, (3) relational cues, or (4) no human cues.

### Participants

We recruited (N = 269) healthy participants between 18 and 30 years old with flyers on the university campus and via social media, who were willing to work on their level of physical activity, had access to a smartphone running iOS or Android, and were sufficiently proficient in English. We excluded participants who were not able to engage in a normal physical activity pattern using the Physical Activity Readiness Questionnaire (PAR-Q) [46]. Students from Leiden University received credits (required to complete their first year) for their participation, and all participants who would complete the study would enroll in a lottery (with the chance of winning one of the three Fitbit devices, or one of the 100 vouchers worth €10,-). Power calculations (G\*Power) [47] identified a minimum sample size of 128 to detect a medium between-group effect ( $f = .25$ ) of cue-type with an alpha of .05 (ANOVA with 4 groups). Given the high attrition rates in similar studies (eg, [48]), we aimed to recruit about double the required number of participants. All participants provided their consent before the start of the experiment, and the study was approved by the Psychology Research Ethics Committee of Leiden University.

### Benefit StepCoach Intervention

#### Intervention

The aim of the intervention was to enhance participants' physical activity levels by increasing daily step counts. The intervention consisted of daily exercises, such as a quiz about the health consequences of physical activity and a decisional balance worksheet (see Appendix 1 for overview of daily exercises). These exercises would take about 5 to 10 minutes each day to complete. The exercises were based on behavior change techniques (BCTs), such as prompts, information about health consequences, review of goals, and social reward [49]. These are intervention components designed to regulate behavior by reinforcing factors that facilitate behavior change, and mitigating factors that hinder behavior change. These behavior change techniques were incorporated following the Transtheoretical Model of health behavior change [50], which views behavioral change as an

upward spiral process involving progress through five stages (i.e. pre-contemplation, contemplation, preparation, action, maintenance). The model has been used to target a wide range of health behaviors [51].

### ***Technical Implementation of the Benefit StepCoach for Android and Apple's iOS Platforms***

The Benefit StepCoach app was implemented with MobileCoach ([www.mobile-coach.eu](http://www.mobile-coach.eu)) [52,53, 54], an open-source software platform for smartphone-based and chatbot-delivered behavioral interventions (eg, [55]) and ecological momentary assessments (eg, [56]). The platform provides a web-based graphical user interface for intervention authors and allows them to implement the intervention logic and content. MobileCoach uses a web server to execute the intervention logic and to deliver the content to the MobileCoach-based mobile applications for Apple's iOS and Android platforms. The mobile app was customized to fit the needs of this study and published in the iOS and Android app stores with the name Benefit StepCoach. One of the important features of this app was to automatically and objectively retrieve step counts of the participants. Google Fit ([www.google.com/fit/](http://www.google.com/fit/)) for the Android app and Apple's Health Kit ([developer.apple.com/documentation/healthkit](http://developer.apple.com/documentation/healthkit)) for the iOS app were used for this purpose. Appropriate interactions were implemented, i.e. asking participants for their permission, to allow the apps to access the step data. Moreover, each experimental group was assigned a dedicated TCA.

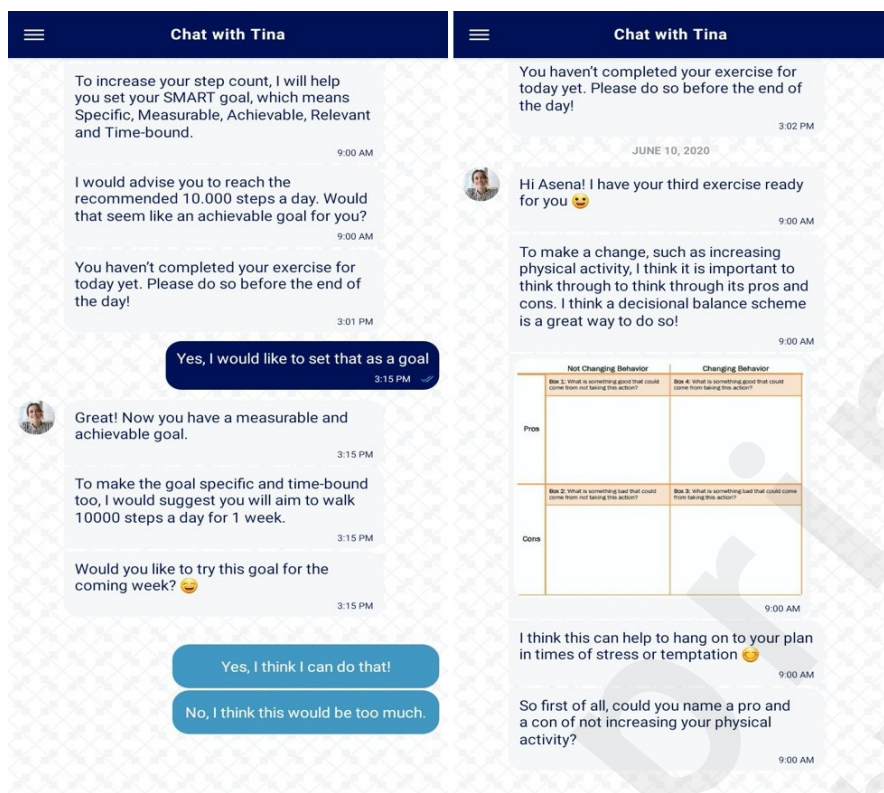
### ***Text-based Conversational Agent***

Participants interacted daily with a TCA, the virtual coach who delivered the intervention and offered various conversational turns. Via the chat feature, the TCA delivered daily exercises in line with the intervention and would respond to messages of the participants via conversational turns (see Figure 1). All conversational turns were scripted. Each day would consist of two to four conversational turns. The first message would be sent in the morning (9:00 am), and the following messages after a reply of the participant. If the participant would not have replied yet, the TCA would send a reminder in the afternoon (3:00 pm).





Across the experimental groups, the intervention (eg, tasks and feedback) was identical, but the conversational turns differed in the type of cues the TCA used. We manipulated two types of human cues: (1) visual cues, which were related to the humanness of the communication style, and the design and appearance of the messages (human avatar, use of emoticons, human tone-of-voice, and response delay), and (2) relational cues, which were related to the content of the messages, and to what extent these followed social scripts and human conversation rules (eg, showing empathy,

self-disclosure, humor, small talk, and meta-relational communication) (see Figure 2).

**Figure 1.** Screenshots of Benefit StepCoach app.



**Figure 2.** Example of conversational turns per condition.

Visual & relational cues	Visual cues	Relational cues	No cues
 <p>Hi [name]! I really appreciate all the work you're doing, so keep it up 😊</p>	 <p>Hi [name]! 😊</p>	 <p>Hi [name]! All the work you're doing is impressive, so keep it up.</p>	 <p>Hi [name]!</p>
<p>I can imagine that it is quite hard to work on your goal alone. We've already discussed many ways to achieve your step goal yourself, but in my experience social support can be very helpful as well 😊</p>	<p>We've already discussed many ways to achieve your step goal yourself, but I think social support can be very helpful as well 😊</p>	<p>It can be quite hard to work on your goal alone. You've already worked out many ways to achieve your step goal yourself, but social support can be very helpful as well.</p>	<p>You've already worked out many ways to achieve your step goal yourself, but social support can be very helpful as well.</p>
<p>Do you have a friend or family member that would be willing to help you? Could you tell me his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could you tell me his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could share his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could share his or her name?</p>

**Measures**

### ***Adherence***

Adherence was measured based on the number of days participants finished the session of conversational turns. Participants were marked as “adherent” for a particular day if they had replied to the final message of the TCA before the end of the day (12:00 pm at midnight). Given the duration of the intervention, the level of adherence over the whole study could range between 1 and 21 days.

### ***Physical Activity***

We measured physical activity through objective step count data retrieved from Apple Health or Google Fit (depending on the smartphone of the participant). Effectiveness was based on the baseline average step count in the week before the intervention, as well as on the step counts retrieved during the intervention itself.

To assess baseline levels of physical activity, the International Physical Activity Questionnaire Short Form (IPAQ-SF) [57] was used. The questionnaire consists of seven items asking the participants about their time spent on vigorous and moderate physical activities, walking, and sitting during the previous week. The output is a MET score, representing the amount of energy used to carry out the reported physical activities. The IPAQ-SF has been shown to have a high reliability, but minimal validity [57, 58]. Therefore we decided to additionally use objective step count as baseline measurement.

### ***Working Alliance***

Working alliance with the TCA was measured with an adjusted version of the Working Alliance Inventory Short Revised form (WAI-SR) [59]. The WAI-SR consists of 12 items measured on a 5-point Likert-type scale ranging from 1 (seldom) to 5 (always), subdivided in 3 subscales: agreement on tasks, agreement on goals, and bond. Questions were adjusted to fit the context of the study by using the words “coach”, “lifestyle” and “intervention” (eg, “The coach and I collaborate on setting lifestyle goals.”). The WAI-SR has been shown to have sufficient reliability and validity [59], and our adjusted version showed to have a high internal consistency (Cronbach’s  $\alpha = .945$ ).

### ***Procedure***

A week before the start of the intervention 282 participants provided digital informed consent and filled in a screening survey assessing the inclusion and exclusion criteria. 226 eligible participants received a link to the iOS or Android app store to download the Benefit StepCoach app. Once the app was downloaded, participants were asked to go through the onboarding procedure to correctly configure the app (eg, allowing push messages and access to step count data via Apple Health or Google Fit), and to complete the baseline survey. Participants were reminded through

emails and text messages to complete the onboarding and baseline survey (measuring demographics and baseline characteristics) after 3, 4 and 5 days, and excluded if they did not do so before the start of the intervention. Participants were allocated to one of the four conditions by an automated mechanism within the app. All participants started simultaneously with the three-week (21 days) intervention. Each day, the TCA would send the participants one or several short exercises to complete that day (eg, quiz or worksheet, see Appendix 1 for overview of daily exercises) via a push notification. After completing the final survey on day 22 (measuring Working Alliance), participants would receive the debriefing.

### Data Analysis

The analyses were preregistered via the Center for Open Science [60]. Intervention effectiveness independent of cue-type was tested using a one-tailed dependent samples *t* test in which we compared average step count during the baseline week with the final week of the intervention. For the first research question (Is there an effect of human cue type (visual and relational cues) on adherence?), we predicted that the condition with both visual and relational cues would lead to the highest adherence, followed by the conditions with either visual or relational cues, and the condition with no human cues. To test this, a one-way between-subjects ANOVA with planned contrasts was conducted. Because the differences in mean adherence of each group were different than expected, the planned contrasts we ran were different from those pre-registered. In the first alternative post-hoc analysis we compared the visual & relational cues- and visual cues-conditions with the relational cues- and no cues-conditions. In the second, we compared the relational cues-condition with the no human cues-condition.

For the second research question (Does working alliance mediate the effect of human cue type on adherence?), we predicted that working alliance with the TCA would be highest in the condition with both visual and relational cues, followed by the conditions with either visual or relational cues, and then the condition with no human cues. In turn, we expected that a higher working alliance would lead to a higher adherence. A one-way between-subjects ANOVA was conducted to compare the working alliance with the TCA between the four different conditions. As we did not find significant differences, we did not conduct our preregistered mediation analysis. We did run an additional regression analysis with working alliance as independent and adherence as dependent variable. Statistical analyses were conducted with SPSS (version 26; IBM Corp) using a *p*-value of .05 as the level of significance.

## RESULTS

### Demographics

A total number of 127 participants downloaded the app, after which 121 participants completed the baseline measurement. These participants were on average 22.7 years (SD = 2.8) old, 84/121 (69%) were female, 73/121 (60%) were of Dutch nationality, and of 91/121 (75%) their current or highest education level was bachelor's degree or higher. One-way ANOVA's showed no difference in baseline physical activity (MET score on IPAQ),  $F(3, 115) = .583$ ,  $P = .63$ , or baseline average steps per day between the experimental groups,  $F(3, 62) = .769$ ,  $P = .52$  (see Table 1).

**Table 1.** Baseline demographic characteristics (N = 121)

Variable	Visual & relational cues (n = 31)	Visual cues (n = 24)	Relational cues (n = 29)	No cues (n = 37)
Age in years, mean (SD)	22.65 (2.84)	22.71 (2.79)	22.76 (2.70)	22.54 (3.01)
Gender, female, n (%)	26 (84)	12 (50)	21 (75)	25 (68)
<b>Nationality, n (%)</b>				
Dutch	19 (61)	15 (63)	14 (48)	25 (67.5)
German	3 (10)	3 (13)	6 (21)	5 (13.5)
Other	9 (29)	6 (25)	9 (31)	7 (19)
<b>Education level, n (%)</b>				
High school	4 (13)	6 (25)	6 (21)	11 (30)
Vocational school	1 (3)	1 (4)	0 (0)	1 (3)
Bachelor's degree	17 (55)	14 (58)	21 (72)	18 (49)
Master's degree or higher	9 (29)	3 (13)	2 (7)	7 (19)
<b>Physical activity level, mean (SD)</b>				
MET score (per week)	4556 (5324)	2928 (5370)	3800 (3373)	3854 (3804)
Average steps per day	3282 (2289)	1912 (1557)	3266 (1601)	3361 (2616)

### Intervention Effectiveness



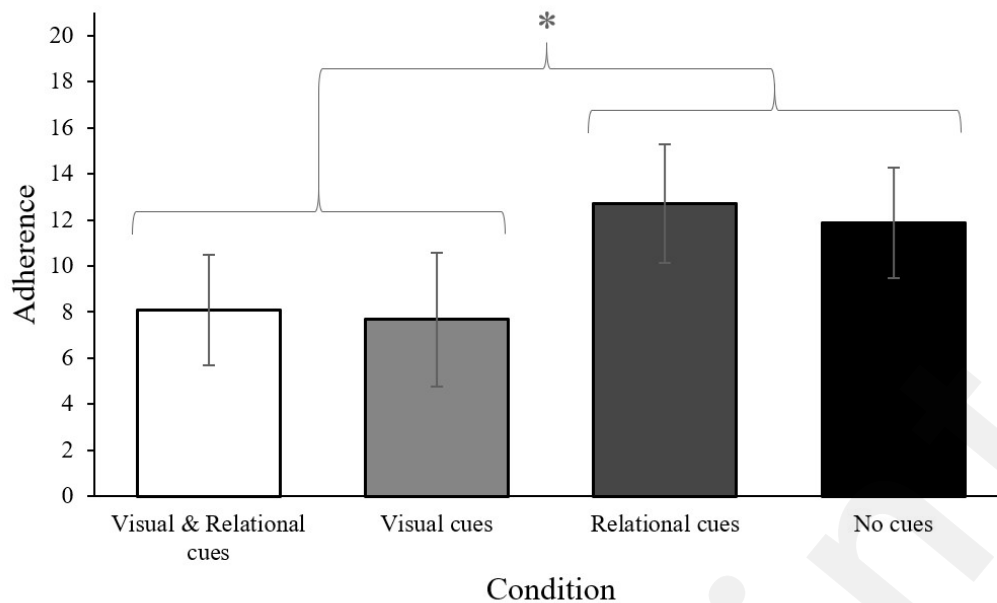
In our preregistration, we proposed to test hypotheses with effectiveness as outcome variable. Our power calculations identified a minimum sample size of 128 to detect the expected effects of experimental group on effectiveness. However, as we had insufficient cases with both a valid baseline step count and a minimum of 5 days of step count registered in the final week, we did not have enough power to detect this effect. We therefore decided to focus in this paper on adherence as outcome variable, and report the analyses with effectiveness as outcome variable in Appendix 2.

To test whether the intervention was effective (independently of the experimental condition), we used the cases that had both a valid baseline step count and enough days of steps registered in the final week to conduct a (one-tailed) paired samples *t* test to compare the average step count in the baseline week with the average step count in the final week. There was a significant increase in the average step count from the baseline week ( $M=3412.37$ ,  $SD=2363.17$ ) to the final week ( $M=4556.77$ ,  $SD=2545.65$ ),  $t(42)=-3.975$ ,  $P < .001$ .

### Adherence

We found a significant difference in adherence between the conditions,  $F(3, 117) = 3.901$ ,  $P = .011$  (see Table 2 for mean and SD per group). By visually inspecting the means, we saw that the differences between groups were not as expected (see Figure 3). The contrast analyses showed that in the relational cues- and no cues-conditions there was a significantly higher adherence than in the other two conditions,  $t(117) = -3.415$ ,  $P = .001$ . However, adherence in the relational cues condition was not higher than in the no human cues condition,  $t(117) = .458$ ,  $P = .65$ . So contrary to what was expected, participants were less adherent to the intervention in the groups in which the TCA used visual cues compared to the groups without visual cues. Furthermore, when the TCA used relational cues, participants were not more adherent than when the TCA used no human cues at all.

**Figure 3.** Mean adherence (number of days participants finished the session of conversational turns) per experimental condition, with 95% confidence intervals, including the post-hoc contrast between groups with and groups without visual cues.



### Working Alliance

There was no significant difference in working alliance between the conditions,  $F(3, 73) = 1.110$ ,  $P = .35$  (see Table 2 for mean and SD per group). However, we did find a positive relationship between working alliance and adherence,  $b = .202$ ,  $t(75) = 4.136$ ,  $P < .001$ . These outcomes indicate that adding human cues did not lead to a better (or worse) reported working alliance with the TCA, but that at the same time, participants who reported a better working alliance were more adherent to the intervention.

**Table 2.** Means and standard deviations per group of adherence (number of days participants finished the session of conversational turns) and working alliance (measured after the final day of the intervention with the WAI-SR).

Variables	Adherence		Working alliance	
	N	Mean (SD)	N	Mean (SD)
<b>Visual &amp; relational cues</b>	37	8.10 (6.70)	19	35.26 (13.15)
<b>Visual cues</b>	24	7.67 (7.08)	14	41.07 (10.80)
<b>Relational cues</b>	29	12.69 (6.93)	22	39.59 (9.08)
<b>No cues</b>	37	11.89 (7.30)	25	35.92 (11.99)

### DISCUSSION

We investigated if and how a TCA could help increase adherence to a self-help eHealth lifestyle

intervention. Regarding our first research question (Is there an effect of human cue type (visual and relational cues) on adherence?), the results of our field experiment showed that, contrary to our expectations, the use of human cues by the TCA did not lead to a higher adherence. In contrast, visual cues even led to a lower adherence. To answer the second research question (Does working alliance mediate the effect of human cue type (visual and relational cues) on adherence?), we found that human cues did not lead to a higher working alliance with the TCA. We did, however, find that a better reported working alliance was related to a better adherence to the intervention.

Our results did not show the positive effects of human cues and visual elements that have been reported in previous studies [30, 39, 42, 43]. On the contrary, we even found a negative effect on intervention adherence for visual cues. One reason for this could be that, before the start of the intervention, we did not tell participants whether they would be coached by a human being or a computer. This lack of transparency might have led to expectations that could not be met by the TCA, which might have led to frustration among users [61]. Many studies however show a positive effect on user perceptions and user behavior of not disclosing the nature of an automated chatbot, or of suggesting that users are interacting with a human being while they are not [62-64]. However, Mozafari and colleagues [65] show that the effects of disclosure depend on whether there are errors in the conversation with a chatbot. In their study with a customer-service bot, they found that when the chatbot was not able to solve a customer's issue, the negative responses to these errors could be prevented by disclosing the chatbots true nature beforehand. Although our study concerned a lifestyle intervention, similar effects could have occurred. As visual cues might have wrongly suggested communication with a human being and our CA was not always able to respond fully correctly (as the messages were preprogrammed), correctly informing participants about the nature of the agent could have prevented the negative effects of the errors within the conversations. Furthermore, the avatar we used in the visual cues conditions might have played a role. We intentionally chose a younger and healthy-looking female agent to resemble the psychology student population, and because a young female peer agent is generally preferred in health coaching tasks [66, 67]. However, some literature suggest that male agents are preferred as athletic trainer, which might have influenced the results if our participants perceived the TCA to be an athletic coach rather than a health coach [68]. Furthermore, another study shows that non-ideal overweight agents are seen as more trustworthy and lead to higher use intentions [69], which suggests our TCA might have been too slender and healthy looking. Additionally, similarity between CA and target population can have a downside when the agent is perceived as unhelpful [70]. The user's perception of the helpfulness of

the agent depends on the goal of the user [71], and as our participants might have participated in the intervention with other goals than increasing physical activity (eg, gaining participant-credits), the similar-looking agent might have been unhelpful to them, leading to a lower adherence. All in all, the negative effect of the visual cues on adherence could therefore have been due to a lack of transparency about the true nature of the CA, and the type of visual cues we applied to our CA.

The absence of an effect for relational cues in our study contradicts previously mentioned studies in which a positive effect was found [42, 43]. However, what is important to note is that these studies concern ECAs, while we used a TCA. ECAs generally outperform text-based ones [44, 72], which can be explained by an additional range of design characteristics an ECA can make use of [73]. In one study though, there was no difference found between a TCA and an ECA, which the authors argued was due to the lack of incorporating non-verbal communication in the latter one [74]. The inability of our (or any) TCA to use non-verbal communication, might be the reason that relational cues did not have the same effects as in studies with ECAs. Similar patterns occur in human-human computer-mediated communication, where people are limited in their use of non-verbal communication. Text-based communication would not be rich enough to transfer ambiguous communication, such as relational behavior [75], and relationship building requires more time in text-based environments to reach the same quality as in face-to-face situations [76]. Possibly, the lack of effect of relational cues could be caused by the text-based nature of our CA, as CAs could possibly need non-verbal communication in order to make relational cues work.

Finally, we found that people who reported a better working alliance with the CA were more adherent to the intervention. This result is in line with studies about regular face-to-face interventions [24, 25], digital therapy or treatment [26, 27], and automated digital interventions [30, 31, 32, 33, 34]. Nonetheless, we did not find an effect of human cues on the reported working alliance with the CA. This lack might also be due to the fact that TCAs are unable to use non-verbal communication. Building a relationship is an ambiguous process, which is more difficult to establish in a less rich text-based communication environment [75]. Moreover, in studies that did find an improved working alliance with a CA either the interactions with the agent or the intervention itself were longer compared to those in our study [30, 32]. In other studies, although a high working alliance was reported within shorter periods of time, the interactions with the CA followed after introduction by a human healthcare professional [33, 34]. It is therefore unclear whether a TCA is less able to build a relationship with the user, or that it requires a longer time or introduction in a face-to-face

introduction to do so. So even though our findings do support that working alliance is an important mechanism within interventions with TCAs, how to foster a relationship with a TCA still remains a question.

### **Limitations and Suggestions for Future Work**

We aimed to test the effects of visual and relational cues in general, and whether a combination of visual and relational cues would lead to a higher adherence than visual or relational cues alone. However, this led to a combination of various visual and relational cues that were simultaneously manipulated. Future studies could dismantle these, and compare the effects of individual visual and relational cues with another to test which have the biggest influence on intervention adherence.

Secondly, to mimic human behavior, we intentionally chose to apply subtle human cues to our CA. However, participants might not have processed the messages of the agent elaboratively enough to notice these subtle cues, resulting in a lack of effects. We suggest that future studies investigate whether stronger cues are needed in TCAs compared to ECAs to have similar effects. Additionally, these studies could investigate whether longer interactions do lead to an improved working alliance, and thus adherence, for interventions with a TCAs that elicit human cues.

Finally, we did not inform our participants beforehand whether they were interacting with a computer or a human being. Therefore, the expectations of people might have varied, which might have affected our results. Future studies could keep these expectations constant by being transparent about the true nature of the automated agent. Another option would be to manipulate the description of the CA to more closely represent a human being or a computer, and ask participants about their expectations towards support by a human being or a computer, to test how these influence the effects of human cues within automated interventions.

Our study was the first to test the effect of human cues within TCA-supported lifestyle interventions. Future studies could investigate the differences in applying human cues between TCAs and ECAs. Our results suggest that the lack of using non-verbal communication limited the capability to successfully apply relational cues. It would be interesting to test this hypothesis, and how to overcome the lack of non-verbal communication within TCAs.

### **Conclusion**

We found that human cues do not improve adherence to TCA-supported interventions, and that visual cues even lead to lower levels of adherence. This is in contrast to the positive results of

human cues found in studies with ECAs. In line with previous studies, we did find a positive relationship between working alliance with the TCA and adherence. The results suggest that being transparent about the computer-based nature of a CA and thereby setting the right expectations might be key. Besides, we found that factors that work for ECAs, in this case human cues, possibly cannot be carelessly copied to TCAs. This knowledge could help us gaining further knowledge that help us design better automated interventions in the future, which lead to better working alliances, higher levels of adherence, and in turn a healthier lifestyle for us all.

### **Authors' Contributions**

Study design (TRCR, TR); intervention and app design and development (TRCR, PS, TK, AK); data acquisition (TRCR, DRdB, AK); data analysis and interpretation (TRCR, TR, AWME); drafting the manuscript (TRCR, TR, PS, TK, AWME); manuscript revision (TRCR, TR, DRdB, PS, TK, VJ, RAK, DEA, AWME). All authors gave final approval and agree to be accountable for all aspects of the work ensuring integrity and accuracy.

### **Acknowledgements**

This work was supported by The Netherlands Cardiovascular Research Initiative: an initiative with support of the Dutch Heart Foundation, CVON2016-12 BENEFIT, ZonMw (The Netherlands Organization for Health Research and Development) and the members of the BENEFIT consortium. Furthermore, we would like to thank Asena Kinik (AK; Master student Psychology, Leiden University) for her assistance in the practical implementation of the study procedure, and during the development of the intervention.

### **Conflicts of Interest**

None declared.

### **Abbreviations**

CA: conversational agent

ECA: embodied conversational agent

TCA: text-based conversational agent

## REFERENCES

1. Bennett JE, Stevens GA, Mathers CD, et al. NCD Countdown 2030: worldwide trends in non-communicable disease mortality and progress towards Sustainable Development Goal target 3.4. *Lancet*. 2018;392(10152):1072-1088. PMID:30264707
2. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J*. 2016;37(29):2315-2381. PMID:27222591
3. Wilkins E, Wilson L, Wickramasinghe K, et al. European cardiovascular disease statistics 2017. European Heart Network. <http://www.ehnheart.org/cvd-statistics/cvd-statistics-2017.html>. Published February 2017. Accessed September 29 2020.
4. Barak A, Klein B, Proudfoot JG. Defining internet-supported therapeutic interventions. *Ann Behav Med*. 2009;38(1):4-17. PMID:19787305
5. Beishuizen CR, Stephan BC, van Gool WA, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. *J Med Internet Res*. 2016;18(3):e55. PMID:26968879
6. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. *J Med Internet Res*. 2018;20(5):e162. PMID:29728346
7. Brotons C, Björkelund C, Bulc M, et al. EUROPREV network. Prevention and health promotion in clinical practice: the views of general practitioners in Europe. *Prev Med*. 2005;40(5):595-601. PMID:15749144
8. Jallinoja P, Absetz P, Kuronen R, et al. The dilemma of patient responsibility for lifestyle change: perceptions among primary care physicians and nurses. *Scand J Prim Health Care*. 2007;25(4):244-9. PMID:17934984
9. Jansink R, Braspenning J, van der Weijden T, Elwyn G, Grol R. Primary care nurses struggle with lifestyle counseling in diabetes care: a qualitative analysis. *BMC Fam Pract*. 2010;11:41. PMID:20500841
10. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: A systematic review and meta-analysis. *Prev Med*. 2017;100:194-207. PMID:28456513

11. Lau Y, Chee DGH, Chow XP, Cheng LJ, Wong SN. Personalised eHealth interventions in adults with overweight and obesity: A systematic review and meta-analysis of randomised controlled trials. *Prev Med.* 2020;132:106001. PMID:31991155
12. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res.* 2012;14(6):e152. PMID:23151820
13. Kelders SM, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. *J Med Internet Res.* 2011;13(2):e32. PMID:21493191
14. Murray E, White IR, Varagunam M, Godfrey C, Khadjesari Z, McCambridge J. Attrition revisited: adherence and retention in a web-based alcohol trial. *J Med Internet Res.* 2013;15(8):e162. Published 2013 Aug 30. PMID:23996958
15. Wangberg SC, Bergmo TS, Johnsen JA. Adherence in internet-based interventions. *Patient Prefer Adherence.* 2008;2:57-65. PMID:19920945
16. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. *J Med Internet Res.* 2011;13(3):e52. PMID:21821503
17. Etzelmueller A, Vis C, Karyotaki E, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. *J Med Internet Res.* 2020;22(8):e18100. PMID:32865497
18. Karyotaki E, Ebert DD, Donkin L, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. *Clin Psychol Rev.* 2018;63:80-92. PMID:29940401
19. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev.* 2012;32(4):329-342. PMID:22466510
20. Smith PN, Gamble SA, Cort NA, Ward EA, He H, Talbot NL. Attachment and alliance in the treatment of depressed, sexually abused women. *Depress Anxiety.* 2012;29(2):123-130. PMID:22065593
21. Brandt CJ, Clemensen J, Nielsen JB, Søndergaard J. Drivers for successful long-term lifestyle change, the role of e-health: a qualitative interview study. *BMJ Open.* 2018;8(3):e017466. PMID:29530904
22. Bordin, ES. The generalizability of the psychoanalytic concept of the working alliance. *Psychol Psychother.* 1979;16(3):252-260. doi:10.1037/h0085885



23. Horvath, AO, Greenberg, L. Development and validation of the working alliance inventory. *J Couns Psychol.* 1989;36(2):223–233. doi:10.1037/0022-0167.36.2.223
24. Goldberg SB, Davis JM, Hoyt WT. The role of therapeutic alliance in mindfulness interventions: therapeutic alliance in mindfulness training for smokers. *J Clin Psychol.* 2013;69(9):936-950. PMID:23775222
25. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Consult Clin Psychol.* 2000;68(3):438-450. PMID:10883561
26. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy (Chic).* 2018;55(4):316-340. PMID:29792475
27. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. *J Med Internet Res.* 2012;14(4):e110. PMID:22858538
28. Nass CI, Moon Y. Machines and mindlessness: social responses to computers. *J. Soc. Issues.* 2000;56(1):81-103. doi:10.1111/0022-4537.00153
29. Reeves B, Nass, CI. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge university press; 1996. ISBN: 978-1575860534
30. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput.* 2010;22(4):289-298. PMID:20628581
31. Clarke J, Proudfoot J, Whitton A, et al. Therapeutic alliance with a fully automated mobile phone and web-based intervention: secondary analysis of a randomized controlled trial. *JMIR Ment Health.* 2016;3(1):e10. PMID:26917096
32. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. *JMIR Mhealth Uhealth.* 2020;8(4):e15806. PMID:32242820
33. Kowatsch T, Lohse KM, Erb V, et al. Hybrid ubiquitous coaching with a novel combination of mobile and holographic conversational agents targeting adherence to home exercises: four design and evaluation studies. *J Med Internet Res.* 2021;23(2):e23612. PMID:33461957
34. Kowatsch T, Schachner T, Harperink S, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. *J Med Internet Res.* 2021;23(2):e25060. PMID:33484114
35. Cassell J, Bickmore T, Billingham M, et al. Embodiment in conversational interfaces. *Conference Proceedings of the SIGCHI conference on Human Factors in Computing Systems.* May

- 1999;520-527. doi:10.1145/302979.303150
36. Kowatsch T, Nißen M, Rügger D, et al. The impact of interpersonal closeness cues in text-based healthcare chatbots on attachment bond and the desire to continue interacting: an experimental design. *Conference Proceedings of the 26th European Conference on Information Systems*. June 2018; <https://www.alexandria.unisg.ch/254284/>. Accessed October 01, 2020.
37. Aharoni E., Fridlund, AJ. Social reactions toward people vs. computers: how mere labels shape interactions. *Comput Hum Behav*. 2007;23(5):2175-2189. doi:10.1016/j.chb.2006.02.019
38. Appel J, von der Pütten A, Krämer NC, Gratch J. Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Adv Hum Comput Interact*. 2012. doi:10.1155/2012/324694
39. Sah YJ, Peng W. Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. *Comput Hum Behav*. 2015;45:392-401. doi:10.1016/j.chb.2014.12.055
40. Kelders SM, Bohlmeijer ET, Pots WT, van Gemert-Pijnen JE. Comparing human and automated support for depression: fractional factorial randomized controlled trial. *Behav Res Ther*. 2015;72:72-80. PMID:26196078
41. Walther JB, D'addario KP. The impacts of emoticons on message interpretation in computer-mediated communication. *Soc Sci Comput Rev*. 2001;19(3):324-347. doi:10.1177/089443930101900307
42. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns*. 2005;59(1):21-30. PMID:16198215
43. Schulman D, Bickmore T. Persuading users through counseling dialogue with a conversational agent. *Conference Proceedings of the 4th International Conference On Persuasive Technology*. April 2009; 1-8. doi:10.1145/1541948.1541983
44. Lisetti C, Amini R, Yasavur U, Rishe, N. I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Trans Manag Inf Syst*. 2013;4(4):1-28. doi:10.1145/2544103
45. Tudor Car L, Dhinakaran DA, Kyaw BM, et al. Conversational agents in health care: scoping review and conceptual analysis. *J Med Internet Res*. 2020;22(8):e17158. PMID:32763886
46. Thomas S, Reading J, Shephard RJ. Revision of the Physical Activity Readiness Questionnaire (PAR-Q). *Can J Sport Sci*. 1992;17(4):338-345. PMID:1330274

47. Faul F, Erdfelder E, Lang AG, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191. PMID:17695343
48. Kramer J, Künzler F, Mishra V, Pesset B, Kotz D, Smith S, Scholz U, Kowatsch T. Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. *JMIR Res Protoc*. 2019;8(1):e11540. PMID:30702430
49. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med*. 2013;46(1):81-95. PMID:23512568
50. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot*. 1997;12(1):38-48. PMID:10170434
51. Prochaska JO, Velicer WF, Rossi JS, et al. Stages of change and decisional balance for 12 problem behaviors. *Health Psychol*. 1994;13(1):39-46. PMID:8168470
52. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: A novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. *Conference Proceedings of the IEEE 2015 Wireless Telecommunications Symposium*. April 2015;1-6. doi:10.1109/WTS.2015.7117255
53. Kowatsch T, Volland D, Shih I, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform Mobilecoach. In: Maedche A, vom Brocke J, Hevner A, eds. *Designing the Digital Transformation*. DESRIST 2017. Lecture Notes in Computer Science, vol 10243. Berlin, Germany: Springer; 2017:485-489. doi:10.1007/978-3-319-59144-5\_36
54. MobileCoach. URL: [www.mobile-coach.eu](http://www.mobile-coach.eu). Accessed January 11 2021.
55. Stieger M, Flückiger C, Rügger D, Kowatsch T, Roberts BW, Allemand M. Changing personality traits with the help of a digital personality change intervention. *Proc Natl Acad Sci U S A*. 2021;118(8):e2017548118. PMID:33558417
56. Tinschert P, Rassouli F, Barata F, et al. Prevalence of nocturnal cough in asthma and its potential as a marker for asthma control (MAC) in combination with sleep quality: protocol of a smartphone-based, multicentre, longitudinal observational study with two stages. *BMJ Open*. 2019;9(1):e026323. PMID:30617104
57. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity

- Questionnaire Short Form (IPAQ-SF): a systematic review. *Int J Behav Nutr Phys Act.* 2011;8:115. PMID:22018588
58. Craig CL, Marshall AL, Sjöström M, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc.* 2003;35(8):1381-1395. PMID:12900694
59. Hatcher RL, Gillaspay JA. Development and validation of a revised short version of the Working Alliance Inventory. *Psychother Res.* 2006;16(1):12-25. doi:10.1080/10503300500352500
60. Cohen Rodrigues TR, Reijnders T. Virtual coaches in eHealth lifestyle interventions. *Open Science Framework.* osf.io/mgw2s. Published August 18 2020.
61. Luger E, Sellen A. "Like having a really bad pa" the gulf between user expectation and experience of conversational agents. *Conference Proceedings of the 2016 CHI conference on human factors in computing systems.* May 2016;5286-5297. doi:10.1145/2858036.2858288
62. Hendriks F, Ou CX, Amiri AK, Bockting, S. (2020). The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience. *Interaction.* 2020;12:15. doi:10.24251/HICSS.2020.034
63. Shi W, Wang X, Oh YJ, Zhang J, Sahay S, Yu Z. Effects of persuasive dialogues: testing bot identities and inquiry strategies. *Conference Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* April 2020;1-13. doi:10.1145/3313831.3376843
64. Skjuve M., Haugstveit IM, Følstad A, Brandtzaeg PB. Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human-chatbot interaction. *Hum Technol.* 2019;15(1). doi:10.17011/ht/urn.201902201607
65. Mozafari N, Weiger WH, Hammerschmidt M. The chatbot disclosure dilemma: desirable and undesirable effects of disclosing the non-human identity of chatbots. *Conference Proceedings of the 41st International Conference on Information Systems.* 2020; [https://aisel.aisnet.org/icis2020/hci\\_artintel/hci\\_artintel/6/](https://aisel.aisnet.org/icis2020/hci_artintel/hci_artintel/6/). Accessed December 08, 2020.
66. ter Stal S, Tabak M, op den Akker H, Beinema T, Hermens H. Who do you prefer? The effect of age, gender and role on users' first impressions of embodied conversational agents in eHealth. *Int J Hum-Comput Int.* 2020;36(9):881-892. doi:10.1080/10447318.2019.1699744
67. Zhou, S., Bickmore, T., Paasche-Orlow, M., & Jack, B. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. *Conference Proceedings of the International Conference on Intelligent Virtual Agents.* August 2014;528-541. doi:10.1007/978-3-319-09767-1\_63
68. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. *Int. J. Hum. Comput. Stud.* 2020;138:102409. doi:10.1016/j.ijhcs.2020.102409

69. van Vugt HC, Konijn EA, Hoorn JF, Veldhuis J. When too heavy is just fine: creating trustworthy e-health advisors. *Int. J. Hum. Comput. Stud.* 2009;67(7):571-583. doi:10.1016/j.ijhcs.2009.02.005
70. Van Vugt HC, Bailenson JN, Hoorn JF, Konijn EA (2008). Effects of facial similarity on user responses to embodied agents. *ACM Trans. Comput.-Hum. Interact.* 2008;17(2):1-27. doi:10.1145/1746259.1746261
71. Van Vugt HC, Hoorn JF, Konijn EA. Interactive engagement with embodied agents: an empirically validated framework. *Comput Animat Virtual Worlds.* 2009;20(2-3):195-204. doi:10.1002/cav.312
72. Zalake M, Tavassoli F, Griffin L, Krieger J, Lok B. Internet-based tailored virtual human health intervention to promote colorectal cancer screening: design guidelines from two user studies. *Conference Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents.* July 2019;73-80. doi:10.1145/3308532.3329471
73. Loveys K, Sebaratnam G, Sagar M, Broadbent E. The effect of design features on relationship quality with embodied conversational agents: a systematic review. *Int. J. Soc. Robot.* 2020;1-20. doi:10.1007/s12369-020-00680-7
74. Friederichs S, Bolman C, Oenema A, Guyaux J, Lechner L. Motivational interviewing in a Web-based physical activity intervention with an avatar: randomized controlled trial. *J Med Internet Res.* 2014;16(2):e48. PMID:24550153
75. Daft RL, Lengel RH. (1986). Organizational information requirements, media richness and structural design. *Manag. Sci.* 1986;32(5):554-571. doi:10.1287/mnsc.32.5.554
76. Walther JB. (1992). Interpersonal effects in computer-mediated interaction: a relational perspective. *Commun. Res.* 1992;19(1):52-90. doi:10.1177/009365092019001003

## Appendix 1.

**Overview of the 3-week physical activity intervention based on Behavior Change Techniques (BCTs) and Transtheoretical Model of health behavior change (TTM)**

Day	TTM stage	BCTs	Exercise
1	Pre-contemplation	Goal setting	Formulate general goal (what do I want to achieve and why? )
2		Information about health consequences	Quiz about behavior and health consequences
3	Contemplation	Pros and cons	Decisional balance worksheet
4	Preparation	Goal setting	Formulate SMART goal
5		Valued self-identity	Self-affirmation exercise
6		Prompts/Cues; Action planning	Formulate 'If-then plan'
7		Problem solving	Identify barriers and coping strategies
8		Self-monitoring	Implement short bursts of activities, and compare step-count to yesterday's
9		Action planning; Social support; Barrier identification	Plan physical activity challenge with other person; identify barriers and coping solutions
10	Action	Instructions on how to perform health behavior	Quiz about performing physical activity
11		Review behavioral goal(s)	Reflect on goals (day 4) and make adjustments (SMART)
12		Identification of self as role model	Identify own role model, and for whom you are a role model
13		Demonstration of the behavior; Social comparison; Credible source	Watch video of Usain Bolt interview
14		Review outcome goal(s)	Reflect on PA challenge (day 9)
15		Review behavioral goal(s)	Reflect on goals (day 11) and make adjustments (SMART)
16		Focus on past success	Reflect on rewarding experience of previous physical activities
17		Reduce negative	Stress management and emotional

		emotions; Monitoring of coping emotional consequences	
18		Self-talk	Positive labelling of upsetting experiences
19	Maintenance	Review outcome goal(s)	Reflect on barriers and coping strategies (day 7)
20		Incompatible beliefs; Discrepancy between current behavior and goal	Imagine future self and set goals to work towards that
21		Monitoring of emotional consequences; Review outcome goal(s)	Meta-reflection of intervention (what did I learn, what did I like the most, how did I change?)
1-21	All stages	Social reward; Feedback on behavior	Praise for effort and progress; inform participant about daily step counts

Appendix 2.

### **Analyses with effectiveness as outcome variable**

Effectiveness was measured through objective step count data retrieved from Apple Health or Google Fit (depending on the smartphone of the participant). We calculated the mean difference between the average baseline step count (measured in the week before the intervention) and the average step count in the final week of the intervention. Participants were included in the analyses if both a valid baseline step count and a minimum of 5 days of step count in the final week were registered.

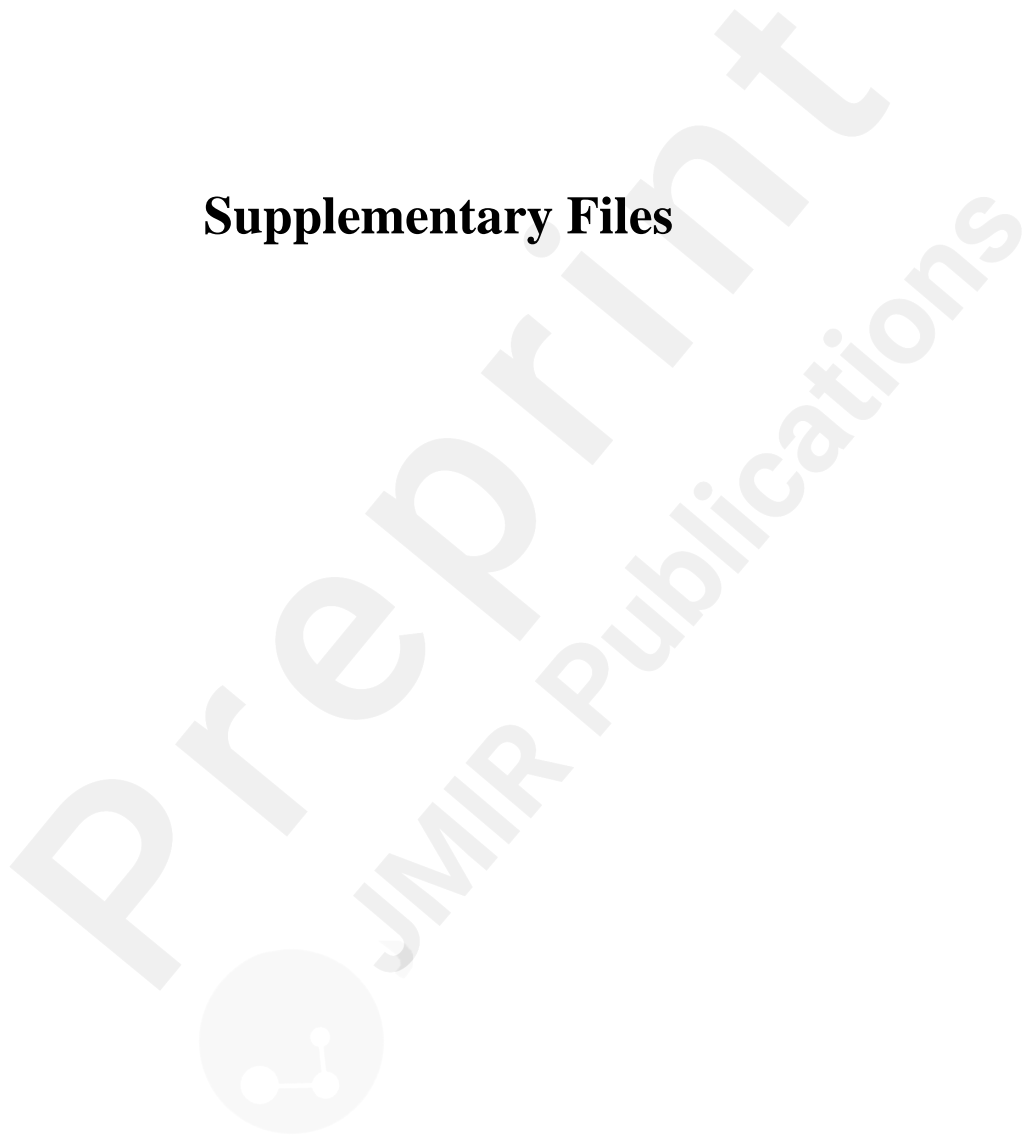
We conducted an one-way between subjects ANOVA to compare intervention effectiveness between the four different conditions. There was no significant difference in effectiveness between the conditions,  $F(3, 43) = .726$ ,  $P = .54$  (see Table 3 for mean and SD per group). Also the post-hoc tests comparing the three human cues conditions with the no human cues condition ( $t(39) = -1.021$ ,  $P = .31$ ), and the test comparing the condition with both visual and relational cues with the visual cues only and relational cues only groups ( $t(39) = .171$ ,  $P = .87$ ) showed no significant differences.

**Table 3.** Means and standard deviations per group of effectiveness (mean difference between average baseline step count and average step count in final week of intervention).

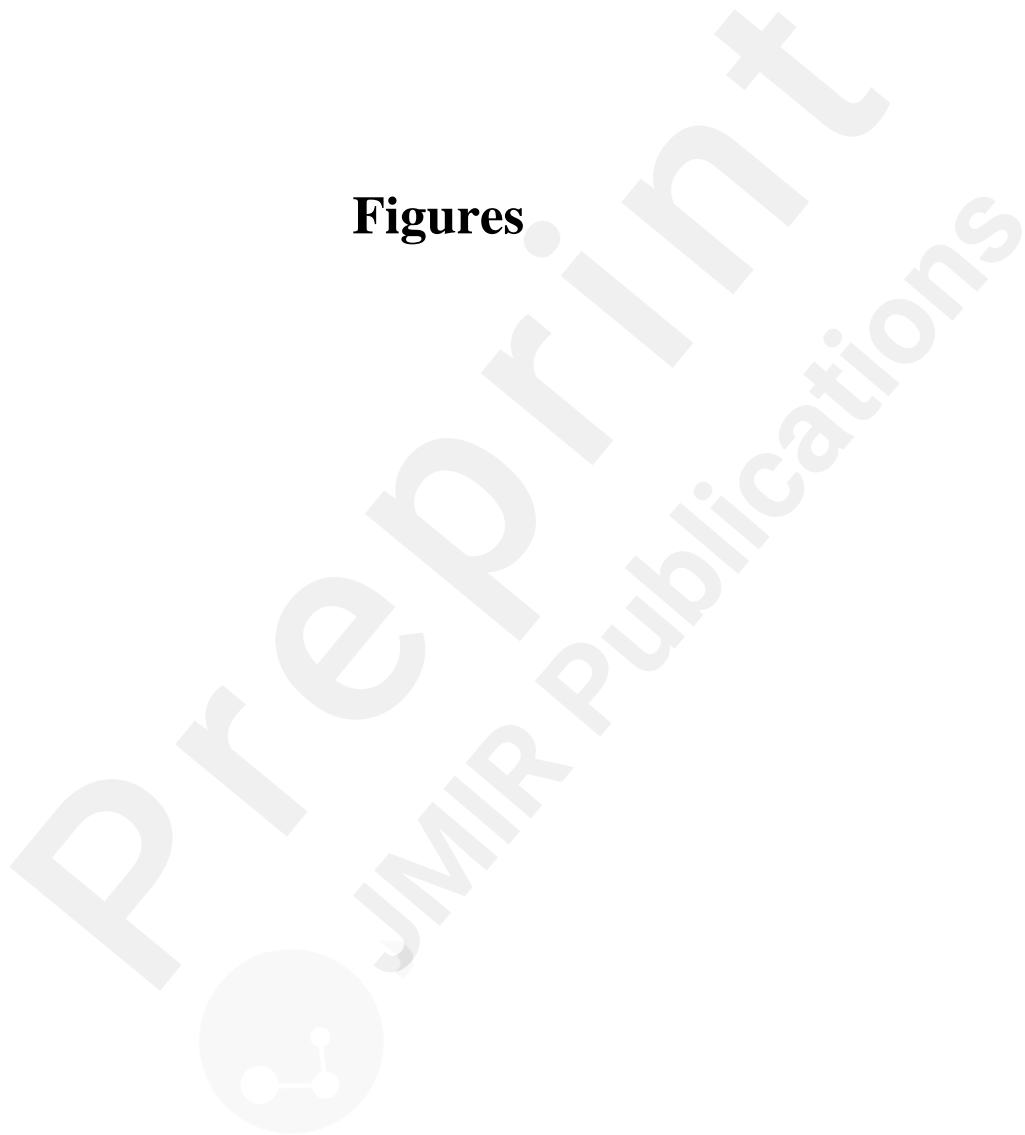
Variable	Effectiveness	
	N	Mean (SD)
<b>Visual &amp; relational cues</b>	11	1351.58 (1193.88)
<b>Visual cues</b>	6	2062.36 (1653.04)
<b>Relational cues</b>	11	898.40 (1933.30)
<b>No cues</b>	15	805.67 (2337.57)



## Supplementary Files




## Figures



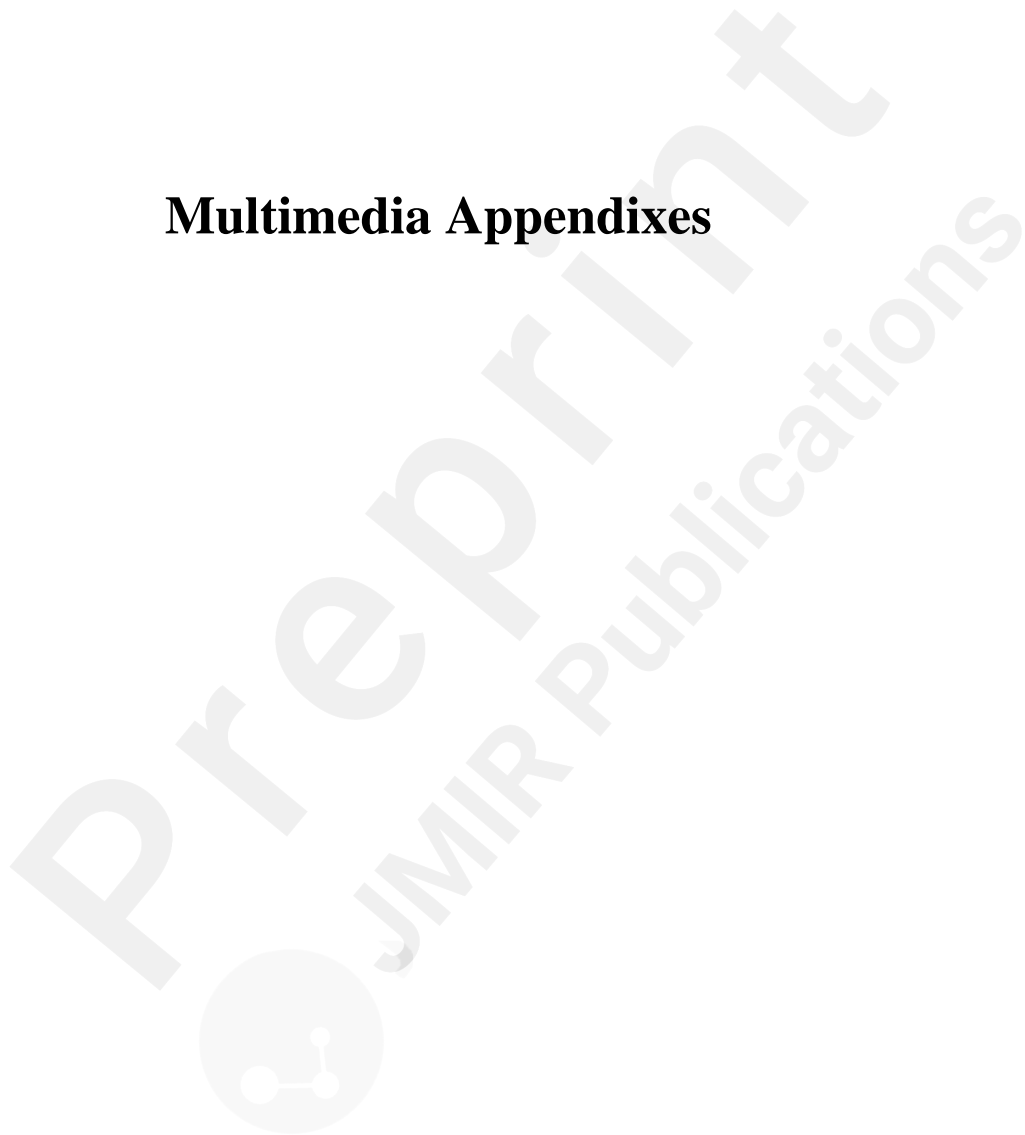
Screenshots of Benefit StepCoach app.



Example of conversational turns per condition.

Visual & relational cues	Visual cues	Relational cues	No cues
 <p>Hi (name)! I really appreciate all the work you're doing, so keep it up 😊</p>	 <p>Hi (name)! 😊</p>	 <p>Hi (name)! All the work you're doing is impressive, so keep it up.</p>	 <p>Hi (name)!</p>
<p>I can imagine that it is quite hard to work on your goal alone. We've already discussed many ways to achieve your step goal yourself, but in my experience social support can be very helpful as well 😊</p>	<p>We've already discussed many ways to achieve your step goal yourself, but I think social support can be very helpful as well 😊</p>	<p>It can be quite hard to work on your goal alone. You've already worked out many ways to achieve your step goal yourself, but social support can be very helpful as well.</p>	<p>You've already worked out many ways to achieve your step goal yourself, but social support can be very helpful as well.</p>
<p>Do you have a friend or family member that would be willing to help you? Could you tell me his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could you tell me his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could share his or her name?</p>	<p>Do you have a friend or family member that would be willing to help you? Could share his or her name?</p>

## Multimedia Appendixes



Overview of the 3-week physical activity intervention based on Behavior Change Techniques (BCTs) and Transtheoretical Model of health behavior change (TTM).

URL: <http://asset.jmir.pub/assets/56f19370f836d60da2866ef5e3d58762.docx>

Analyses with effectiveness as outcome variable.

URL: <http://asset.jmir.pub/assets/e22e4c95bc657373adb2227507cec718.docx>

