Friends or Funds: Comparing Accountability Partners and Deposit Contracts for Increasing Step Counts

Lefeure, Diana

Psychologie
Faculteit der Sociale Wetenschappen

Friends or Funds: Comparing Accountability Partners and Deposit Contracts for Increasing Step Counts

(Diana Lefeure)

Master Thesis Health and Medical Psychology

Faculty of Behavioural and Social Sciences – Leiden University

(March 2025)

First Examiner: (David de Buisonjé, D.R.), Health, Medical and Neuropsychology Unit; Leiden University

Second Examiner: (Margot P. van der Doef, PhD), Health, Medical and Neuropsychology Unit; Leiden University

Abstract

This retrospective cohort study analized the effectiveness of financial versus social incentives for increasing average daily step counts. Data involving 7,675 WayBetter app users participating in two-week challenges were examined. Participants selected either a financial incentive by paying $10 in a deposit contract or a social incentive by inviting an accountability partner. Step counts were tracked by gyroscopic sensors or fitness wearables (e.g., FitBit). The users were instructed to achieve step count goals to win the game. Our analysis compared which of the two conditions (independent variable) was more effective in increasing step counts (dependent variable) and success outcomes (dependent variable). In IBM SPSS 27, a paired samples t-test showed that the overall daily step count increased significantly during the challenge compared to the average recorded before the challenge (baseline). Furthermore, an independent samples t-test showed significant differences between the two groups, where deposit contracts showed significant step count increases of 8%, while accountability partners showed significant step count decreases of -16%. A cross-tab for an independent chi-square test revealed significant differences in success outcomes between the two groups, with a success rate of 70% in the deposit contract group and only 13% in the accountability partner group. Exploratory analysis showed that winners increased their steps by 16%, while losers decreased them by 13%. These correlational findings highlight the importance of careful implementation of incentives in digital health interventions. Further research on improved success rates and variations of desing of the accountability partner condition must be explored.

Layman's Abstract

In this study we investigated whether a social or a financial incentive is more effective to motivate individuals to increase their daily step counts. We received the data recorded from 7,675 users who tracked their daily step count progress with the WayBetter app. Participants were able to choose their incentive. To join with a financial incentive in the form of a deposit contract they were requested to pay a $10 deposit. Participants were at risk of losing their deposit but they were assured of getting a refund if they won the challenge. As an alternative with no financial risk, participants could chose to join with a social incentive in the form of an accountability partner. They were requested to send a link to a friend to invite them to encourage them throughout the challenge as their accountability partner. Compared to the step counts before

the challenge, we found that the majority of people in the deposit contract group won challenges and increased their steps significantly, while the majority in the accountability group lost the challenges and significantly decreased their step counts. Thus, this study shows that financial incentives in the form of deposit contracts have potential to increase motivation for improved physical activity, but accountability partners might not have been effective in increasing motivation. Therefore, future research should investigate ways to implement social incentives more effectively.

Friends or Funds: Comparing Accountability Partners and Deposit Contracts for Increasing Step Counts

A pedometer called "*manpo-kei",* which translates from Japanese as "10,000-step meter" was designed in Japan around the time of the 1964 Tokyo Olympics which is known to have started the concept of walking 10,000 steps a day (Tudor-Locke et. al., 2002). Step counting was initially introduced to combat sedentarism and encourage more walking (Tudor-Locke et. al., 2002). However, it has since evolved into a widely accessible method of quantifying physical activity (PA), promoting health (Pedersen & Saltin, 2015) and reducing chronic disease risk (Banach et al., 2023; Saint-Maurice et al., 2020; World Health Organization: WHO, 2018, 2020). Studies show that walking reduces mortality (Paluch et al., 2022), cancer mortality (Del Pozo Cruz et. al., 2022; WHO, 2018), and dementia (Del Pozo Cruz et. al., 2022b). Moreover, it improves mental health by reducing depression and anxiety, regulating mood and increasing cognitive functions (Hsu et al., 2021; Mammen & Faulkner, 2013; WHO, 2020). However, despite this knowledge, around a third of the worldwide population is physically inactive (Guthold et al., 2018) or fails to meet the minimum recommended PA levels (WHO, 2020). This raises the question of how to increase people's engagement in PA.

Health-promoting interventions (Mitchie et al., 2011), often delivered via e-health platforms such as mobile apps, use diverse features to keep their users engage in a new behaviour (Dugas et al., 2020). Mobile interventions are location independent and accessible to a large audience. They are specifically well suited when they introduce activities such as walking, since smartphone integrated sensors can easily record steps (Fuller et al., 2020). Mobile interventions often involve gamification (Cugelman, 2013). Gamification refers to the use of game-like elements in non-game context such as e-health or education. Gamification uses rewards like points for completing tasks, rankings to encourage competitive behaviour, badges to visually represent accomplishments. This type of features makes mobile interventions appealing and help sustain participation in new behaviours. On the other hand, the downside of mobile interventions is the lack of personal face-to-face feeling which leads to low adherence (Hurmuz-Bodde et al., 2024) and people drop-out before the benefits of PA are detectable (Collins et al., 2022). Therefore, the implementation of the strategies must be well-addressed.

Strategies to increase participation and achieve behavioural change (e.g. increase PA), can include financial or social incentives. Financial incentives involve monetary rewards such as

bonuses, profit-sharing, or deposit contracts for achieving predefined goals (e.g., promoting health-related behaviours) (Kurti et al., 2016; Mantzari et al., 2015; Pirotta et al., 2019). On the other hand, social incentives offer rewards derived from social interactions, which help motivate individuals by fulfilling their need for acceptance and recognition. They motivate individuals through public statements, peer comparisons or accountability partners (Fehr & Falk, 2002). This study examines how social and financial incentives are used in a smartphone app, WayBetter, which was designed to help users encourage health behaviours such as to increase step counts. As financial incentives WayBetter uses deposit contracts. Deposit contracts require people to contribute a predetermined sum of their own money up front, which they can only get back if they fulfil their predefined goals (Boonmanaunt et al., 2022) and are effective in increasing PA (Budworth et al., 2019; de Buisonjé et al., 2022; Donlin Washington et al., 2016; Patel et al., 2016).

Deposit contracts as financial incentives are effectively using the present bias. Present bias influences people to engage in behaviours that lead to immediate payoffs (Laibson, 1997), thus, people will prioritize short term rewards to long term benefits (Laibson, 1997; Wang & Sloan, 2018). In the context of PA, people tend to avoid the short-term discomfort of PA disregarding the long-term benefits (Hunter et al., 2018). However, when using deposit contracts, interventions are refocusing the attention of an individual, making them focus on new short-term discomfort which in the case of deposit contracts would be losing their own money. Thus, the immediate payoff is in this case getting the deposit back. Additionally, the phenomenon of loss aversion (Kahneman and Tversky, 1979) may make deposit contracts the most effective form of financial rewards (Boonmanaunt et al., 2022).

On the other hand, deposit contracts also have some disadvantages. Their working mechanism is through loss aversion, which leads to a low uptake, especially for the individuals who have low self-efficacy and fear the risk of losing their deposit (de Buisonjé et al., 2023; Kerrigan et al., 2019). For example, in a study using deposit contracts for gym attendance in the workplace, only 12% of workers joined the challenge (Royer et al., 2015). Financial risk, which can be stress or anxiety inducing and is not beneficial to behavioural change (Tamir & Gutentag, 2017). Furthermore, they may serve as a financial barrier to people who cannot contribute the required sum, which is problematic, as the objective of scientific research is to reduce socioeconomic position (SEP) related health disparities (Lunze & Paasche-Orlow, 2013).

To overcome some of the drawbacks of deposit contracts, WayBetter uses social incentives as an alternative (*WayBetter*, n.d.). Social incentives coupled with gamification elements have demonstrated significant step count increases compared to control groups in studies using social elements such as competition, support and collaboration (Harrison et al., 2019; Patel et al., 2017; Patel et al., 2019). Furthermore, public praise (e.g., on social media) for achieving health objectives may improve mental well-being and provide psychological resources needed to reach goals (Cohen et al., 2000). Social interactions foster a sense of accountability and belonging, motivating people to stay committed to their goals (Feeney & Collins, 2015; Rusbult et al., 2009). In online communities, social engagement was associated with increased success rates as it encourages external feedback and provides opportunities for comparison with others, which helps individuals become more self-aware, confident and motivated (Song & Xu, 2023).

WayBetter specifically uses accountability partners as a form of social incentive (*WayBetter*, n.d.). An accountability partner is someone who regularly checks in with the participant, offering support, giving constructive criticism, and expecting updates or progress reports (Oussedik et al., 2017). The key to how accountability works is that it creates a sense of responsibility, which increases while an individual anticipates the next follow-up. Knowing that they will be held accountable for their actions, participants get motivated to follow through with the predefined goals (Oussedik et al., 2017). Additionally, feedback from an accountability partner strengthens one's self-regulatory processes such as self-monitoring and helps one adjust their behaviour in order to achieve the desired state (Oussedik et al., 2017).

Furthermore, accountability partners are linked to public commitments, whether face-to-face or online (Bailey, 2009), where participants openly share their intentions to change behaviours, thereby creating social accountability (Munson et al., 2015). This creates a sense of pressure that motivates individuals to follow through on their commitments to protect their reputation and avoid failure in front of others (Munson et al., 2015). However, long-lasting effective accountability focuses on encouragement and support, which in turn creates autonomous motivation. The Self-Determination Theory (SDT) (Deci & Ryan, 1985) suggests that people who are autonomously motivated are more likely to stay long-term committed to their goals because they feel in control and personally invested in their actions (Feeney & Collins, 2015). Thus, accountability partners who provide support encourage participants to

achieve their goals by fostering their drive to succeed. Additionally, accountability partners are linked to the effect of social facilitation, which suggests that people perform better on tasks when they are observed by others (Bond & Titus, 1983). Finally, interventions based on accountability partners combine the social pressure of the public commitment with the individualized social support which in turn creates a powerful system that encourages people to stick to their goals (Fortuna et al., 2019; Munson et al., 2015; Santarossa et al., 2018; Song & Xu, 2023).

On the other hand, main drawbacks of accountability partners are that they rely on a third person who serves as an incentive by encouraging progress and goal achievement (Oussedik et al., 2017). The presence of a third person implies unpredictability from a lower level of control because interhuman relationships are dynamic and involve communication and feelings which are subjective (Berscheid et al., 1989).

Accountability Partners Versus Deposit Contracts

A comparison of the two incentive strategies, accountability partner versus deposit contract, allows for a more tailored intervention design that considers incentives' adherence potentials and suitability to a target population. From a practical perspective, deposit contracts use a straightforward, measurable incentive: avoid financial loss by adhering to a goal (Boonmanaunt et al., 2022), but they require monetary risks which may act as a financial barrier, limiting participation of some populations (e.g., low SEP). In contrast, accountability partner interventions are more accessible, they do not require money, but rely on consistent social engagement (Oussedik et al., 2017), which makes them less structured as they depend on personal interactions, that can be unpredictable.

Furthermore, deposit contracts leverage present bias and loss aversion that shift participants' focus toward achieving their goal to avoid the loss of their deposit. In contrast, participants with an accountability partner rely on encouragement, responsibility and social engagement, which can foster autonomous motivation for behavioural change, however, its effectivity may depend on the partnership's dynamics and therefore the effects can vary (Berscheid et al., 1989).

In conclusion, deposit contracts may be suitable for individuals motivated by financial consequences and when a high-impact incentive is needed to achieve a strong behavioural

change. In contrast accountability partners may be more suitable for populations with limited resources and with a tendency to avoid risk.

This Study

Direct comparisons between accountability partner and deposit contract interventions are limited, yet evaluating them is crucial as digital health interventions often rely on social or financial incentives that differ significantly in how they motivate behaviour. This study provides a new perspective by directly comparing these strategies in the context of a two-week gamified challenge from WayBetter designed to increase step counts in individuals intending to improve their levels of moderate PA. This study expands understanding of motivation and behavioural economics and offers insights for designing interventions that encourage sustained physical activity. To explore this, three Research Questions (RQ) will help examine the impact of the two incentives on step counts. Answering these questions can benefit apps like WayBetter in understanding what motivates their users more effectively and provide valuable insights into improving incentive strategies. Specifically, due to the differences between deposit contracts and accountability partners, understanding their different mechanisms and effects can guide intervention designers in implementing one strategy to overcome the limitation of the other.

First, to investigate differences in behaviour, we differentiate between average daily step count during the two-week challenge and the average step counts recorded prior to starting the challenge, which we refer to as baseline. We will analyse the effect of the mobile intervention on the average daily step count and examine: How do step count change during the two-week challenge compared to baseline (RQ1)? Based on studies on behaviour change with deposit contracts and accountability partners (e.g., de Buisonje et al., 2022; Oussedik et al., 2017), we expect an overall increase in the average step counts (H1).

Second, we investigate: How do the two conditions (deposit contracts vs. accountability partners) differ in step counts increases during the challenge (RQ2) compared to baseline? A study on adolescents' adherence to health programs which compares social versus financial rewards shows that financial rewards are more effective incentives (Van Dooren et al., 2018) and argues that money can be used as value outside of the intervention itself (Van Dooren et al., 2018) making financial incentives more attractive. Furthermore, due to the strong effect of loss aversion (Kahneman & Tversky, 1979) deposit contracts are stronger than any other financial

incentive (Boonmanaunt et al., 2022), which leads us to expect participants in deposit contract condition to achieve higher step counts compared to those in the accountability partner condition (H2).

Finally, we inspect: Which group (deposit contract vs. accountability partner) has a higher success rate (RQ3)? Similarly to the previous hypothesis, this hypothesis is based on literature showing that deposit contracts are highest financial motivators (Boonmanaunt et al., 2022) which overall demonstrated to be stronger than social motivators (Van Dooren et al., 2018), we expect that the success rates are higher in the deposit contract group than in the accountability partner group (H3).

Methods

This study is a retrospective cohort study, of daily step counts data collected over the duration of two-week challenges on the WayBetter app, which includes participants in two different incentive groups which serve in this study as independent variables (IV). One group used accountability partners, while the other group used deposit contracts. As dependent variables (DV) we analyse step count increases (continuous) compared to baseline and success ratios (binary). Since this study did not include any randomized variables and was purely observational, no causality can be inferred, and conclusions will be based solely on associations.

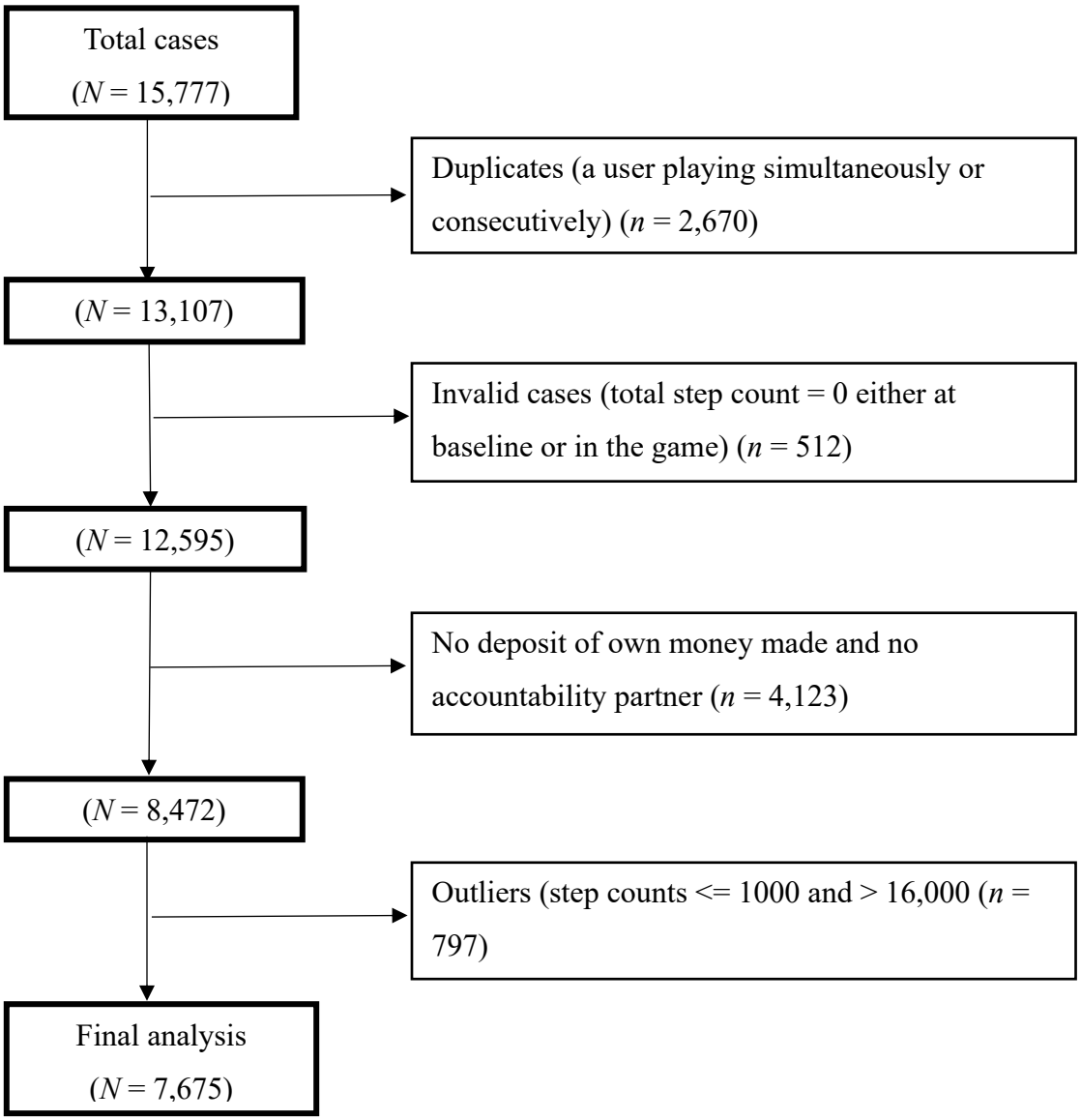
Participants

A total of 7,675 unique participants were analysed. The original file received from WayBetter contained 15,777 cases of users participating at step count challenges between 2023-04-01 and 2023-09-04. A flowchart displayed in *Figure 1* describes the data cleaning process. A total of 2,670 simultaneously or consecutively played challenges were flagged as duplicates and were removed. Another 512 cases were removed as invalid for representing zero step counts at baseline or during the game. Our data included also cases pertaining to a third incentive condition that was not analysed in this study and therefore a total of 4,123 cases were removed for not meeting the criteria described by the two conditions that are compared in this study. Finally, 797 outliers were identified using the interquartile range (IQR) method. Tukey (1977) introduced this concept as a robust statistical technique, resistant to extreme values (see *Appendix 2* for more information about the calculation of the outliers). This method is well suited

for large sample sizes, it is not sensitive to unequal group sizes, and it aligns with the focus of our sample, which is the general public. Thus, by flagging extreme values, we ensured that the focus remains on behavioural patterns in the general population. Based on IQR cutting rule we removed participants with extreme values below 1,000 (n = 542) and above 16,000 steps (n = 255) (see *Appendix 2* for a detailed calculation of outliers and *Appendix 3* for a sensitivity check report where these outliers are included).

Figure 1

Flowchart of Data Cleaning Process



Procedure

The data was collected by the WayBetter app and was provided to researchers from Leiden University. The participants were users of the WayBetter app who joined the challenges on their own initiative without researcher's interventions, which makes the data collection a convenience sampling. To participate in a challenge a user had to download the WayBetter app and allow it to record their step count information. To access the games, users needed a membership priced at $69 for six months. However, the app was also available as a free trial for the first week. With the free trial, users were able to participate in and finish the challenge they started during the free trial week and collect their winnings, even if they did not buy a membership. To track their steps in a challenge, participants had to own a smartphone and could use an additional wearable device (e.g. Fitbit, Garmin or Apple Health). The users were required to choose a group, either paying a $10 deposit, or inviting an accountability partner instead. To win a challenge, users were required to complete step count activities (see *Waybetter app* for details on step count goals) during a pre-defined two-week period. In most cases, the first week of the two-week challenge was a warm-up week, which means that to win, participants only had to complete the activities in the second week, because the first one counted as warming-up week. Failure to reach the required step count activities over the second week, resulted in losing the challenge. Participants in the deposit contract group lost their deposit if they failed the challenge. On the other hand, participants who won the challenge, received their deposit back and a small bonus from the money pot. The pot refers to the total amount of money deposited by all users for a specific game, that was split between all winners at the end of a game. Participants in the accountability partner group did not lose or win any money regardless of if they won or lost the challenge. Additionally, during the two weeks, the users received personalized notifications on their smartphones when they completed their daily step count goals or to remind them to rescue their game when they did not complete them. Finally, participants could join games as early as two weeks before the start date or shortly after the challenge began. If they changed their mind, they could quit a game within three days after a game started and receive a refund for their paid deposit.

For the analysis we did not obtain informed consent before the start of the study since we used anonymous research data collected by WayBetter. The company informs its users about the possibility of academic research on anonymized data in their privacy policy statement and did

not require an additional informed consent. The study protocol was preregistered on Open Science Framework: doi:10.17605/OSF.IO/D237C. This study was approved by the Psychology Research Ethics Committee of Leiden University (2024-08-02).

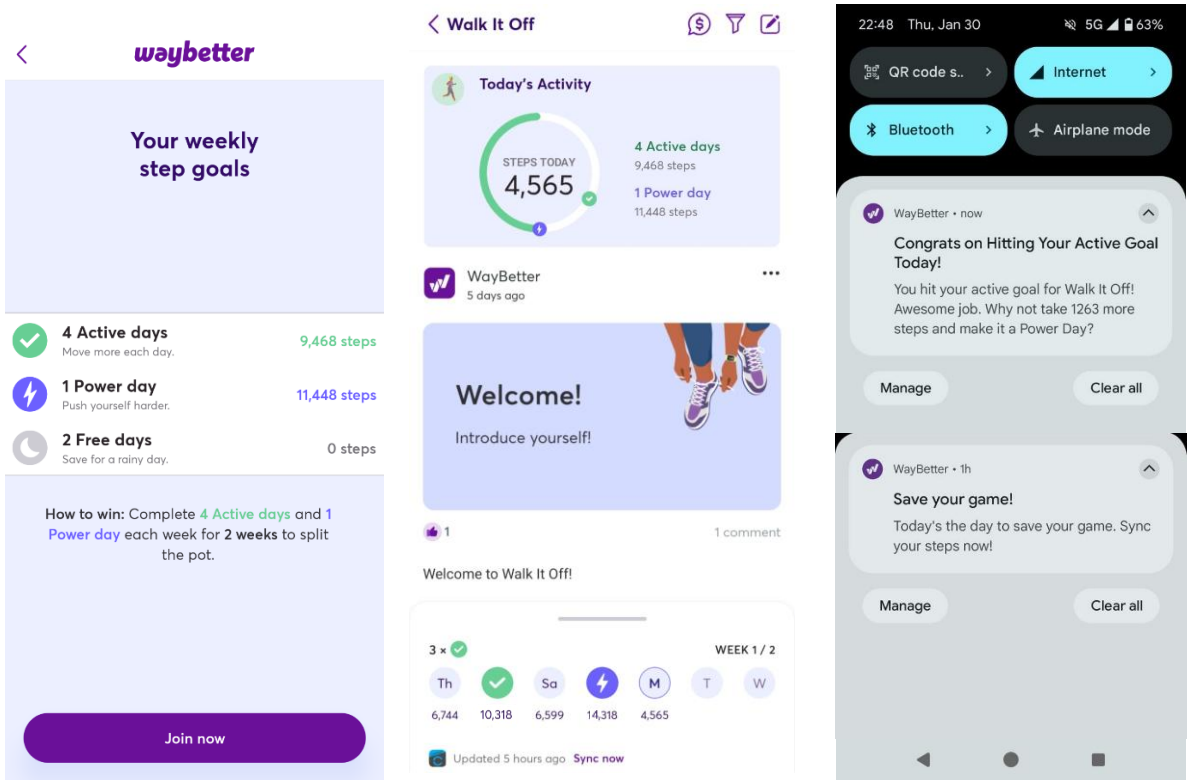
Materials

WayBetter app

After downloading the app participants were able join a challenge. Then, the app synchronized the step counts from the tracker, calculated and displayed the individual weekly step goals (see *Figure 2* for screenshots) and the requirements to win the game: "*Complete 4 Active days and 1 Power Day each week for 2 weeks to split the pot.*". For determining the goals, WayBetter retrieved the historical step count data from the past seven days and calculated the baseline. Further, the algorithm determined the step count goals using the baseline. Goals were split into step count activities: a week (seven days) was divided in four "*Active days*" (110% of baseline), one "*Power day*" (130% of baseline) and two "*Free days*" (no step counts required) (see *Figure 2*). Once the challenge officially started, the users were able to follow their progress in the app. A weeklong calendar showed them their step counts and checkmarks were placed to suggest if active days or power days were reached (see *Figure 2*).

Figure 2

Screenshots Waybetter App and Smartphone Notifications



a b c

[a] Visualisation of weekly goals split into daily activities of active and power and free days

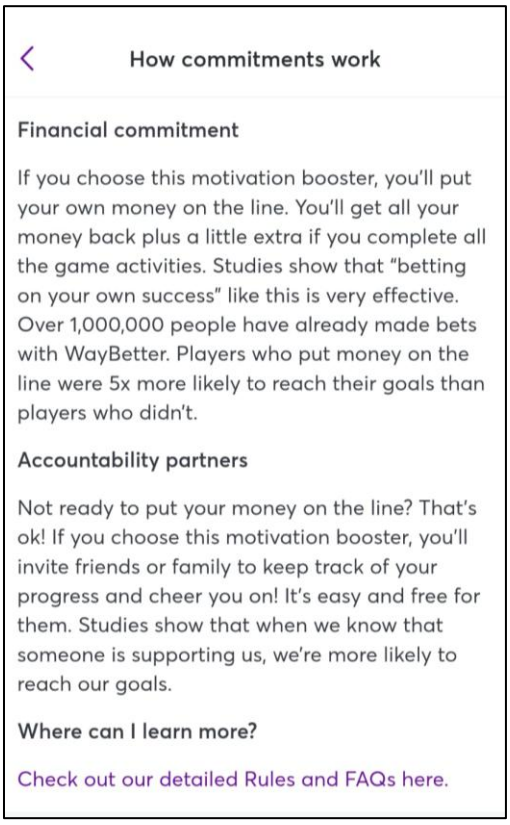
[b] Visualisation of daily and weekly progress

[c] Daily smartphone notifications when reaching a daily goal (upper notification) and when daily goal still must be attained (lower notification)

To choose the condition (incentive) the app presented the user with the question: *"How do you want to stay motivated?"*. Participants could select one of two possibilities: either by putting down a $10 deposit ("Make a financial commitment") or by sharing the link to their challenge with a person of their choice, inviting them to be their accountability partner ("Invite an accountability partner"). For more information about the different conditions, the user could click on a link ("*Learn more*") where they could read how the two incentives work (see *Figure 3*). Participants who selected the financial commitment option were once again reminded that they will receive their money back if they complete all game activities ("*Remember, you will get*

Figure 3

WayBetter App Screenshot: Description of the Two Incentives



it all back if you complete all the game activities.") and were requested to paying the deposit through PayPal or Credit card. The participants who selected to join with an accountability partner, shared the link as an E-mail or a message (e.g. SMS, WhatsApp) saying: *"Hey! Will you be my accountability partner in my WayBetter game".* The message was able to be modified by the sender. Once the motivator (deposit contract or accountability partner) was chosen, and either a payment was made or a link to a partner was sent, the user had officially joined the game, and their name was listed with other users who joined the same game.

Measures

The primary outcomes were step count increases (continuous DV) and challenge success (binary DV) (win vs. lose). The step counts were determined with gyroscopic sensors in smartphones or wearable devices such as, Fitbit, Garmin or Apple Health, and translated into daily step counts (continuous DV). First, the step count increase was calculated by subtracting

the baseline (average daily step count from the seven days prior to joining the game) from the average of daily steps counted during the challenge. An increase in steps was confirmed if the average of step count increase during the challenge was greater than 0. Second, the challenge success (binary DV) was determined automatically by applying individual weekly goal rules with specified number of daily step goals (4 active days and 1 power day). A challenge was won if all activities were achieved and failed when one or more activities were missed.

Primary analysis focused on comparison of the challenge outcomes between two groups, (binary) independent variables (IV): deposit contract group and accountability partner group. Secondary analysis focused on additional binary IV: membership type (free trial vs. paid membership).

Analysis

Due to the large sample size (N = 7,675) of this study, even small effects may reach statistical significance (Cumming, 2014) without ensuring practical relevance (Lin et al., 2013). Therefore, we defined a threshold for practical significance at a change in steps of 1,000 steps (see *Appendix 5* for the rationale). Furthermore, the effect sizes are expressed in their original measurement units and confident intervals (Cumming, 2014). Data analysis was done using IBM SPSS 27, and the standard $p < .05$ criterium for determining statistical significance.

Hypothesis 1 (H1): The step counts of both groups combined increase during the challenge compared to baseline.

To determine if the step counts increased (DV) during the challenge compared to baseline (IV) a paired samples t-test was computed. The effects were interpreted with effect size Cohens d $\geq .2$, $\geq .5$, and $\geq .8$ as small, moderate, and large, respectively (Cohen, 1988). Furthermore, the assumptions for the t-test, including the assumption of normality and equal variances were checked. In case of violation of assumptions, a Wilcoxen Signed-Ranks test was additionally planned.

Hypothesis 2 (H2): People in deposit contracts condition have higher step count increases compared to people in the accountability partner condition.

To determine the effects observed in the type of the incentive (deposit contract vs. accountability partner) (IV) on the step count increases (DV) a two-tailed independent samples t-

test was performed. The effects were interpreted with effect size Cohens d ≥ .2, ≥ .5, and ≥ .8 as small, moderate, and large, respectively (Cohen, 1988). Just as before, the assumptions for the t-test, including the assumption of normality and equal variances were checked. In case of violation of assumptions, a non-parametric Mann-Whitney-U test was additionally planned.

Hypothesis 3 (H3): Challenge success is higher in the deposit contract condition than in the accountability partner condition.

To investigate if challenge success (DV) differs between the two incentive types (deposit contract vs accountability partner) (IV) a crosstabs test for a chi-square test of independence was performed. The effects were interpreted with effect size Phi (φ) (df = 2) ≥.10, ≥.30, ≥.50 as small, moderate and large, respectively (Cohen, 1988). Furthermore, the assumptions of a chi-square test of independence which includes the independence, random samples and mutual exclusivity were checked. In case of violation of assumptions, a Mann-Whitney U Test was planned to confirm the results.

Exploratory analysis

In additional analyses, we explored with an independent samples t-test, whether step count increases are different between participants who won versus the participants who lost a challenge. Another exploratory analysis was done to check if free trial memberships was associated with step count increases or with the outcome of the challenges. Therefore, two crosstabs test for a chi-square test of independence, a paired sample t-test and an independent sample t-test were performed to test H1, H2 and H3 while differentiating between the two membership types (free trial vs. paid membership) as binary IV.

Results

A total of 7,675 participants were included in the analysis after exclusion of outliers. The data set had no specific information about the demographics, but based on previous research (De Buisonjé et al., 2023) we estimated that the sample consisted of 87% females, a mean age of 38 years old and 93% North American, and the remaining mostly European (4.6%). Most of the participants (95%) took part at two-week challenges with a warm-up week. The average daily step count before the game (baseline step count) was 6,632 (*SD* = 3,026) which increased by 7,030 (*SD* = 3,154) or by 6% during the challenges. Most of the participants (60%) had a paid

WayBetter membership while the rest (40%) were in a free trial. A portion of 74% of participants in the accountability partner group were free trial members, while 36% of participants in the deposit contract group were free trial members. See *Table 1* for more descriptive statistics.

Table 1

Descriptive Results (N = 7,675)

	Baseline Steps [a]	**Game Steps [b]**	**Step Change [c]**	**Relative Change (%) [d]**	**Win ratio (%) [e]**	**Number of participants (N)**
All Players	6,632 (3,026)	7,030 (3,154)	398 (2.075)	+6.00%	61.8%	7,675
Outcome						
Winners	7,066 (3,057)	8,145 (2,751)	1,109 (1,632)	+15.69%	100%	4,742
Losers	5,932 (2,838)	5,180 (2,873)	-752 (2,196)	-12.68%	0%	2,933
Condition						
Deposit Contract	6,780 (3,025)	7,305 (3,066)	525 (2,013)	+7.73%	66.9%	6,955
Accountability Partner	5,209 (2,638)	4,378 (2,732)	-832 (2,258)	-15.97%	12.8%	720
Membership type						
Free Trial	5,639 (2,811)	5,890 (3,068)	251 (2,507)	+4.45%	46.3%	3,046
Paid Membership	7,288 (2,984)	7,782 (2,980)	494 (1,726)	+6.78%	72%	4,626

Note. Data are frequencies, means (SD) and percentages.

[a] Baseline Steps = Daily steps before starting the challenge.

[b] Game Steps = Average daily steps during challenge.

[c] Steps Change = Relative change in steps.

[d] Relative Change % = Change steps/baseline steps.

[e] Win ratio % = Percentage of winners within group

Hypothesis Testing H1: The step counts of both groups combined increase during the challenge compared to baseline.

No major violation against the assumption of normality was detected (see A*ppendix 4* for detailed assumptions check). In line with the hypothesis, the overall step count of both groups combined significantly increased during the challenge, $t(7,674) = 16.790$, $p < .001$, $d = .192$, 95% CI [.169; .214]. See *Table 1* for an overview of the descriptive results. The daily average step counts during a challenge (M = 7,030 steps, SD = 3,026) increased by 398 steps (SD = 2,075) (95% CI [351, 444]) (relative increase of 6%), compared to baseline (M = 6,632 steps, SD = 3,026), however, the effects were small and the increase in step count did not reach the practical relevance threshold of 1,000 steps.

Exploratory analysis H1: Step counts increase separately for winners and losers

Additionally, the daily average step counts changes were analysed for winners and losers (see *Table 1* for an overview of the descriptive results). The analysis suggests that there was a significant difference in step count change between winners and losers, $t(7,673) = 42.409$, $p < .001$, $d = .996$, 95% CI [.948, 1.045], $d = -.996$, 95% CI [.948; 1.045]. To determine if the changes were significant for winners and losers separately, a further analysis revealed that for winners, daily step count during the challenge increased significantly, $t(4,741) = 46.779$, $p < .001$, $d = .679$, 95% CI [.648; .711]. The daily average step counts increased by 1,109 steps (SD = 1,632) (95% CI [1,062, 1,155]) (15.69% increase) compared to baseline. The effect size is moderate to large, and the increase in step count reached the practical relevance threshold of 1,000 steps. On the other hand, participants who lost the challenge decreased their step count significantly, $t(2,932) = -18.547$, $p < .001$, $d = -.342$, 95% CI [-.380; -.305]. The daily average step counts during a challenge decreased by -752 steps (SD = 2,196) (95% *CI* [-832, -673]) (12.68% decrease) compared to baseline. The effect size is moderate, but it is under the 1,000 steps threshold for practical relevance.

Hypothesis testing H2: People in deposit contracts group have higher step count increases compared to people in the accountability partner group.

In line with the hypothesis, participants in the deposit contract condition showed a significantly higher step count increase compared to the accountability partner group, with a mean difference of 1,357 steps (95% CI [1,200.350, 1,513.037]). See *Table 1* for descriptive statistics. Specifically, the deposit contract group increased their step count by 525 steps (*SD* = 2,013) while the accountability partner group decreased their step counts by -832 steps (*SD* = 2,258), $t(7{,}673) = 17.011$, $p < .001$, $d = -.666$, 95% CI [.588; .743]. The effect size is moderate, but the step count changes did not reach the practical relevance threshold of 1,000 steps.

Because we had indications that the assumption of equal variances was violated (see *Appendix 4* for details on assumptions check) we ran a non-parametric Mann-Whitney U Test. The results indicated significant step count differences between the two groups (deposit contract vs. accountability partner), $Z = -15.875$, $p < .001$, $r = 0.18$. The deposit contract group had a higher mean rank than the accountability partner group (3,967 vs. 2,590), as expected, participants in the deposit contract group showed a greater step count improvement than the accountability partner group but the effects were small.

Hypothesis testing H3: Challenge success is higher in the deposit contract group than in the accountability partner group.

To measure challenge success, a crosstabulation (see *Table 1*) showed that 61.8% of challenges were successful (winners) while 38,2% were not successful (losers). No major violation against the assumptions of a chi-square test of independence was detected (see A*ppendix 4* for detailed assumptions check). In line with the hypothesis, the chi-square test of independence showed a significantly higher success rate of challenges in the deposit contract group (66.9%) compared to challenges in the accountability partner group (12.8%), $\chi^2(1, N = 7{,}675) = 808.197$, $p < .001$, $\varphi = .325$. The effect size is moderate and as expected, deposit contracts are associated with higher success rates than accountability partners.

Sensitivity check H1: Step count increase for free trial and paid membership separately

The analysis suggests that there was a significant difference in step count change between free trial and paid membership, $t(7{,}674) = 16.765$, $p < .001$, $d = .191$, 95% CI [.169,.214]. The step count changes were significant for the two membership types separately. The free trial

participants increased their steps significantly, $t(3,048) = 5.531$, $p < .001$, $d = .1$, 95% CI [.065; .136]. The daily average step counts increased in the game by 251 steps ($SD = 2,507$) (95% CI [162.138, 340.212]) (4.5% increase) compared to baseline. Also, participants with a paid membership increased their step count significantly, $t(4,625) = 19.480$, $p < .001$, $d = .286$, 95% CI [.257; .316]. The daily average step counts during a challenge increased by 494 steps ($SD =$ 1,726) (95% *CI* [444.524, 544.011]) (6.8% increase) compared to baseline. However, the effect sizes were small, and the step count increase did not reach the practical relevance threshold of 1,000 steps.

Sensitivity check H2: Step count increase per group for paid membership (free trial exclusion)

To exclude eventual effects of the membership type (free trial vs. paid membership), a sensitivity check was used to investigate if the results in H2 are confirmed when only participants with a paid membership played in the challenges. The results remain in line with H2. See *Table 1* for descriptive statistics. There is a significant difference in step count changes with a mean difference = 846 steps, (95% CI [595.625, 1,102.383) between the deposit contract group ($M = 528$ step increase, $SD = 1,712$) and accountability partner group ($M = -321$ step decrease, $SD = 1,860$), $t(4,624) = 6.569$, $p < .001$, $d = .494$, 95% CI [.346; .642]. These results have a moderate effect size and show that deposit contracts are associated with higher step count increase than accountability partners, regardless of membership type, but the step count changes did not reach the practical relevance threshold of 1,000 steps.

Sensitivity check H3: Winning rates for paid membership only (free trial exclusion)

Similarly, as for the step count increase, we investigated if we could confirm the findings in H3 for paying members only. After exclusion of the free trial participants, the results remain in line with H3. See *Table 1* for descriptive statistics. In the deposit contract group the success rate was significantly higher (74%) compared to the accountability partner group (22%), $\chi^2(1, N = 4,626) = 239.563$, $p < .001$, $\varphi = .228$. The effect size is small to moderate and shows that deposit contracts are associated with higher success rates than accountability partners regardless of membership type.

Discussion

While the benefits of sustained moderate physical activity are widely recognized, (Pedersen & Saltin, 2015; WHO, 2020), some individuals may require incentives to stay motivated (Budworth et al, 2019; Gneezy et al., 2011; Mitchell et al., 2020). This study explored the effectiveness of social vs. financial incentives on step counts in a 2-week challenge. We expected an overall step count increase for both groups, additionally we expected that the deposit contract group shows higher step count increases and higher success rates than the accountability partner group. In line with our hypotheses, we observed an overall increase in average daily step counts of 6% during the two-week challenge (7,030 steps) compared to baseline (6,632 steps). This may be explained with the presence of the financial and social incentives which may have provided participants with a sense of commitment to increase their step count.

Furthermore, we found that compared to baseline, participants in the deposit contract group had an average daily step count increase of 525 steps, while the participants in the accountability group decreased them by -832 steps. Furthermore, the deposit contract groups also showed higher winner rates (67%) compared to participants in the accountability partner group (13%), supporting the hypothesis that loss aversion plays a critical role in motivating individuals to obtain successful outcomes. However, this study was observational, and we cannot draw causal conclusions.

First, the finding that participants at the challenges increased their steps (by 398 steps) compared to baseline, aligns with previous research showing that incentives can effectively increase PA levels (Budworth et al, 2019; de Buisonjé et al., 2022; Gneezy et al., 2011; Mitchell et al., 2020). This finding is in line with earlier studies that demonstrate that deposit contracts enhance behaviour change (Boonmanunt et al., 2022) but this step count increase did not reach the practical relevance threshold of 1,000 steps. Moreover, the step count increase is lower than the results found in a meta-analysis of randomized controlled trials with other types of financial incentives which demonstrated a much higher increase in step counts with a mean of 607 steps (10-15% increase compared to baseline) (Mitchell et al., 2019). Furthermore, a recent study by de Buisonjé et al. (2023) which also investigated step count increases in a naturalistic setting, found that participants in the two-week challenges with deposit contracts increased their step counts by 31.2% compared to baseline. This increase is much higher than observed in our study

(6% increase), suggesting that the two distinct incentives (deposit contracts and accountability partners) we investigated may have had different effects than the deposit contracts alone.

Interestingly, when differentiating between winners and losers, we found that people who won the challenge had an increase in step count by 1,109 steps (16%), while those who lost, decreased their steps count by -752 (-13%). Winning a challenge was related to a significantly increased step count, reaching the practical relevance threshold of 1,000 steps, while losing in a challenge was related to a significant decreased step count. Since this study was observational, no causal assumptions can be made, but we can speculate on possible explanations. The challenges were designed on individual goal achievements on a weekly basis, and we suspect that participants who experienced an initial failure may have been affected by setback effects or the losing a streak effect, where one failure reduces motivation and increases the likelihood of subsequent failures (Silverman & Barasch, 2022; Wenzel et al., 2020). Early setbacks may have lowered their self-efficacy (ten Broeke & Adriaanse, 2023) and motivation (Silverman & Barasch, 2022; Wenzel et al., 2020). Additionally, an "all-or-nothing" mindset makes it difficult for people to accept partial failure (Wulsin et al., 1999), which may have led individuals to attribute failure to internal factors (e.g., laziness), and influenced them to drop up or give up potential improvements in the second week. In contrast, the individuals who succeeded achieving their weekly goals during the first week might have been more inclined to sustain their efforts during the second week and achieve higher step counts. The current data set did not allow for a detailed observation of step counts day by day during the challenge; therefore, we do not have observations of how participants behaved after a challenge was failed, we can only suspect that some people might have uninstalled the app or quit recording the steps. Future studies could investigate how step count behaviours change after a challenge is lost and help explain contributing factors to step count decreases.

Second, as expected, participants in the deposit contract group showed a significantly greater step count increase compared to the accountability partner group. This aligns with previous research participants in gamified health related behaviour intervention, where financial rewards were preferred due to their monetary nature as money can be used beyond the scope of the intervention (Van Dooren et al., 2018). Additionally, the step count increase in the deposit contract can likely be attributed to the financial pressure associated with loss aversion, participants perceiving the consequences of loss higher than potential gains (Kahneman and

Tversky, 1979) which can lead to a more sustained effort to goal attainment (Boonmanunt et al., 2022). Interestingly, the accountability partner group not only performed worse than the deposit contract group, which saw an increase of 525 steps (7.7%), but it experienced a significant decrease in step counts of -832 steps (16%). While we expected the deposit contract group to outperform the accountability group, the highly significant decrease in step counts in the accountability partner group was surprising. Given that the dataset lacks details to explain the step count decrease, further research is needed to examine behavioural differences. It is possible that some participants might chose the accountability partner option as a low-commitment alternative to avoid financial risk, suggesting weaker initial intentions to increase step counts and mere curiosity about the challenge (Modirshanechi et al., 2023). This underscores the role of goal commitment in behavioural interventions, which, may have been weaker in the accountability partner group. To confirm this, future studies could incorporate questionnaires to assess commitment levels before, during and after the challenge.

Third, as hypothesized, participants in the deposit contract group also had a significantly higher winning ratio (67%) compared to those in the accountability partner group (13%). This aligns with our results in H2 and speculate that the loss perception was higher with financial risk (Patel et al. 2016), because monetary rewards were preferred over social rewards (Van Dooren et al., 2018). In addition, the present bias (Wang and Sloan, 2018) likely influenced the deposit contract group's engagement, as they were motivated by the immediate financial rewards associated with winning the challenge. While the expected pattern of behaviour was largely confirmed, this correlational study found deposit contracts to be superior to accountability partners, but highlighted concerns about a potential link between participants' commitment to step count increases and incentive choice.

Strengths and Limitations

This study was conducted on a large demographically heterogenous sample in a naturalistic observation which makes the results highly representative of the real-world behaviour in general population. Additionally, the participation to the challenges was based on user's own personal intention to join in a step count challenge which ensures that our findings are generalizable and applicable to similar situations. However, the results only indicate short-

term effects, therefore longitudinal studies are necessary to investigate long-term behaviour changes.

Moreover, participants were able to choose how they join a challenge, either by paying a $10 deposit or by getting an accountability partner instead. The self-selection of participants may have introduced biases, as individuals with different psychological traits, commitment levels or low SEP may have gravitated towards one condition over the other. Future studies can use extensive questionnaires to investigate how eventual differences may influence the selection of an incentive.

Furthermore, there may have been high variability in the accountability partner condition caused by a lack of control of how the accountability partnership was maintained after the invitation link was sent to a partner. In addition, there were no guidelines to structure the choice of a partner, the relationship or the quality and quantity of their communication. Without clear guidance or sufficient communication, the motivational power of an accountability partner may have been reduced, highlighting the need for more structured social incentive designs. Future research should explore methods to optimize social incentive designs by fostering stronger partner engagement. WayBetter could include additional gamification and social media elements (badges, avatars, comments, likes) to facilitate engagement between the participant and their accountability partner directly on the app.

Additionally, we suspect that some players who lost might have stopped playing after experiencing setbacks. For example, offering a partial deposit refund for continuous participation could encourage those who lost to stay engaged. Future testing of strategy adjustments would provide further insights into how these changes impact participation.

Implications

These findings contribute significantly to both behavioural economics and gamification literature by highlighting the varying effectiveness of social and financial incentives on health-related behaviour change. However, they are of correlational nature, and they need to be confirmed by future experimental studies. We found that deposit contracts were superior to accountability partners, and we mainly attribute the effects to the direct financial commitment of deposit contracts, which likely triggered loss aversion. Moreover, we suspect that the effects of accountability partners (e.g., support, encouragement) may have been perceived at a lower value,

because social incentives could not be reached in the current model, due to the very simplistic interaction with the accountability partner (sending an invitation link). This may have not provided enough social engagement. Furthermore, low commitment participants may have chosen the accountability partner as a free alternative to avoid genuine commitment. However, despite their complexity, accountability partners offer a more accessible option for people with a low SEP. Interventions with social incentives are therefore inclusive to those with limited finance, making it crucial for smartphone-based intervention designers to ensure steady and genuine engagement with accountability partners.

This correlational study found that participants, who won the challenge increased their average step counts by 1,109 steps, which is exceeding the threshold (> 1,000 steps) for reducing mortality risk (Banach et al., 2023). However, losers decreased their step count by -752 steps which demonstrates a large gap between winners and losers. This gap highlights the need for interventions that encourage participation, facilitate success but also offers alternatives to prevent dropouts. For example, future interventions could avoid an eventual step count decrease if they offer additional rewards even in case of losing, to ensure engagement even after they expect to lose.

Conclusion

In conclusion, this study highlights the superiority of financial incentives through deposit contracts in promoting increased physical activity and successful outcomes in gamified challenges. The analysed sample included data on behaviour changes regarding step count increases in individuals who participated at challenges with the intention to improve their levels of moderate activity. The associated steps count increases in the deposit contract group is in line with numerous studies demonstrating that financial incentives are effective motivators. However, the accountability partner group was associated with significant step count decreases. While this cohort observational study cannot establish causality, its findings underscore the need for carefully designed interventions that optimize social incentive mechanisms. Specifically, social incentives can offer an alternative to overcoming financial barriers and risk aversion found in deposit contracts. Therefore, it is crucial to explore interventions with potential to reduce health disparities and provide equal opportunities and access to interventions that help individuals drive behaviour change to reach better health outcomes.

References

Banach, M., Lewek, J., Surma, S., Penson, P. E., Sahebkar, A., Martin, S. S., Bajraktari, G., Henein, M. Y., Reiner, Ž., Bielecka-Dąbrowa, A., & Bytyçi, I. (2023). The association between daily step count and all-cause and cardiovascular mortality: A meta-analysis. *European Journal of Preventive Cardiology*, *30*(18), 1975–1985. https://doi.org/10.1093/eurjpc/zwad229

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-l

Berscheid, E., Snyder, M., & Omoto, A. M. (1989). The relationship closeness inventory: Assessing the closeness of interpersonal relationships. *Journal of Personality and Social Psychology*, *57*(5), 792–807. https://doi.org/10.1037/0022-3514.57.5.792

Beyer, H. (1981b). Tukey, John W.: Exploratory data analysis. Addison-Wesley Publishing Company Reading, Mass. Menlo Park, Cal., London, Amsterdam, Don Mills, Ontario, Sydney 1977, XVI, 688 s. *Biometrical Journal*, *23*(4), 413–414. https://doi.org/10.1002/bimj.4710230408

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*(2), 265–292. https://doi.org/10.1037/0033-2909.94.2.265

Boonmanunt, S., Pattanaprateep, O., Ongphiphadhanakul, B., McKay, G., Attia, J., Vlaev, I., & Thakkinstian, A. (2022). Evaluation of the effectiveness of behavioral economic incentive programs for goal achievement on healthy diet, weight control and physical activity: A systematic review and network meta-analysis. *Annals of Behavioral Medicine*, kaac066. https://doi.org/10.1093/abm/kaac066

Budworth, L., Prestwich, A., Sykes-Muskett, B., Khatun, K., Ireland, J., Clancy, F., & Conner, M. (2019). A feasibility study to assess the individual and combined effects of financial incentives and monetary contingency contracts on physical activity. *Psychology of Sport and Exercise*, *44*, 42–50. https://doi.org/10.1016/j.psychsport.2019.04.021

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, S., Underwood, L. G., & Gottlieb, B. H. (2000). *Social Support Measurement and Intervention*. Oxford University Press. https://doi.org/10.1093/med:psych/9780195126709.001.0001

Cohen, S., & Janicki-Deverts, D. (2009). Can we improve our physical health by altering our social networks? *Perspectives on Psychological Science*, *4*(4), 375–378. https://doi.org/10.1111/j.1745-6924.2009.01141.x

Collins, K. A., Huffman, K. M., Wolever, R. Q., Smith, P. J., Siegler, I. C., Ross, L. M., Hauser, E. R., Jiang, R., Jakicic, J. M., Costa, P. T., & Kraus, W. E. (2022). Determinants of dropout from and variation in adherence to an exercise intervention: The STRRIDE Randomized Trials. *Translational Journal of the American College of Sports Medicine*, *7*(1). https://doi.org/10.1249/tjx.0000000000000190

Cox, D. (2018) Watch your step: why the 10,000 daily goal is built on bad science The Guardian. Available from https://www.theguardian.com/lifeandstyle/2018/sep/03/watch-your-step-why-the-10000-daily-goal-is-built-on-bad-science [Accessed on 25 Feb 2025]

Creasy, S. A., Lang, W., Tate, D. F., Davis, K. K., & Jakicic, J. M. (2018). Pattern of daily steps is associated with weight loss: Secondary analysis from the step-up randomized trial. *Obesity*, *26*(6), 977–984. https://doi.org/10.1002/oby.22171

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L., & Taylor, N. G. A. (2020b). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. *JMIR Mhealth and Uhealth*, *8*(9), e18694. https://doi.org/10.2196/18694

Cumming, G. (2013). The new statistics. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

de Buisonjé, D. R., Brosig, F., Breeman, L. D., Bloom, E. L., Reijnders, T., Janssen, V. R., Kraaijenhagen, R. A., Kemps, H. M., & Evers, A. W. (2023). Put your money where your feet are: The real-world effects of StepBet gamified deposit contracts for physical activity. *Internet Interventions*, *31*, 100610. https://doi.org/10.1016/j.invent.2023.100610

de Buisonjé, D.R., Reijnders, T., Cohen Rodrigues, T.R., Prabhakaran, S., Kowatsch, T., Lipman, S.A., Bijmolt, T.H.A., Breeman, L.D., Janssen, V.R., Kraaijenhagen, R.A., Kemps, H.M.C., Evers, A.W.M., 2022. Less carrot more stick? Investigating rewards and deposit contract financial incentives for physical activity behavior change using a smartphone application: randomised controlled trial (Preprint). *Journal of Medical Internet Research.* https://doi.org/10.2196/preprints.38339

Deci, E. L., & Ryan, R. M. (1985). Intrinsic motivation and self-determination in human behavior. *Springer eBooks*. https://doi.org/10.1007/978-1-4899-2271-7

Del Pozo Cruz, B., Ahmadi, M., Naismith, S. L., & Stamatakis, E. (2022). Association of daily step count and intensity with incident dementia in 78 430 adults living in the UK. *JAMA Neurology*, *79*(10), 1059. https://doi.org/10.1001/jamaneurol.2022.2672

Del Pozo Cruz, B., Ahmadi, M. N., Lee, I., & Stamatakis, E. (2022b). Prospective associations of daily step counts and intensity with cancer and cardiovascular disease incidence and mortality and all-cause mortality. *JAMA Internal Medicine*, *182*(11), 1139. https://doi.org/10.1001/jamainternmed.2022.4000

Donlin Washington, W., McMullen, D., Devoto, A., 2016. A matched deposit contract intervention to increase physical activity in underactive and sedentary adults. *Translational Issues in Psychological Science*, *2*(2), 101–115. https://doi.org/10.1037/tps0000069

Dugas, M., Gao, G., & Agarwal, R. (2020). Unpacking mHealth interventions: A systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. *Digital Health*, *6*. https://doi.org/10.1177/2055207620905411

Faries, M. D. (2016). Why We Don't "Just Do It": Understanding the intention-behavior gap in lifestyle medicine. *American Journal of Lifestyle Medicine*, *10*(5), 322–329. https://doi.org/10.1177/1559827616638017

Feeney, B. C., & Collins, N. L. (2015). A new look at social support: A theoretical perspective on thriving through relationships. *PubMed*, *19*(2), 113–147. https://doi.org/10.1177/1088868314544222

Fehr, E., & Falk, A. (2002). Psychological foundations of incentives. *European Economic Review*, *46*(4–5), 687–724. https://doi.org/10.1016/s0014-2921(01)00208-2

Fortuna, K. L., Brooks, J. M., Umucu, E., Walker, R., & Chow, P. I. (2019). Peer support: A human factor to enhance engagement in digital health behavior change interventions. *Journal of Technology in Behavioral Science*, *4*(2), 152–161. https://doi.org/10.1007/s41347-019-00105-x

Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., Buote, R., Van Heerden, D., Luan, H., Cullen, K., Slade, L., & Taylor, N. G. A. (2020). Reliability and validity of commercially available wearable devices for measuring steps, energy

expenditure, and heart rate: systematic review. *JMIR Mhealth and Uhealth*, *8*(9), e18694. https://doi.org/10.2196/18694

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives*, *25*(4), 191–210. https://doi.org/10.1257/jep.25.4.191

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2018). Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global Health*, *6*(10), e1077–e1086. https://doi.org/10.1016/s2214-109x(18)30357-7

Haff, N., Patel, M. S., Lim, R., Zhu, J., Troxel, A. B., Asch, D. A., & Volpp, K. G. (2015). The role of behavioral economic incentive design and demographic characteristics in financial incentive-based approaches to changing health behaviors: A meta-analysis. *American Journal of Health Promotion*, *29*(5), 314–323. https://doi.org/10.4278/ajhp.140714-lit-333

Harrison, J. D., Jones, J. M., Small, D. S., Rareshide, C. A., Szwartz, G., Steier, D., Guszcza, J., Kalra, P., Torio, B., Reh, G., Hilbert, V., & Patel, M. S. (2019). Social incentives to encourage physical activity and understand predictors (STEP UP): Design and rationale of a randomized trial among overweight and obese adults across the United States. *Contemporary Clinical Trials*, *80*, 55–60. https://doi.org/10.1016/j.cct.2019.04.001

Hsu, M., Lee, S., Yang, H., & Chao, H. (2021). Is brisk walking an effective physical activity for promoting taiwanese adolescents' mental health? *Journal of Pediatric Nursing*, *60*, e60–e67. https://doi.org/10.1016/j.pedn.2021.03.012

Hunter, R. F., Tang, J., Hutchinson, G., Chilton, S., Holmes, D., & Kee, F. (2018). Association between time preference, present-bias and physical activity: Implications for designing behavior change interventions. *BMC Public Health*, *18*(1). https://doi.org/10.1186/s12889-018-6305-9

Hurmuz-Bodde, M. Z., Jansen-Kosterink, S. M., Hermens, H. J., & Van Velsen, L. (2024). Attrition of older adults in web-based health interventions: Survival analysis within an observational cohort study. *Journal of Health Psychology*. https://doi.org/10.1177/13591053241274097

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kurti, A. N., Davis, D. R., Redner, R., Jarvis, B. P., Zvorsky, I., Keith, D. R., Bolivar, H. A., White, T. J., Rippberger, P., Markesich, C., Atwood, G., & Higgins, S. T. (2016). A review of the literature on remote monitoring technology in incentive-based interventions for health-related behavior change. *Translational Issues in Psychological Science*, *2*(2), 128–152. https://doi.org/10.1037/tps0000067

Kerrigan, S. G., Forman, E. M., Patel, M., Williams, D., Zhang, F., Crosby, R. D., & Butryn, M. L. (2019). Evaluating the feasibility, acceptability, and effects of deposit contracts with and without daily feedback to promote physical activity. *Journal of Physical Activity and Health*, *17*(1), 29–36. https://doi.org/10.1123/jpah.2018-0589

Laibson, D., 1997. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics 112*(2), 443–478. https://doi.org/10.1162/003355397555253

Lin, M., Lucas, H. C., & Shmueli, G. (2013). Research Commentary—Too big to fail: Large samples and the P-Value problem. *Information Systems Research*, *24*(4), 906–917. https://doi.org/10.1287/isre.2013.0480

Lunze, K., & Paasche-Orlow, M. K. (2013). Financial incentives for healthy behavior: Ethical safeguards for behavioral economics. *American Journal of Preventive Medicine*, *44*(6), 659–665. https://doi.org/10.1016/j.amepre.2013.01.035

Mammen, G., & Faulkner, G. (2013). Physical activity and the prevention of depression. *American Journal of Preventive Medicine*, *45*(5), 649–657. https://doi.org/10.1016/j.amepre.2013.08.001

Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J. P. T., & Marteau, T. M. (2015). Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. *Preventive Medicine*, *75*, 75–85. https://doi.org/10.1016/j.ypmed.2015.03.001

Modirshanechi, A., Kondrakiewicz, K., Gerstner, W., & Haesler, S. (2023). Curiosity-driven exploration: Foundations in neuroscience and computational modeling. *Trends in Neurosciences (Regular Ed.)*, *46*(12), 1054–1066. https://doi.org/10.1016/j.tins.2023.10.002

Munson, S. A., Krupka, E., Richardson, C., & Resnick, P. (2015). Effects of public commitments and accountability in a technology-supported physical activity intervention. *Association for Computing Machinery,* 1135–1144. https://doi.org/10.1145/2702123.2702524

Müssener, U. (2021). Digital encounters: Human interactions in mHealth behavior change interventions. *Digital Health*, *7*. https://doi.org/10.1177/20552076211029776

Oussedik, E., Foy, C. G., Masicampo, E. J., Kammrath, L. K., Anderson, R. E., & Feldman, S. R. (2017). Accountability: A missing construct in models of adherence behavior and in clinical practice. *Patient Preference and Adherence*, *Volume 11*, 1285–1294. https://doi.org/10.2147/ppa.s135895

Paluch, A. E., Bajpai, S., Bassett, D. R., Carnethon, M. R., Ekelund, U., Evenson, K. R., Galuska, D. A., Jefferis, B. J., Kraus, W. E., Lee, I., Matthews, C. E., Omura, J. D., Patel, A. V., Pieper, C. F., Rees-Punia, E., Dallmeier, D., Klenk, J., Whincup, P. H., Dooley, E. E., Fulton, J. E. (2022). Daily steps and all-cause mortality: A meta-analysis of 15 international cohorts. *The Lancet Public Health*, *7*(3), e219–e228. https://doi.org/10.1016/s2468-2667(21)00302-9

Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., Sproat, S., Hyson, C., Haff, N., Lee, S. M., Wesby, L., Hoffer, K., Shuttleworth, D., Taylor, D. H., Hilbert, V., Zhu, J., Yang, L., Wang, X., & Volpp, K. G. (2016). Framing financial incentives to increase physical activity among overweight and obese adults: A randomized, controlled trial. *Annals of Internal Medicine*, *164*(6), 385. https://doi.org/10.7326/M15-1635

Patel, M. S., Small, D. S., Harrison, J. D., Fortunato, M. P., Oon, A. L., Rareshide, C. a. L., Reh, G., Szwartz, G., Guszcza, J., Steier, D., Kalra, P., & Hilbert, V. (2019). Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the United States. *JAMA Internal Medicine*, *179*(12), 1624. https://doi.org/10.1001/jamainternmed.2019.3505

Patel, M. S., Benjamin, E. J., Volpp, K. G., Fox, C. S., Small, D. S., Massaro, J. M., Lee, J. J., Hilbert, V., Valentino, M., Taylor, D. H., Manders, E. S., Mutalik, K., Zhu, J., Wang, W., & Murabito, J. M. (2017). Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families. *JAMA Internal Medicine*, *177*(11), 1586. https://doi.org/10.1001/jamainternmed.2017.3458

Pedersen, B. K., & Saltin, B. (2015). Exercise as medicine – evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scandinavian Journal of Medicine and Science in Sports*, *25*(S3), 1–72. https://doi.org/10.1111/sms.12581

Pirotta, S., Joham, A., Hochberg, L., Moran, L., Lim, S., Hindle, A., & Brennan, L. (2019). Strategies to reduce attrition in weight loss interventions: A systematic review and meta-analysis. *Obesity Reviews*, *20*(10), 1400–1412. https://doi.org/10.1111/obr.12914

Royer, H., Stehr, M., & Sydnor, J. (2015). Incentives, commitments, and habit formation in exercise: Evidence from a field experiment with workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, *7*(3), 51–84. https://doi.org/10.1257/app.20130327

Rusbult, C. E., Finkel, E. J., & Kumashiro, M. (2009). The Michelangelo phenomenon. *Current Directions in Psychological Science*, *18*(6), 305–309. https://doi.org/10.1111/j.1467-8721.2009.01657.x

Santarossa, S., Kane, D., Senn, C. Y., & Woodruff, S. J. (2018). Exploring the role of In-Person components for online health behavior change interventions: Can a digital Person-to-Person component suffice? *Journal of Medical Internet Research*, *20*(4), e144. https://doi.org/10.2196/jmir.8480

Silverman, J., Barasch, A., 2022. On or Off Track: How (broken) streaks affect consumer decisions. *Journal of Consumer Research, 49*(6), 1095–1117. https://doi.org/10.1093/jcr/ucac029

Song, J., & Xu, P. (2023). Healthier together: Social support, self-regulation and goal management for chronic conditions in online health communities. *Information & Management*, *60*(7), 103830. https://doi.org/10.1016/j.im.2023.103830

Tamir, M., & Gutentag, T. (2017). Desired emotional states: their nature, causes, and implications for emotion regulation. *Current Opinion in Psychology*, *17*, 84–88. https://doi.org/10.1016/j.copsyc.2017.06.014

ten Broeke, P., Adriaanse, M.A., (2023). Understanding the setback effect in everyday self-regulation. *European Journal of Social Psychology*, *53*(4), 703–719. https://doi.org/10.1002/ejsp.2931

Tudor-Locke, C., Corbin, C. B., Pangrazi, R. P., & Franks, D. (2002). Taking steps toward increased physical activity: Using pedometers to measure and motivate [Dataset]. In *PsycEXTRA Dataset*. https://doi.org/10.1037/e603452007-001

Tudor-Locke, C., Craig, C. L., Brown, W. J., Clemes, S. A., De Cocker, K., Giles-Corti, B., Hatano, Y., Inoue, S., Matsudo, S. M., Mutrie, N., Oppert, J., Rowe, D. A., Schmidt, M. D., Schofield, G. M., Spence, J. C. H., Teixeira, P. J., Tully, M. A., & Blair, S. N. (2011). How many steps/day are enough? for adults. *International Journal of Behavioral Nutrition and Physical Activity*, *8*(1), 79. https://doi.org/10.1186/1479-5868-8-79

Tudor-Locke, C., & Lutes, L. (2009). Why do pedometers work? *Sports Medicine*, *39*(12), 981–993. https://doi.org/10.2165/11319600-000000000-00000

Tukey, J.W. (1977) Exploratory Data Analysis.

Van Dooren, M. M., Visch, V. T., & Spijkerman, R. (2018). Rewards that make you play: The distinct effect of monetary rewards, virtual points and social rewards on play persistence in substance dependent and non-dependent adolescents. *2018 IEEE 6th International Conference on Serious Games and Applications for Health (SeGAH),* 1–7. https://doi.org/10.1109/segah.2018.8401312

Wang, Y., & Sloan, F. A. (2018). Present bias and health. *Journal of Risk and Uncertainty*, *57*(2), 177–198. https://doi.org/10.1007/s11166-018-9289-z

Wenzel, M., Rowland, Z., Hofmann, W., Kubiak, T., 2020. Setbacks in self-control: Failing not mere resisting impairs subsequent self-control. *Social Psychological and Personality Science*, *11*(6), 782–790. https://doi.org/10.1177/1948550619888875

World Health Organization. (2020). *WHO guidelines on physical activity and sedentary behaviour: at a glance*. World Health Organization.

World Health Organization: WHO. (2018). *Physical activity WPRO*. https://www.who.int/westernpacific/health-topics/physical-activity#:~:text=Physical%20activity%20in%20the%20Western%20Pacific&text=Conversely%2C%20physical%20inactivity%20has%20been,estimated%203.2%20million%20deaths%20globally.

World Health Organization: WHO. (2022). *Global Status Report on Physical Activity 2022*. World Health Organization. https://www.who.int/teams/health-promotion/physical-activity/global-status-report-on-physical-activity-2022

Wulsin, L. R. (1999). Practicing cognitive therapy: A guide to interventions. *The Journal of Psychotherapy Practice and Research*, *8*(4), 318–319.

Xu, Z., Zahradka, N., Ip, S., Koneshloo, A., Roemmich, R. T., Sehgal, S., Highland, K. B., & Searson, P. C. (2022). Evaluation of physical health status beyond daily step count using a wearable activity sensor. *Npj Digital Medicine*, *5*(1). https://doi.org/10.1038/s41746-022-00696-5

Appendix 1: Exploratory Analysis on Free Trial Distribution per Group

A Chi-Square test of independence indicated that the participants with a free trial membership versus participants with a paid memberships were not evenly distributed across the two conditions (deposit contract versus accountability partner) (see *Table 2* for descriptive results), $\chi^2(1, N = 7{,}675) = 399.962$, $p < .001$, $\varphi = .228$. Specifically, a higher proportion of participants in the accountability partner group were free trial members (74,4%) compared to the deposit contract group (36.1%).

Table 2

Descriptive Results Per Incentive Group (N = 7,675)

	Deposit Contract	Accountability Partner	**Total**
Membership type			
Free Trial	2,513	536	3,049
Paid Membership	4,442	184	4,626
Free trial ratio	36.1%	74.4%	39.7% [a]
Total	6,955	720	7,675

Note. Sample excludes outliers

[a] Ratio of free memberships within all members.

Appendix 2: Outlier Detection

To identify outliers in daily step count data, a lower and an upper cutoff was computed with the interquartile range (IQR) method. The IQR method is a robust statistical approach that identifies outliers as values falling below 1.5 times the IQR from the first quartile (Q1) or above 1.5 times the IQR from the third quartile (Q3) (Beyer, 1981). The following calculations were performed in IBM SPSS 27: For the baseline steps the Q1 (25th percentile) had a value of 4,069.50 and the Q3 (75th percentile) had a value of 8,631.50. The IQR was computed: IQR = Q3 - Q1 = 8,631.50 - 4,069.50 = 4,562. The lower cutoff was calculated Q1 - (1.5 × IQR) = 4,069.50 - (1.5 × 4,562) = -2,773.50 but since this value was negative, it was not applicable. The upper cutoff was calculated Q3 + (1.5 × IQR) = 8,631.50 + (1.5 × 4,562) = 15,474.50. For the steps during the challenge, the same formulas were applied for Q1 and Q3 with IQR = Q3 - Q1 = 9,126.50 - 4,310.00 = 4,816.50. The lower cutoff (Q1 - (1.5 × IQR) = 4,310.00 - (1.5 × 4,816.50) = -3,514.75) was negative and not applicable. The upper cutoff (Q3 + (1.5 × IQR) = 9,126.50 + (1.5 × 4,816.50)) resulted in 16,051.25.

Finally, we determined which datapoints were considered as outliers based on the IQR results (15,474.50 and 16,051.25). We rounded the values to 16,000 as the upper cutoff. Since IQR detected no lower cutoff due to negative values, we set the minimum at 1,000 steps. Average step counts bellow this value indicate technical error or illness (Xu et al., 2022) rather than sedentary behaviour, which was beyond the scope of this study. This threshold still includes values bellow the sedentarism benchmark (Tudor-Locke et al., 2011) while it ensures inclusivity. Furthermore, in appendix 3 we report sensitivity checks including the outliers.

Appendix 3: Sensitivity Checks

Several sensitivity checks were computed to increase result's robustness and determine if our findings are confirmed also when extreme values are included. For H1 a paired samples t-test was used to confirm the previous findings with an additional Wilcoxen Signed Rank test to ensure a more comprehensive analysis which includes extreme values. For H2 we computed an independent samples t-test and a non-parametric Mann-Whitney U Test to compare the two groups (deposit contract vs accountability partner). We considered significance at $p < .05$ criterium, and effect size of Cohen's $d \geq .2$, $\geq .5$ and $\geq .8$ as small, medium, and large respectively (Cohen, 1988) and of $r \geq .1$, $\geq .3$, $\geq .5$ as weak, moderate, and strong respectively with r = z/sqrt(N).

Sensitivity Check H1

We expected both groups to increase their step counts compared to baseline. A paired sample t-test was computed including the outliers. See *Table 3* for descriptive results. In line with our previous findings, the steps count increased (M = 7,018, SD = 3,962) significantly compared to baseline (M = 6,698, SD = 3,769) by 320 steps (SD = 2,228), $t(8{,}471) = 13.233$, $p < .001$, d = .144, 95% CI [.122; .165]. However, the results were small. The Wilcoxen Signed Rank test confirmed that more people increased their step counts compared to baseline with positive ranks (N = 5,131) higher than negative ranks (N = 3,339) suggesting that most people increased their step counts, Z = -18.099, $p < .001$ with a small to medium effect size r = .197.

Sensitivity Check H2

We expected deposit contract group to show higher their step counts increases compared to accountability partner. This hypothesis was again confirmed that the deposit contract group showed higher increased step counts compared to the accountability partner group. Specifically, the deposit contract group increased their step count by 476 steps (SD = 2,158) while the accountability partner group decreased their step counts by -1,002 steps (SD = 2,367), $t(8{,}470)$ = 19.125, $p < .001$, d = -.677, 95% CI [.607; .747]. The effect size was moderate and the step increase of the deposit contract group did not reach the practical relevance threshold of 1,000 steps, but the decrease of the accountability partner did. Furthermore, the Mann-Whitney U Test confirmed the findings. The deposit contract group had a higher mean rank than the accountability partner group (4,407 vs. 2,788), Z = -18.686, $p < .001$ suggesting participants in

the deposit contract group experienced greater step count increases than the accountability partner group. The result of the robust, non-parametrical test showed that the effect size was small to medium $r = .203$.

Table 3

Descriptive Results Including Outliers (N = 8,472)

	Baseline Steps [a]	**Game Steps** [b]	**Change Steps** [c]	**Relative Change (%)** [d]	**Winn ratio** [e]	**Number of participants (N)**
All Players	6,698 (3,769)	7,018 (3,962)	320 (2228)	+4.78%	59.7%	7,675
Condition						
Deposit Contract	6,925 (3,801)	7,400 (3,887)	476 (2,158)	+6.87%	65.3%	7,581
Accountability Partner	4,769 (2,824)	3,767 (2,975)	-1,002 (2,367)	-21.00%	11.6%	891

Note. Data are frequencies, means (SD) and percent.

[a] Baseline Steps = Daily steps before starting the challenge.

[b] Game Steps = Daily steps in the game.

[c] Step Change = Relative change in steps.

[d] Relative Change % = Change steps/baseline steps.

[e] Win ratio % = Percentage winners within group.

Appendix 4: Assumptions Check

Assumptions of normality

For H1 we compared the step count increases between challenge steps and baseline steps. With a Kolmogorov-Smirnov test and a Q-Q plot we tested the assumption of the normality of the step differences (challenge steps - baseline steps). The test showed a significant result (D = 0.078, N = 7,675, $p < .001$) indicating violation of normality, but given the large sample size, further descriptive analysis was computed. The skewness and kurtosis showed values of -0.320 and 2.735, respectively, indicating that the data is approximately normally distributed.

Assumptions of normality and homogeneity

To compare step count changes between the deposit contract group and accountability partner group, the assumptions of normality and homogeneity were assessed. For normality checks, Q-Q plots and histograms and a Shapiro-Wilk test were conducted. Visual inspection of the Q-Q plots and histograms suggested that both groups were approximately normal. The Shapiro-Wilk test indicated significant deviations from normality for both groups (deposit contracts vs. accountability partner), $p < .001$. However, due to the large sample size (N = 7,675) further checks were performed. Descriptive statistics suggested a near-normal distribution: the skewness of the deposit contract was -0.234 and kurtosis was 2.98, while the accountability partner group a skewness of -0.555 and kurtosis of 1.297. Furthermore, a Levene's test indicated violation of the assumption of equal variances, $F(1, 7{,}673) = 289.36$, $p < .001$.

In an exploratory analysis a comparison of the step count changes between the free trial members and paying members was computed. No violation of normal distribution was observed for both, paying members (skewness = -0.004, kurtosis = 3.372) and free trial members (skewness = -0.352, kurtosis = 1.498), however a Levene's test indicated that the assumptions of homogeneity was violated, $F(1, 7{,}673) = 347.798$, $p < .001$.

Assumptions of independence and expected frequencies

For the Chi-Square Test which compared the winning rates between de deposit contract group and the accountability partner group, the assumptions of independence and expected frequencies were assessed. The results showed sufficient expected cell frequencies (all expected counts ≥ 5, with the smallest expected count being 275.1). Additionally, to help control for the

assumption of independence of observations we excluded simultaneous and subsequent games of each user during the data cleaning, ensuring that each case was independent of others.

Appendix 5: Step Count Change Threshold Rationale

Due to the large sample size (N = 7,675) of this study, we anticipated that even small effects could reach statistical significance (Cumming, 2014). To ensure that our findings also reflect practical relevance (Lin et al., 2013), we introduced a threshold of 1,000 steps to interpret practical relevance. The decision to use 1,000 steps as the minimum threshold for significant behavioural change is based on a meta-analysis by Banach et al., (2023), which suggests that an increase of 1,000 correlates with reduced risk of all-cause mortality by 15%. Given that the data includes participants willing to take part in a challenge to increase their PA levels, 1,000 steps is a reasonable threshold for a meaningful change. By setting this threshold we ensure that the step count change is impactful but also achievable for individuals driven to become more active.