Article:
Budworth, L, Prestwich, A orcid.org/0000-0002-7489-6415, Sykes-Muskett, B et al. (4 more authors) (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. pp. 42-50. ISSN 1469-0292

https://doi.org/10.1016/j.psychsport.2019.04.021

A FEASIBILITY STUDY TO ASSESS THE INDIVIDUAL AND COMBINED EFFECTS OF FINANCIAL INCENTIVES AND MONETARY CONTINGENCY CONTRACTS ON PHYSICAL ACTIVITY

Luke Budworth*, [1 2]

Andrew Prestwich**, [2]

Bianca Sykes-Muskett[2]

Khaledha Khatun[2]

James Ireland[2]

Faye Clancy[2]

Mark Conner[2]

[1]Bradford Institute for Health Research, Bradford Royal Infirmary, Temple Bank House, Duckworth Lane, Bradford, BD96RJ, UK.

[2]School of Psychology, University of Leeds, 4 Lifton Place, Leeds, LS29JZ, UK

*Corresponding author

** Corresponding author

Highlights

- Of 5 approaches, combining loss and gain-framed monetary rewards most raised steps
- Combining loss and gain-framed monetary rewards (vs. loss-framed) increased deposits
- Larger deposits led to higher step counts
- Interventions were rated as equally acceptable

Abstract

Objectives. To assess the feasibility, and demonstrate the preliminary relative efficacy of, individual and/or combined financial incentives interventions for physical activity.

Design. Eighty participants were randomized to conditions receiving either: (i) a monetary contingency contract (where individuals deposit money, forfeited or returned depending on goal achievement) plus a standard financial incentive (simple reward upon achievement), (ii) a monetary contingency contract only, (iii) a standard incentive only, or controls groups (iv) with or (v) without a set behavioural goal. Feasibility was investigated through assessment of intervention acceptability, cost-effectiveness, study retention, contamination and missing data. The effects of the interventions on (i) physical activity (daily steps over 2-weeks) and (ii) potential mediators (e.g. intentions) were assessed also.

Results. Indicators of feasibility were generally positive, with high acceptability ratings, low drop-out and low missing data. Participants receiving monetary contingency contracts plus standard financial incentives had (i) increased steps above controls (with some evidence of superiority over monetary contingency contract-only participants), (ii) the highest prevalence of goal achievement and cost-effectiveness (being between 57-317% cheaper per goal achiever versus other conditions) and (iii) larger deposits than contingency contract-only participants (with some evidence that higher deposits increased steps). There was evidence of contamination between participants, but the results were mostly robust after excluding 'contaminated' participants. No differences were observed on psychological mediators.

Conclusion. This feasibility trial found promising results for a combined strategy approach to physical activity incentivisation, though a larger confirmatory trial is required.

Keywords. Behavioural Economics, Behaviour Change, Financial Incentives, Money Contingency Contracts, Physical Activity, Walking Behaviour.

Introduction

Across the developed world, physical inactivity is endemic (Dumith et al., 2011), contributing to increasing obesity and associated diseases (González, Fuentes & Márquez, 2017). In the USA, physical inactivity may account for around 8.7% of healthcare expenditure (Carlson, Fulton, Pratt, Yang & Adams, 2015) and in the UK, the tax-payer funded National Health Service spends around £1.2 billion annually due to inactivity related conditions (BHF, 2017) (see also Ding et al. 2016). Perhaps the most accessible way of buffering against inactivity related conditions is to increase walking; a 20% increase in steps-per-day may produce tangible health benefits (Dasgupta et al. 2017; Hajna, Ross & Dasgupta, 2017; Ewald, Oldmeadow & Attia, 2017). As a rule of thumb at least 10,000 daily steps, which many governments and organisations endorse, appears a healthy aim for most people (Wattanapisit & Thanamee, 2017).

A challenge in changing physical activity is that its long-term rewards (e.g., improved fitness) compete against more immediate rewards offered by other behaviours (e.g., watching television). Offering individuals financial incentives (FIs) for physical activity may address this issue. Promberger and Marteau (2013) suggest if FIs are sufficiently large, certain, and close enough in time to a behaviour, they may tilt the balance in its favour (see Leonard & Shuval, 2017). While FIs, as an extrinsic motive, *may* undermine intrinsic motivation (a key determinant of sustained behaviour change) (Johnston & Sniehotta, 2010; Deci, Koestner & Ryan, 1999) Promberger and Marteau (2013) argue that this is unlikely. First, individuals offered FIs are likely those inherently low in intrinsic motivation (because they do not exercise regularly) and second, intrinsic motivation may not map strongly onto physical activity because, for example, individuals could exercise for less intrinsically motivating reasons (e.g., to improve appearance). Further, while non-predictable versus fixed rewards may be more effective in generating habitual behaviour (Wood & Neal, 2016), it may be that

FIs can at least help establish habitual physical activity; after which the cessation of extrinsic rewards may not threaten long term maintenance (Acland & Levy, 2015; Charness & Gneezy, 2009).

Although public attitudes towards the use of FIs to change health behaviours are generally negative (especially compared to alternative interventions) (Promberger, Brown, Ashcroft & Marteau, 2011), they are judged more acceptable when incentives are perceived as effective (Promberger, Dolan & Marteau, 2012). Further, from the perspective of the target individual, incentives are likely to be at least as acceptable as no incentives for various health behaviours (Giles, Becker, Ternant, Sniehotta, McColl & Adams, 2016).

Studies have compared FIs against control groups on physical activity with promising results (e.g., Courneya, Estabrooks, & Nigg, 1997; Finkelstein et al. 2008; Pope & Harvey-Berino, 2013). Larger rewards and rewards (i) received only upon achieving a goal and (ii) in non-lottery-based structures, appear more effective (Barte & Wendel-Vohs, 2017; Mitchell et al., 2013). However, the impact of FIs have often been tested as *additions* to existing programmes (e.g., Shin et al., 2017) which is problematic because it is unclear whether the FI is effective *individually*. Studies also typically investigate specific FI modalities only, such as changes in fuel prices (e.g., Hou et al., 2011), discounted gym use (Tanham, Murphy & Breslin, 2014), and paying participants contingent on their levels of physical activity (e.g., Finkelstein, Brown, Brown & Buchner 2008). While testing specific types of FIs versus controls is useful, it may be more useful to directly compare different types of FIs; some FIs may be more cost-effective.

In a rare within-study comparison of different types of FIs, Patel et al. (2016) reported that loss-framed monetary incentives (losing $1.40 daily for not meeting a step goal) but not gain-based incentives (earning $1.40 for every day a step goal was met) or cash-based lotteries outperformed a control group. This is consistent with a series of studies by Tversky

and Kahneman (1981) showing that individuals are more willing to make risky choices to avoid losses than to achieve equivalent-sized monetary gains. Subsequent work (e.g., Rothman & Salovey, 1997), however, has suggested that the frame (gain/loss) effect may be moderated by behaviour type. For example, loss-framed messages are considered more effective for risky (i.e., detection: cancer screening attendance) versus less risky behaviours (i.e., prevention: exercise). In support, Latimer, Brawley and Bassett (2010) reported that of six studies that compared gain versus loss-framed health information regarding physical activity, two studies showed a positive main effect favouring gain-framed messages (versus zero for loss-framed messages). However, while this review provides some weak evidence favouring financial reward for performing, versus financial losses for not performing physical activity, none of the included studies actually manipulated monetary losses or gains. This is problematic because there is evidence that interventions that capitalise on participants' aversion of monetary loss, such as monetary contingency contracts (MCCs), may boost physical activity.

MCCs have the potential to be less expensive, and thus more cost-effective than standard FIs; MCCs involve depositing money which could be lost contingent on not meeting a pre-specified goal. Based on the concept of negative reinforcement, MCCs introduce an aversive stimulus (the threat of losing money) which is removed upon the performance of the desired behaviour (see Baum, 2017). Through removing the aversive stimulus, the desired behaviour becomes more likely to be performed in the future. Meta-analytic evidence suggests MCCs may aid weight loss and that participants are willing to pay into them (Sykes-Muskett, Prestwich, Lawton, & Armitage, 2015; Sykes-Muskett, Prestwich, Lawton, Meads & Armitage, 2017).

Comparisons of standard FIs and MCCs in changing health behaviours have produced mixed results. In Halpern et al. (2015) reward-based programmes versus MCCs were more

accepted, and more effective for smoking abstinence (15.7% vs. 10.2%) (both outperformed usual care: 6%). However, after accounting for differences in acceptance, MCCs were superior. These findings are complicated because the MCC was *combined* with a FI. Similarly, Donlin Washington, McMullen and Devoto (2016) required participants to make a MCC deposit of $25 but offered an additional cash reward should they achieve a behavioural goal (10,000 steps). Thus, the willingness to pay a deposit into a MCC alone (particularly for physical activity) is not clear. Moreover, while Donlin Washington et al. reported MCCs combined with a FI were as effective as FI-only, there was no control condition so firm conclusions regarding effectiveness are precluded.

In the only direct comparison of standard FI and MCCs (not combined with cash rewards) on physical activity, Burns and Rothman (2018) demonstrated cash rewards and MCCs, regardless of using a fixed or variable reward schedule, yielded similar increases in walking, each outperforming a control condition. However, they did not test whether combining the two types of FI conferred any additional benefit. Matching the amount that individuals are willing to pay into a MCC with an additional FI may yield stronger effects, at least in part, because this approach may increase the amount of money that individuals are willing to deposit into a MCC. Indeed, encouraging greater deposits - and thus making goal failure more aversive may be one (previously unstudied) mechanism by which MCC efficacy is modified.

Other mechanisms through which different types of FIs promote physical activity require examination also. Although FIs may not change autonomous motivation (e.g., Promberger & Marteau, 2013) this requires formal investigation. Instead, by providing an external reward, it may be that incentives increase controlled motivation. While controlled motivation (in line with self-determination theory) may not be most conducive to enduring physical activity, it has been demonstrated to be associated with exercise intention formation

(e.g. Willem, De Rycke & Theeboom, 2017) and may develop into autonomous motivation over time (Deci & Ryan, 2000). Controlled motivation, by increasing the importance of the behaviour, may also lead to self-monitoring of behaviour which is highly associated with achieving behavioural goals (Harkin et al., 2016).

In summary, while there is evidence showing FIs promote physical activity, there are several unknowns regarding (i) the impact of individual and combined types of FIs delivered outside of existing programmes and (ii) possible mediators including autonomous (intrinsic-related) and controlled (extrinsic-related) motivation, as well as (iii) how acceptable different types of FIs are. Without such data, it is unclear how feasible or justified testing such interventions are within longer-term studies. Here, a feasibility study was conducted at a UK-based university to examine whether interventions using different FI strategies differentially modified physical activity (pedometer-assessed daily walking).

Objectives

The present study had four objectives, collectively addressing each was necessary to inform fully-powered and longer-term trials of different types of FIs: (i) To evaluate data collection procedures: whether participants completed the relevant measures, whether levels of activity untracked by a pedometer was equivalent across groups and to assess the risk of contamination between conditions. This was important to minimise threats to the internal validity of the study; (ii) To evaluate the acceptability and suitability of the intervention and study procedures: in particular, participant retention, responses to quantitative and qualitative measures of acceptability and to identify the proportion of participants willing to pay into a MCC and whether this influenced through the offer of additional rewards. This was important because widespread acceptability of an (effective) intervention may maximise intervention effects; (iii) To present a preliminary evaluation of effectiveness: specifically, whether different types of FIs lead to different levels of physical activity change, as well as a

preliminary estimate of the relative strategy cost effectiveness. This may have indicated, notwithstanding potential (particularly Type-II) errors, whether further testing of these interventions is warranted; (iv) To identify potential mechanisms (intention, intrinsic/autonomous motivation, extrinsic/controlled motivation, amotivation, frequency of self-monitoring, deposit size) which could explain why particular types of FIs could be effective. Understanding potential mechanisms is useful; serving to indicate how particular behaviour change effects can be maximized, for example by indicating whether different behaviour change techniques targeting other important mechanisms (unaffected by the intervention) are warranted (Prestwich, Kenworthy, & Conner, 2017).

Although the study was not a-priori statistically powered, it was predicted that all types of FIs would promote steps, motivation (increased controlled motivation and behavioural intentions, reduced amotivation) and progress monitoring relative to the comparison conditions. It was also predicted that participants in a combined MCC and standard FI condition would deposit more money on average than participants in a MCC-only condition. Differences in controlled motivation, amotivation, progress monitoring and deposit size were anticipated to mediate the effects of the interventions on steps. Whether the interventions served to reduce autonomous motivation or to have no negative effect on autonomous motivation was explored.

Method

Participants

Eighty participants were recruited. Participants were identified via email responses to poster/email advertisements and screened via an online eligibility questionnaire. Included participants were (i) aged 18-65 and (ii) classified as having low/moderate levels of physical activity (IPAQ-Short Form category one and two, respectively). Excluded were those (i) at

risk of negative health impacts following increased activity (based on the Physical Activity Readiness Questionnaire), (ii) with insufficient English language skills.

Design

A factorial (+ 1 condition), five-arm, parallel groups, randomised controlled trial was conducted from June 2016 to October 2017. Twelve participants per condition was the sample size aim, though to account for attrition, 16 were recruited. Allocations (1:1:1:1:1) were specified and listed in a randomly generated order. Each consecutive participant recruited into the study was then assigned to the corresponding condition next in the list. Participants were masked to conditions, experimenters (including data analysts) were not.

Participants attended the laboratory three times; *time 1* on day 0, *time 2* on day 8 and *time 3* on day 23. At *time 2* all but one condition was set a goal of increasing their daily steps over a period of two weeks (between *time 2* and *3*). Goals were based on the median number of steps each participant achieved between *time 1* and *2*; participants walking ≤8000 median daily steps were given a goal of 10,000 daily steps, participants walking >8000 steps were given a goal of 12,000. These were chosen to maximally enhance goal achievement in line with Locke and Latham (2002); they were *specific* and *difficult* (at minimum 2000 steps more than participants' current average) - but not too difficult (for participants averaging fewer than 8000 steps, 12,000 steps was considered likely to undermine goal commitment through insufficient self-efficacy).

Participants could receive a payment of £25 worth of shopping vouchers for completing relevant measures (£15 at *time 2*, £10 at *time 3*), though allocation to certain conditions gave participants the chance of earning more (or less).

Condition 1. Financial incentive (FI). Study completers could earn £25 or £40; participants earned £15 in vouchers (in addition to the standard £15 at *time 2* and £10 at *time 3*) if they met their set goal on 12 of the 14 days (between *time 2* and *3*).

Condition 2. Monetary contingency contract (MCC). Completers could earn between £10 and £25; participants had the option of putting £0, £5, £10 or all of their *time 2* £15 voucher aside. Failing to meet their goal on 12 out of 14 days forfeited whatever they put aside (though still received the standard *time 3* £10 voucher). Participants were given a 24-hour period to withdraw their decision to put vouchers aside and retrieve the full value. Note that the option to put zero money aside was given to participants (i) due to ethical reasons and (ii) to assess real-world engagement with the paradigm (i.e., ecological validity).

Condition 3. Monetary contingency and financial incentive (MCC+FI). Completers could earn between £10 and £40; as above, participants could put some of the voucher aside (with a 24-hour window), and forfeited that amount if they failed to achieve their goal. However, if they achieved their goal on 12 of 14 days they received double the amount they put aside at *time 2*, as well as the standard £10 participation voucher at *time 3*.

Condition 4. Control with goal (CG). Completers earned £25; participants were given a goal but were offered no financial incentive to achieve it, receiving only the standard participation vouchers.

Condition 5. Control without goal (CWOG*).* Completers earned £25; participants were given no step goal and received participation vouchers only.

To investigate effects on potential psychological mediators of behaviour change, behavioural intentions and three facets of motivation (controlled, autonomous and amotivation) were assessed at *time 1, time 2* and once at *time 3* Additionally, to investigate whether monitoring was related to daily step increases, participants noted the frequency which they had monitored their daily step count *per se* and the frequency they monitored their step count to judge goal progress, at *time 3*. At *time 3* participants also completed measures of intervention acceptability, contamination and were encouraged to add any further comments on their experience with the study.

Measures

Daily steps. Yamax/Yamasa PZ-271 pedometers assessed steps. The range has good evidence of reliability and validity (e.g., Crouter, Schneider, Karabulut & Bassett, 2003). Participants were instructed to attach the pedometer (which reset at 2am each morning) to their hip upon morning awakening until bedtime. Participants could monitor their steps *ad libitum*.

To adjust for hours of walking not recorded through pedometer non-compliance, participants were asked to complete a non-wear time questionnaire at *time 2* and *3*. Daily steps were therefore assessed (i) adjusted for non-wear time: (total *n* steps over *n* days/total *n* of waking hours in *n* days – *n* hours pedometer was not worn)*16, where 16 hours was assumed to be the typical time spent awake by the participants, and (ii) unadjusted: total *n* steps over *n* days/*n* of days worn.

To capture physical activity not obtained by the pedometer, those stating they had not worn the pedometer for ≥1 hours were required to describe any exercise and walking journeys above 10 minutes they engaged in without the pedometer .Participants were categorized as doing vs. not doing at least some moderate physical activity (i.e. brisk/fast walking or any exercise) while not wearing the pedometer at *time 3*.

Motivation. The 15-item Treatment Self-Regulation Questionnaire (TSRQ) (Levesque et al. 2006) assessed autonomous motivation, controlled motivation (six items each) and amotivation (three items). Responses were rated from 1 (not at all) to 7 (very true). Internal consistency for each subscale was good (see supplementary Table S1), apart from the amotivation subscale at *time 1* and *3*. Dropping one item considerably improved consistency (e.g. *time 3* Mcdonald's ω = .68 to .78), thus a truncated version was used.

Behavioural intentions. Participants rated four items (due to differing step goals) related to their *n* = 8000, 10,000, 12,000 and 14,000 daily step intentions ("*I intend to take at*

least ____ over the next 14 days") from 1 (strongly disagree) to 7 (strongly agree). Internal consistency was excellent (Table S1).

Self-monitoring & progress monitoring. Participants reported how frequently they had checked their step count in the previous 14 days, from 1 ("*never*") to 7 ("*several times a day*"). Progress monitoring, the frequency participants checked how close they were to achieving their step goal was assessed likewise.

Intervention acceptability. Participants rated eight statements assessing the degree they found the intervention, for example, "*helpful*" from 1 (strongly disagree) to 5 (strongly agree). Those in the MCC, FI and MCC+FI conditions were asked to rate a further two statements related to vouchers (e.g. "*I liked earning vouchers*"). Written participant feedback about the study was also encouraged to provide insight into improving participant experience.

Contamination check. Participants reported whether they knew of any other participants, and if applicable (i) how many, (ii) whether they discussed the study with them, and (iii) whether they knew what another participant was required to do. Those who stated they knew what others had to do were asked to write a description of the requirements/rewards of the other participant(s).

Procedure

Day 0: Time 1 (baseline). After providing informed consent, participants completed behavioural intention/motivation measures and were fitted with a pedometer.

Day 8: Time 2 (pre-intervention). Participants completed measures of behavioural intentions, motivation and pedometer non-wear-time before being provided with condition-specific instructions. The pedometer was retrieved while participants completed the measures, and step counts where taken from the memory.

Day 23: Time 3 (post-intervention). Participants completed measures of intentions, motivation, pedometer non-wear-time, self/progress monitoring, intervention acceptability,

304 contamination and offered any comments regarding their experience. The pedometer was
305 retrieved while participants completed the measures, and step counts taken from the memory.
306 Depending on their condition, participants received their respective voucher rewards.
307 Participants were debriefed via email.

308 **Analysis strategy**

309 Analyses were conducted using JASP and Jamovi software. Throughout, alpha was
310 set at .05 (two tailed). Between subjects AN(C)OVA was used for analyses with one DV and
311 one IV with >2 levels (e.g., step counts between conditions), MAN(C)OVA when >1 DV
312 (e.g., randomisation/attrition checks), and a within-between ANCOVA with repeated
313 measures (motivation and intentions between conditions over time). Chi-Squared tests
314 assessed differences when IVs and DVs were categorical (e.g. contamination between
315 conditions), with logistic regression further investigating condition goal achievement and
316 multinomial regression investigating deposit amount between MCC versus MCC+FI
317 conditions. Exploratory bootstrapped (1000 resamples) mediation models formally assessed
318 potential mechanisms for the effect of condition on step counts. This was done only when
319 there was evidence that the potential mediating variable (e.g. intentions or deposit size) was
320 significantly affected by condition allocation and that it predicted step counts. It was
321 acknowledged that with a feasibility study sample size significance testing may be prone to
322 Type-II error. To partially offset this in primary analyses, Bayes factor (BF_{10}) equivalents of
323 all the above frequentist tests were used - thus evidence for null effects could be quantified.
324 Here, BF_{10} 1-3 = anecdotal (+), 3-10 = moderate (++), 10+ = strong (+++) evidence in favour
325 of the alternative over the null hypothesis, and 0.33-1 = anecdotal (=), 0.10-0.33 = moderate
326 (==) and 0.03-0.10 = strong (===) evidence for the null over the alternative hypothesis (c.f.
327 Quintana & Williams, 2018; Wagenmakers et al. 2018). Sensitivity analyses assessed the
328 results with set priors of r = 0.3, 0.5 and 0.7 (or equivalent). The BF_{10} interpretation (e.g.

anecdotal versus moderate) did not meaningfully change between priors for any primary analysis (results using JASP defaults were reported c.f. Quintana & Williams, 2018; Wagenmakers et al. 2018).

Bayesian post-hoc analyses were conducted using standard *t*-test equivalents with the JASP default Cauchy prior (0.71) (Wagenmakers et al. 2018). Post-hoc tests conducted within the frequentist tradition were adjusted for multiple comparisons to mitigate Type-I error inflation. Subscripts proceeding *p* value labels denote the adjustment method used (Tukey or Holm methods, depending on the analysis).

Condition cost-effectiveness was determined by dividing the total expenditure per condition by the number of successful goal achievements per condition (assuming that the CWOG condition had a goal based on their baseline steps); generating a measure of expenditure per goal achievement. Each condition was compared against the highest achieving condition (MCC+FI) by dividing its expenditure per goal achievement by the corresponding MCC+FI figure (relative expenditures are reported as percentages).

Results

Sixteen participants per condition were recruited (n = 80). Participants were mostly female (90%) and were an average age of 34.2 (SD = 11.6) years. Thirty percent were students and 66.3% were university employees (see Table 1).

Objective 1: Assessing data collection procedures

Missing data. Data were well captured across measures. For all motivation/monitoring measures at all time-points, data were missing for 4 participants (5%) or fewer. Pre-intervention average steps data were missing for 2 (2.5%) participants, post-intervention 6 (7.5%). Pedometer non-wear time data were missing for 5 (6.3%) or fewer participants across time-points. Six (7.5%) participants had missing data for the first contamination item. Acceptability ratings had 5 (6.3%) or fewer participants missing data.

For demographic data, age and student status was missing for 3 (3.8%) participants, sex for zero.

Validity checks. Between-condition proportions of participants reporting any physical activity deemed at or above METs 'moderate' (brisk/fast walking and exercise/dancing) while not wearing the pedometer at *time 3* were investigated. There were no significant differences between those involving MCCs versus others, and no differences between individual conditions. However, there was a risk that conditions with FIs would be incentivised not only to walk more, but to wear their pedometer more often whilst doing physical activities (increasing their chances of goal achievement and monetary reward). This was found to be the case, $\chi^2(1) = 4.49$, $p = .03$, after conditions including financial (MCC+FI, FI) versus no financial incentive (MCC, CG, CWOG) were compared.

There were no significant differences between conditions in the (i) proportion of participants knowing another participant, (ii) total number known, and (iii) proportion of participants reporting that they discussed the study with another participant. However, there was a significant between-condition difference in the number of participants reporting that they knew what another participant had to do in the study ($p = .05$). Inspection of the further comments section on the contamination measure revealed that two participants were only exposed to others in the same condition. After recoding these participants as non-exposed, between-condition differences remained as above, $\chi^2(4) = 10.07$, $p = .04$. No participants in the MCC+FI, FI or CG conditions were exposed to other manipulations; five CWOG and one MCC participant were.

The potential impact of participants being more or less likely across conditions to do physical activity while not wearing the pedometer was taken into account by considering pedometer steps adjusted for non-wear time. Sensitivity analyses were also conducted to examine whether effects changed (i) excluding individuals who engaged in moderate or

above physical activity without the pedometers and (ii) excluding individuals who had become aware of what participants in different conditions were required to do (see Objective 3).

Objective 2: Acceptability and suitability of the intervention and study procedures

Quantitative feedback. There were no significant differences between groups on the acceptability related items completed by all conditions, or the two additional voucher themed items completed by those in the FI, MCC, and MCC+FI conditions. All mean scores were greater than 4 (see supplementary Table S2) (indicating above middle ratings, thus seemingly high intervention acceptability), with exceptions on the '*fun*' item (FI-only = 3.40, MCC-only = 3.88, CG = 3.71), and in the MCC-only condition, the item asking whether vouchers increased their walking (3.38; likely because not all participants deposited their vouchers).

Qualitative feedback. Various informative participant comments were made (see supplementary Table S3). Briefly, nine participants suggested that an incentive was highly motivating, though one suggested this may only be true in the short term. Having a pedometer and thus being able to monitor step counts was considered useful: "[it] *made me think more about walking"*, "[it] *was eye opening* […] *helpful in raising awareness*". Though some participants did have reservations about the pedometer: (e.g. "…*it was not easy to use, too bulky*").

Retention. Seventy-six (95%) participants completed the study. Two dropped out pre-intervention (*time 2*), two post-intervention (between *time 2* and *3*) (see supplementary *Figure S1*). Drop outs and completers did not differ significantly by condition or any other measured variable (e.g. age, baseline steps, baseline intentions).

MCC adherence. No participants in any condition asked for their vouchers back. There was a significant omnibus effect, and moderate evidence for, between-condition differences (MCC+FI versus MCC-only) in deposit amount, $\chi^2(3) = 7.87$, $p = .05$, $BF_{10} =$

4.25 (++) (see supplementary *Figure S2*). Those in the MCC+FI condition, compared to those in the MCC-only condition, were significantly more likely to deposit £15 versus £0 (OR = 3, $p = .02$), but equally likely, versus £0, to deposit £10 ($p = .95$) and £5 ($p = .24$). Multinomial model post-hoc tests showed that those in the MCC-only condition versus MCC+FI condition were 43% more likely to deposit £0 ($p_{Holm} = .03$), 35% less likely to deposit £15 ($p_{Holm} = .08$) and 1% ($p_{Holm} = .95$) and 7% ($p_{Holm} = .34$) less likely to deposit £5 and £10 respectively.

Objective 3: Preliminary evaluation of effectiveness

Age differed significantly across conditions, $F(4, 70) = 2.61$, $p = .04$, FI participants were marginally younger than MCC participants ($M = 28.67$ years, $SD = 8.83$ versus $M = 39.53$, $SD = 12.19$, $p_{Tukey} = .06$). Sex marginally varied across conditions, $\chi^2(4) = 9.17$, $p = .06$; of all 8 men in the study, half were in the CG condition, and zero were in the MCC-only or MCC+FI conditions. There were no other pre-intervention between-condition differences. Subsequent analyses controlled for age (note that results were similar when controlling for age and sex and are available on request).

Adjusted and unadjusted steps. Controlling for age and baseline steps, there were significant between-condition differences on adjusted, $F(4, 64) = 3.70$, $p < .01$, $\omega^2 = .08$, $\eta^2 = .12$, $\eta^2p = .19$ and unadjusted steps, $F(4, 64) = 4.06$, $p < .01$, $\omega^2 = .09$, $\eta^2 = .12$, $\eta^2p = .20$, with moderate-sized Bayes factors (adjusted: $BF_{10} = 4.88$, ++; unadjusted: $BF_{10} = 7.56$, ++). MCC+FI participants achieved marginally more adjusted steps than CG participants, adjusted: $t(64) = 2.67$, $p_{Tukey} = .07$, mean difference = 2772 steps, SE = 1038, $BF_{10} = 1.32$ (+), though significantly more than both the MCC-only, $t(64) = 2.94$, $p_{Tukey} = .04$, mean difference = 2909, SE = 990, $BF_{10} = 0.81$ (=) and CWOG participants, $t(64) = 3.03$, $p_{Tukey} = .03$, mean difference = 2897, SE = 956, $BF_{10} = 3.32$ (++). MCC+FI also achieved significantly more unadjusted steps than both controls See Table 2 for an overview of the results.

Robustness checks. After reanalysis of steps outcomes following the removal of 'contaminated' participants, all main effects remained significant, and evidence against the null remained moderate. However, post-hoc tests showed that the difference between MCC+FI and the CG condition were mostly only marginally significant, with anecdotal Bayes factors. After removing participants who engaged in exercise without the pedometer, all main effects remained significant and Bayes factors favoured the alternative hypothesis. The only significant post-hoc comparison that remained significant was more steps in MCC+FI versus CWOG condition (see Table 2).

Goal achievement. Goal achievement was significantly different between conditions, $\chi^2(4) = 10.13$ $p = .04$, with a moderate-sized Bayes factor ($BF_{10} = 4.25$, ++) (++ remained with prior concentrations of 1, 3, 5 and 7). Those in the MCC+FI had significantly higher odds of goal achievement versus the CWOG condition (reference condition) (odds ratio (OR) = 11.84, $p = .01$). Other conditions had non-significantly higher odds (all $p > .13$, OR: FI = 4.25, CG = 2.12, MCC = 2.49) (see supplementary *Figure S3*).

Cost effectiveness. Per participant pay-outs (mean, SD; including money paid up to the point of drop-out) per condition were: MCC+FI = £515 (£32.2, £10.5), MCC = £380 (£23.8, £3.9), FI = £440 (£27.5, £10.8), CG = £375 (£23.4, £6.3) and CWOG = £390 (£24.4, £2.5). Cost per successful goal completion was: MCC+FI = £46.82, MCC = £76, FI = £73.33, CG = £93.75 and CWOG = £195. Compared to the highest goal achieving condition (MCC+FI, 11/16 participants), others were thus 62% (MCC, 5/16), 57% (FI, 6/16), 100% (CG, 4/16) % and 317% (CWOG, 2/16) more expensive per goal achiever.

Objective 4: Identifying potential intervention mechanisms

Psychological variables. There were no significant main or time *x* condition effects (and moderate to strong Bayes factor supporting null effects) on intentions, autonomous and

controlled motivation, or amotivation (supplementary Table S1). The same was true of self- or progress- monitoring.

Deposit size. Given those in the MCC+FI versus MCC-only condition were significantly more likely to deposit £15 versus £0 (see Objective 2), deposit size represented a viable mechanism explaining between-condition step count differences (see Objective 3). Furthermore, there was evidence to suggest between-deposit (£0, £5, £10, £15) differences in adjusted $F(3, 23) = 3.28$, $p = .04$, $\omega^2 = .16$, $\eta^2 = .23$, $\eta^2p = .30$, $BF_{10} = 4.48$ (++) and unadjusted steps $F(3, 23)= 2.50$, $p = .09$, $\omega^2 = .13$, $\eta^2 = .22$, $\eta^2p = .25$, $BF_{10} = 2.57$ (+). For the former, a £15 deposit led to significantly more adjusted steps than £0, $t(23) = 3.11$, $p_{Tukey} = .02$, mean difference = 3721 steps, SE = 1197, $BF_{10} = 10.43$ (+++).

In formal mediation analyses, step count by condition (MCC or MCC+FI) analyses (the direct effects) were not significant (adjusted steps: $p = .43$; unadjusted steps: $p = .58$). All other path estimates were significant; deposit size by condition ($B = 6.3$, 95% CI: 1.7 - 10.7, $p = .007$) and steps by deposit size (adjusted: $B = 247.7$, 95% CI: 4.2 - 451.8, $p = .03$; unadjusted: $B = 259.5$, 95% CI: 52 - 435.9, $p = .009$). While the overall indirect effects were not statistically significant (adjusted: $B = 1549$, 95% CI: -56.3 - 3850, $p = .12$; unadjusted: $B = 1624$, 95% CI: 167.7 - 3822, $p = .09$), there was some evidence of mediation; 58.1% (adjusted steps) and 70.5% (unadjusted steps) of the total effect in each analysis was explained by indirect effects.

Discussion

This is the first study to separate out the individual and combined effects of MCCs and standard FIs on physical activity. All interventions were well accepted by participants; retention was high; measures well adhered to; and there were some interesting (and robust) preliminary results. There was moderate evidence from Bayes factors and frequentist analyses to suggest between-condition differences on steps over a two-week period. Those in

the MCC+FI condition achieved more steps than both control groups, and there was some evidence of superiority over MCC-only. There was however no evidence of superiority of a combined intervention above standard FIs, and no significant other between-condition differences. MCC+FI participants had the highest level of goal attainment, indeed, this condition was by far the most cost-effective intervention as assessed by cost per participant goal achieved. MCCs+FIs led to more full £15 deposits than the MCC-only condition, and MCC-only led to more £0 deposits than the MCC+FI condition; notable considering £15 versus £0 deposits led to more steps. There was no statistical evidence of between-condition differences on three facets of motivation, intentions or self-monitoring.

The combination of standard FIs plus MCCs appeared highly promising. Though not all statistically significant, the finding that differences between the MCC+FI and other conditions were all (and often well above) 2000 steps is noteworthy; studies have shown that such an increase can lead to tangible health benefits (Dasgupta et al. 2017; Hajna, Ross & Dasgupta, 2017; Ewald, Oldmeadow & Attia, 2017) meaning, if stable and replicable, it may be a clinically significant difference. Similarly, the cost effectiveness of MCC+FI relative to other conditions in terms of goal achievement was impressive.

Those in the FI conditions were less likely than those in the non-FI conditions to do some form of (at least) moderate physical activity while not wearing the pedometer. While this difference did not explain the effect of the interventions on step counts, it is clearly a potential confound which researchers should pay close attention to in future trials. There was no large and notable difference, through frequentist or Bayesian analysis, in the number of steps achieved following FI-only or MCC-only interventions compared to the control group, which is contrary to previous studies (e.g. Burns & Rothman, 2018; Finkelstein et al. 2008; Patel et al. 2016). It should be noted, however, that the small pilot sample size may have introduced imprecision (Bayes factors) and lack of power.

The non-significant mediation effects of deposit size on MCC vs. MCC+FI and steps may also have been driven by imprecise estimates given the relatively small feasibility study sample size. Indeed, it is acknowledged that testing for mechanisms in small samples may be problematic, and that larger replication is needed. As such, given observed differences between MCC+FI and MCC-only conditions in both steps and deposit amounts (and large indirect effect sizes), the relative effects between MCC+FI and MCC may be still explained by deposit amounts. Higher deposits may lead to increases in loss aversion, as well as an increased exposure to the synergistic influence of both positive and negative reinforcement. In effect, the significance of an added financial incentive could be jointly explained by its ability to increase both perceived value and adherence to the MCC paradigm. Previous studies which have had deposits as a participant *requirement* (i.e. 100% adherence) as well as studies where participants would stand to lose their own money (increased negative reinforcement), may explain why they saw increased MCC-only effects (e.g., Donlin Washingon, Mcmullen & Devoto, 2016). For the present study, the participants own money was not risked and zero deposit and was allowed. This may represent a more accurate picture of real-world effectiveness of, and adherence to MCCs. For instance, if MCCs were offered in a healthcare service, it could never be through coercion – patients would always have the option of opting out. Confidence in the ecological validity of the study was boosted further by the fact that no participants requested their lost earnings back. This suggests that participants respected the rules of the paradigm and treated it as 'real'.

For the standard FI-only condition, the value of the reward *alone* may not have been enough of a positive reinforcer to overcome non-trivial costs (evidenced by some participants suggesting that the goals were quite hard for them to achieve). Indeed, larger incentives *per se* (Mitchell et al. 2013), as well as the *perceived* value of incentives (Burns & Rothman, 2018) have been shown to produce larger effects on physical activity. Among a host of other

environmental and psychological variables, this cost-benefit analysis is likely to be associated with personal income (which may be an interesting variable to investigate in future research). Additionally, individuals, if given the choice, may choose a lower sum of money instantly, than a larger sum in the future; here, the week delay in reward may have been vulnerable to this delay discounting effect (Odum, 2011), reducing the effective incentive value.

As demonstrated by between-condition ratings of intervention acceptability, low attrition (equivalent across conditions), and mostly positive qualitative feedback, the present study demonstrated that MCC-only, FI-only and combined FI interventions are similarly accepted by participants (and similarly acceptable to minimal interventions delivered within the control conditions). The present results also showed no impact of interventions on any psychological variables including that FIs had no deleterious effect on autonomous motivation (consistent with Promberger and Marteau, 2013). The present study cannot however determine the psychological effects of FIs over the long term; it may be that sustained external reward only undermines intrinsic motivation when the individual comes to develop a *reliance* on this for a source of motivation at the expense of their own, internal motivation.

As with previous analyses, more power may be able to detect, what may be subtle changes in all measured psychological-related variables - though this is unlikely given the observed Bayes factors favouring the null. One key reason for null results may be that because incentives were only offered over two weeks, this may have been too short of a time frame to truly observe changes in, for example, controlled motivation. Similarly, it is acknowledged that participants, overall, had relatively high levels of certain facets of motivation (e.g. intentions) at baseline - so the degree of potential modifiability may have been small. This latter factor may have impacted the primary results; if participants had lower baseline motivation, speculatively, the intervention effects may have been increased.

Nevertheless, regarding autonomous motivation, this study was interested in examining whether certain types of financial rewards undermine (rather than enhance) motivation. From this perspective, the relatively high levels of autonomous motivation was not particularly problematic.

There were several study limitations. First, there was a relatively high degree of contamination, which may be an inherent issue with single organisation-based studies. Mindful of this, contamination was assessed between conditions, and while the between condition proportions of contamination *per se* were negligible, there was a significant difference in the number of participants knowing what another participant had to do. Although main effects were unaffected, reanalyses excluding these participants led to some post-hoc comparisons to become non-significant. Given the contamination risk, multi-centre trials may be particularly helpful for future trials. Second, females were over-represented in the sample (9:1) which is common in physical activity focused incentive studies (e.g., Sykes-Muskett et al. 2015; Burns & Rothman, 2018) and may reflect between gender differences in physical activity/weight loss attitudes (Azevedo, Araujo, Reichert, Siqueria, da Silva & Hallal, 2007; Burton, Walsh & Brown, 2008). There is however little evidence of between-gender efficacy of physical activity interventions (Williams, Wood, Collins & Callister, 2015). Third, the use of vouchers in this study may not have had equal incentive salience as cash – intrinsically (e.g. Raghubir & Srivastava, 2008) but also because the vouchers limited the number of products/service available for purchase. Fourth, while the cost effectiveness of the combined condition relative to other conditions was impressive, the study ran over a limited timeframe; how this may have changed over time is unknown.

This study provided initial evidence that adding standard FI to MCC may boost their efficacy, possibly through increases in both adherence to and salience of MCCs. The study did not find any evidence that autonomous motivation would be undermined by external

reward, and that different types of FIs were equally accepted by participants. Future trials should be mindful of limitations described above and try to replicate findings using a larger sample size, longer follow-up period, and in the knowledge that contamination (single organisation research) and pedometer adherence (between conditions incentivised to wear it, and those not so) may be potential confounds.

References

Acland, D., & Levy, M. R. (2015). Naiveté, projection bias, and habit formation in gym attendance. *Management Science*, *61*(1), 146-160. doi: 10.1287/mnsc.2014.2091.

Azevedo, M. R., Araújo, C. L. P., Reichert, F. F., Siqueira, F. V., da Silva, M. C., & Hallal, P. C. (2007). Gender differences in leisure-time physical activity. *International Journal of Public Health*, *52*(1), 8. doi: 10.1007/s00038-006-5062-1.

Barte, J. C., & Wendel-Vos, G. W. (2017). A systematic review of financial incentives for physical activity: the effects on physical activity and related outcomes. *Behavioral Medicine*, *43*(2), 79-90. doi: 10.1080/08964289.2015.1074880.

British Heart Foundation. (2017). Physical inactivity report 2017. Retrieved 18th May from https://www.bhf.org.uk/publications/statistics/physical-inactivity-report-2017.

Baum, W. M. (2017). *Understanding behaviorism: Behavior, culture, and evolution*. John Wiley & Sons.

Burns, R. J., & Rothman, A. J. (2018). Comparing Types of Financial Incentives to Promote Walking: An Experimental Test. *Applied Psychology: Health and Well-Being*. doi: 10.1111/aphw.12126.

Burton, N. W., Walsh, A., & Brown, W. J. (2008). It just doesn't speak to me: mid-aged men's reactions to'10,000 Steps a Day'. *Health Promotion Journal of Australia*, *19*(1), 52-59.

Carlson, S. A., Fulton, J. E., Pratt, M., Yang, Z., & Adams, E. K. (2015). Inadequate Physical

Activity and Health Care Expenditures in the United States. *Progress in Cardiovascular Diseases, 57*(4):315-323.

Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica*, *77*(3), 909-931. doi: 10.3982/ECTA7416.

Courneya, K. S., Estabrooks, P. A., & Nigg, C. R. (1997). A simple reinforcement strategy for increasing attendance at a fitness facility. *Health Education & Behavior*, *24*(6), 708-715. doi: 10.1177/109019819702400606.

Crouter, S., Schneider, P., Karabulut, M., Bassett, D. (2003). Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Medicine and Science in Sports and Exercise, 35*, 1455-1460.

Dasgupta, K., Rosenberg, E., Joseph, L., Cooke, A. B., Trudeau, L., Bacon, S. L., ... & SMARTER Trial Group. (2017). Physician step prescription and monitoring to improve ARTERial health (SMARTER): a randomized controlled trial in patients with type 2 diabetes and hypertension. *Diabetes, Obesity and Metabolism*, *19*(5), 695-704. doi: 10.1111/dom.12874.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227-268.

Ding, D., Lawson, K. D., Kolbe-Alexander, T. L., Finkelstein, E. A., Katzmarzyk, P. T., Van Mechelen, W., ... & Lancet Physical Activity Series 2 Executive Committee. (2016). The economic burden of physical inactivity: a global analysis of major non-communicable diseases. *The Lancet*, *388*(10051), 1311-1324.

Donlin Washington, W., McMullen, D., & Devoto, A. (2016). A matched deposit contract

628 intervention to increase physical activity in underactive and sedentary adults.
629 *Translational Issues in Psychological Science*, *2*(2), 101. doi: 10.1037/tps0000069.
630 Dumith, S. C., Hallal, P. C., Reis, R. S., & Kohl III, H. W. (2011). Worldwide prevalence of
631 physical inactivity and its association with human development index in 76 countries.
632 *Preventive Medicine, 53*(1-2), 24-28.
633 Ewald, B. D., Oldmeadow, C., & Attia, J. R. (2017). Daily step count and the need for
634 hospital care in subsequent years in a community-based sample of older Australians.
635 *The Medical Journal of Australia*, *206*(3), 126-130. doi: 10.5694/mja16.00640.
636 Finkelstein, E. A., Brown, D. S., Brown, D. R., & Buchner, D. M. (2008). A randomized
637 study of financial incentives to increase physical activity among sedentary older
638 adults. *Preventive Medicine*, *47*(2), 182-187. doi: 10.1016/j.ypmed.2008.05.002.
639 Giles, E. L., Becker, F., Ternent, L., Sniehotta, F. F., McColl, E., & Adams, J. (2016).
640 Acceptability of financial incentives for health behaviours: a discrete choice
641 experiment. *PloS One*, *11*(6), e0157403. doi: 10.1371/journal.pone.0157403.
642 González, K., Fuentes, J., & Márquez, J. L. (2017). Physical inactivity, sedentary behavior
643 and chronic diseases. *Korean Journal of Family Medicine, 38*(3), 111-115.
644 Hajna, S., Ross, N. A., & Dasgupta, K. (2017). Steps, moderate-to-vigorous physical activity,
645 and cardiometabolic profiles. *Preventive Medicine*, *107*, 9-74. doi:
646 10.1016/j.ypmed.2017.11.007.
647 Halpern, S. D., French, B., Small, D. S., Saulsgiver, K., Harhay, M. O., Audrain-McGovern,
648 J., ... & Volpp, K. G. (2015). Randomized trial of four financial-incentive programs
649 for smoking cessation. *New England Journal of Medicine*, *372*(22), 2108-2117. doi:
650 10.1097/01.ogx.0000470840.40563.9a.
651 Harkin, B., Webb, T. L., Chang, B. P., Prestwich, A., Conner, M., Kellar, I., ... & Sheeran, P.
652 (2016). Does monitoring goal progress promote goal attainment? A meta-analysis of

653 the experimental evidence. *Psychological bulletin*, *142*(2), 198.
654 Hou, N., Popkin, B. M., Jacobs Jr, D. R., Song, Y., Guilkey, D. K., He, K., ... & Gordon-
655 Larsen, P. (2011). Longitudinal trends in gasoline price and physical activity: the
656 CARDIA study. *Preventive Medicine*, *52*(5), 365-369. doi:
657 10.1016/j.ypmed.2011.02.007.
658 Johnston, M., & Sniehotta, F. (2010). Financial incentives to change patient behaviour.
659 *Journal of Health Services Research & Policy, 15*(3), 131-132. doi:
660 10.1258/jhsrp.2010.010048.
661 Latimer, A. E., Brawley, L. R., & Bassett, R. L. (2010). A systematic review of three
662 approaches for constructing physical activity messages: what messages work and
663 what improvements are needed?. *International Journal of Behavioral Nutrition and*
664 *Physical Activity*, *7*(1), 36.
665 Leonard, T., & Shuval, K. (2017). Behavioral economic tools for promotion of physical
666 activity. In *Behavioral Economics and Healthy Behaviors: Key Concepts and Current*
667 *Research*. Routledge London and New York.
668 Levesque, C. S., Williams, G. C., Elliot, D., Pickering, M. A., Bodenhamer, B., & Finley, P.
669 J. (2006). Validating the theoretical structure of the Treatment Self-Regulation
670 Questionnaire (TSRQ) across three different health behaviors. *Health Education*
671 *Research*, *22*(5), 691-702. 10.1093/her/cyl148.
672 Locke, E.A. & Latham, G.P. (2002). Building a practically useful theory of goal setting and
673 task motivation: a 35-year odyssey. *American Psychologist*, *57*(9), 705-717.
674 Mitchell, M. S., Goodman, J. M., Alter, D. A., John, L. K., Oh, P. I., Pakosh, M. T., &
675 Faulkner, G. E. (2013). Financial incentives for exercise adherence in adults:
676 systematic review and meta-analysis. *American Journal of Preventive Medicine*,
677 *45*(5), 658-667. doi: 10.1016/j.amepre.2013.06.017.

678 Odum, A. L. (2011). Delay discounting: I'm ak, you're ak. *Journal of the Experimental*
679 *Analysis of Behavior*, *96*(3), 427-439. doi: 10.1901/jeab.2011.96-423.
680 Patel, M. S., Asch, D. A., Rosin, R., Small, D. S., Bellamy, S. L., Heuer, J., ... & Wesby, L.
681 (2016). Framing financial incentives to increase physical activity among overweight
682 and obese adults: a randomized, controlled trial. *Annals of Internal Medicine*, *164*(6),
683 385-394. doi: 10.7326/M15-1635.
684 Pope, L., & Harvey-Berino, J. (2013). Burn and earn: a randomized controlled trial
685 incentivizing exercise during fall semester for college first-year students. *Preventive*
686 *Medicine*, *56*(3-4), 197-201. doi: 10.1016/j.ypmed.2012.12.020.
687 Prestwich, A., Kenworthy, J., & Conner, M. (2017). *Health behavior change: Theories,*
688 *methods and interventions*. Routledge.
689 Promberger, M., & Marteau, T. M. (2013). When do financial incentives reduce intrinsic
690 motivation? comparing behaviors studied in psychological and economic literatures.
691 *Health Psychology*, *32*(9), 950. doi: 10.1037/a0032727.
692 Promberger, M., Brown, R. C., Ashcroft, R. E., & Marteau, T. M. (2011). Acceptability of
693 financial incentives to improve health outcomes in UK and US samples. *Journal of*
694 *Medical Ethics*, *37*(11), 682-687. doi: 10.1136/jme.2010.039347.
695 Promberger, M., Dolan, P., & Marteau, T. M. (2012). "Pay them if it works": Discrete choice
696 experiments on the acceptability of financial incentives to change health related
697 behaviour. *Social Science & Medicine*, *75*(12), 2509-2514. doi:
698 10.1016/j.socscimed.2012.09.033.
699 Quintana, D., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis
700 significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry,*
701 *18*(1), 178. doi: 10.1186/s12888-018-1761-4.
702 Raghubir, P., & Srivastava, J. (2008). Monopoly money: The effect of payment coupling and

form on spending behavior. *Journal of Experimental Psychology: Applied*, *14*(3), 213.

Rothman, A. J. & Salovey, P. (1997). Shaping perceptions to motivate health behaviour: The role of message framing. *Psychological Bulletin, 121*(1), 3.

Shin, D. W., Yun, J. M., Shin, J. H., Kwon, H., Min, H. Y., Joh, H. K., ... & Cho, B. (2017). Enhancing physical activity and reducing obesity through smartcare and financial incentives: a pilot randomized trial. *Obesity*, *25*(2), 302-310. doi: 10.1002/oby.21731.

Sykes-Muskett, B. J., Prestwich, A., Lawton, R. J., & Armitage, C. J. (2015). The utility of monetary contingency contracts for weight loss: a systematic review and meta-analysis. *Health Psychology Review*, *9*(4), 434-451. doi: 10.1080/17437199.2015.1030685.

Sykes-Muskett, B. J., Prestwich, A., Lawton, R. J., Meads, D. M., & Armitage, C. J. (2017). Exploration of likely engagement with Monetary Contingency Contracts for weight loss: a questionnaire study. *Psychology, Health & Medicine*, *22*(10), 1278-1283. doi: 10.1080/13548506.2017.1332373.

Tanham, J., Murphy, M., & Breslin, G. (2014). Using financial incentives to increase physical activity, weight loss and well being: a randomized control trial. *Working Papers in the Health Sciences*, *1*(9), 1-5.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211, 453-458.67

Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Matzke, D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*(1), 35-57.

Wattanapisit, A., & Thanamee, S. (2017). Evidence behind 10,000 steps walking. *Journal of Health Research*, *31*(3), 241-248. doi: 10.14456/jhr.2017.30.

Williams, R. L., Wood, L. G., Collins, C. E., & Callister, R. (2015). Effectiveness of weight

loss interventions–is there a difference between men and women: a systematic review. *Obesity Reviews*, *16*(2), 171-186. doi: 10.1111/obr.12241.

Willem, A., De Rycke, J., & Theeboom, M. (2017). The Role of Autonomous and Controlled Motivation in Exercise Intentions of Participants in a Mass Cycling Event. *Frontiers in psychology*, *8*, 354.

Wood, W., & Neal, D. T. (2016). Healthy through habit: interventions for initiating & maintaining health behavior change. *Behavioral Science & Policy, 2*(1), 71-83.

Table 1.
Mean (SD) Sample Characteristics between Conditions at Baseline.

Variable	FI	MCC	MCC+FI	CG	CWOG	*p*
Age	28.67 (8.83)	39.53 (12.19)	37.25 (13.02)	29.67 (9.57)	35.75 (10.83)	.04
Age range	20-45	21-64	23-61	18-51	23-59	-
Sex (% female)	93.3	100	100	66.6	76.9	.06
Steps (unadjusted)	7672 (3386)	8390 (2758)	7752 (2711)	9044 (3864)	7850 (3015)	.47
% ≥ 10,000 daily average (unadjusted)	20	19	25	33	13	.70
% ≤ 8,000 daily average (unadjusted)	47	38	50	53	63	.71
Steps (adjusted)	8146 (3489)	9052 (2770)	8259 (2407)	9895 (3809)	8150 (3220)	.70
% ≥ 10,000 daily average (adjusted)	27	31	25	40	13	.53
% ≤ 8,000 daily average (adjusted)	47	31	38	40	50	.84
Total hours not worn the pedometer	10.47 (18.59)	10.84 (18.66)	11.38 (16.28)	9.33 (13.63)	4.84 (4.46)	.50
% doing some physical activity without wearing the pedometer	40	25	13	33	19	.41
Intention	4.63 (1.19)	4.78 (1.88)	4.83 (1.68)	4.47 (1.12)	4.33 (1.17)	.83
Autonomous motivation	4.88 (1)	5.04 (0.81)	4.98 (1.47)	5.25 (1.12)	4.96 (0.77)	.90
Controlled motivation	2.13 (0.76)	2.44 (1)	2.47 (1.41)	2.35 (1.05)	2.26 (1.18)	.87
Amotivation	2.64 (1.16)	2.27 (0.90)	2.54 (1.23)	2.48 (0.79)	2.60 (1.02)	.99

Note. FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal.

Table 2.
Adjusted and Unadjusted Steps Analyses (Controlling for Age).

Condition	Adjusted steps mean (SD)	ANCOVA (BF_{10})	Post-hoc *t*-tests (Tukey)	Post-hoc Bayesian *t*-tests	Unadjusted steps mean (SD)	ANCOVA (BF_{10})	Post-hoc Bayesian *t*-tests	Post-hoc *t*-tests (Tukey)
FI	1. 9690 (3769) 2. 9690 (3769) 3. 10562 (2891)	1. *** (4.9, ++) 2. *** (3.6, ++) 3. ** (3, +)	1. MCC+FI > MCC** & CWOG** & CG* 2. MCC+FI > MCC** & CG* & CWOG* 3. MCC+FI > CWOG**	1. FI = all conditions (=); MCC = all conditions (=) except < MCC+FI (=); MCC+ FI > CG (+) & CWOG (++) 2. FI = all conditions (=); MCC = all conditions (=); MCC+FI > CG (+) & CWOG (+) 3. FI = all conditions (=); MCC = all conditions (=); MCC+FI > CG (++) & CWOG (+)	1. 9725 (3598) 2. 9725 (3598) 3. 10558 (2833)	1. *** (7.6, ++) 2. *** (4.9, ++) 3. ** (3.8, ++)	1. FI = all conditions (=); MCC = all conditions (=); MCC+ FI > CG (+) & CWOG (+) 2. FI = all conditions (=); MCC = all conditions (=); MCC+FI > CG (+) & CWOG (+) 3. FI = all conditions (=); MCC = all conditions (=); MCC+FI > CG (+) & CWOG (+)	1. MCC+FI > MCC* & CG** & CWOG*** 2. MCC+FI > MCC* & CG* & CWOG** 3. MCC+FI > CWOG**
MCC	1. 9504 (4186) 2. 9544 (4341) 3. 10841 (3110)				1. 9551 (3598) 2. 9596 (4130) 3. 10718 (3058)			
MCC+FI	1. 11755 (2984) 2. 11755 (2984) 3. 12103 (2794)				1. 11456 (3034) 2. 11456 (3034) 3. 11811 (2839)			
CG	1. 9386 (2967) 2. 9386 (2967) 3. 8707 (2917)				1. 9125 (2819) 2. 9125 (2810) 3. 8630 (2828)			
CWOG	1. 8680 (3268) 2. 8481 (3400) 3. 9039 (3604)				1. 8559 (3234) 2. 8349 (3360) 3. 8934 (3574)			

Note. 1 = adjusted or unadjusted steps (total sample); 2 = adjusted or unadjusted steps (without contaminated participants); 3 = adjusted or unadjusted steps (without participants engaging in exercise without the pedometer); BF_{10} = Bayes factor; FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal.
= (equals sign) = Bayes factor evidence for the null hypothesis of no condition difference (where =, anecdotal, ==, moderate); > or < = statistical evidence of more or fewer steps in corresponding conditions; + = anecdotal, ++ =, moderate Bayes factor evidence in favour of condition differences.
* ≤ p = .10, ** ≤ p = .05, *** ≤ p = .01.

Table S1.
Mean (SD) Potential Psychological Mediator Scores and Internal Consistency.

Measures	FI		MCC		MCC+FI		CG		CWOG		*F*	BF_{10}
ω (T1, T2, T3)	T2	T3	T2	T3	T2	T3	T2	T3	T2	T3	*(p)*	
Intention	4.92	4.85	4.97	4.92	5.28	5.35	5.07	4.88	4.69	4.55	0.12	.07
(.92, .93, .92)	(0.93)	(1.32)	(2.07)	(1.73)	(1.46)	(1.60)	(1.07)	(1.56)	(1.33)	(1.15)	(.97)	
Aut. motivation	4.94	4.78	5.49	5.66	5.10	4.99	5.39	5.19	5.06	5.33	1.19	.32
(.85, .89, .87)	(1.05)	(1.37)	(.76)	(.79)	(1.34)	(1.53)	(1.20)	(1.54)	(1.07)	(.98)	(.33)	
Con. motivation	2.40	2.34	2.05	2.04	2.24	2.07	2.49	2.29	2.31	2.08	0.05	.06
(.86, .87, .89)	(.94)	(1.13)	(.78)	(1.01)	(1.50)	(1.65)	(1.01)	(1.18)	(1.32)	(.80)	(1)	
Amotivation	2.70	2.90	2.41	2.44	3.09	3	2.73	2.46	2.59	2.50	0.45	.09
(.66, .80, .78)	(1.22)	(1.57)	(1.14)	(1.56)	(1.99)	(2.14)	(1.41)	(1.31)	(1)	(1.22)	(.77)	

Note. Aut. = autonomous; Con. = controlled; ω = McDonald's ω (internal consistency); T1 = time 1, T2 = time 2, T3 = time 3; FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal; ANCOVA *F* value is time*condition; BF_{10} = Bayes factor: figure provided is the likelihood of the time*condition model after reducing error variance attributed to main effects, versus a model consisting of both main effects (condition, and time). Note both the frequentist and Bayesian ANCOVAs here are controlling for age.

Table S2.

Mean (SD) Intervention Acceptability Scores.

Item	FI	MCC	MCC+FI	CG	CWOG	*F (p)*
Fair	4.07 (.96)	4.31 (.95)	4.47 (.64)	4.29 (1.33)	4.13 (.81)	0.33 (.86)
Fun	3.40 (.74)	3.88 (1.02)	4.13 (.64)	3.71 (1.20)	4.06 (.77)	1.71 (.16)
Helpful	4.40 (.63)	4.25 (1.13)	4.60 (.63)	4.43 (.94)	4.38 (.62)	0.49 (.74)
Easy to use	4.57 (.65)	4.25 (1.06)	4.60 (.74)	4.57 (.94)	4.44 (.89)	0.52 (.72)
Convenient	4.33 (.62)	3.63 (1.09)	4.07 (.88)	4.36 (.93)	4.13 (1.20)	1.01 (.41)
Acceptable	4.53 (.64)	4.50 (.73)	4.60 (.63)	4.64 (.84)	4.50 (.52)	0.27 (.90)
Recommend	4.27 (.70)	4.25 (1)	4.60 (.63)	4.07 (1)	4.38 (.72)	1.06 (.37)
Liked monitoring	4.40 (.74)	4.63 (.50)	4.47 (.83)	4.00 (1.18)	4.31 (.79)	1.47 (.22)
Liked earning vouchers	4.73 (.59)	4.88 (.50)	4.80 (.56)	-	-	0.26 (.77)
Vouchers increased activity	4.07 (1.28)	3.38 (1.54)	4.07 (.88)	-	-	2.13 (.13)

Note. FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal; All scales rated from 1 (strongly disagree) to 7 (strongly agree); All analyses here controlled for age.

Table S3.
Further Comments from Participants.

Condition	Sex	Age	Student/Employee	Baseline Steps	Goal Achievement	Comments
FI	Female	25	Student	6079	Achieved	It was a good motivating study which increased my daily steps
FI	Female	24	Student	9530	Failed	12000 steps a day was too far to feasibly manage everyday. It would have been easier to average 12000 over the week/14days with similar exercise benefits. I could probably have done it though if I hadn't gotten ill
FI	Female			1326	Failed	Only downside was how clunky the pedometer was. Would prefer to wear one or have one that showed less
FI	Female	24	Student	5940	Failed	I enjoyed taking part, thank you
FI	Female	24	Employee	8699	Failed	Thank you for letting me take part!
FI	Female	20	Student	10207	Achieved	It didn't ask too much, as in only 2000 step increase which will be possible for most people
FI	Female	22	Student	12279	Achieved	It is motivating at the time but maybe not in the long run
FI	Female	22	Employee	8168	Failed	Good, easy study. Not too disruptive of normal day to day life - pedometer sometimes uncomfy when sitting/can't wear a dress
FI	Female	45	Employee	9258	Achieved	Really enjoyed walking more on this study
FI	Female	29	Employee	691	Achieved	The voucher incentive probably encouraged me to increase my steps more than any other reason!
FI	Female	44	Employee	9904.5	Failed	As a fairly active person, it was very difficult for me to increase my steps without difficulty swapping cycling for walking

						(cycling counts fewer steps) energetic housework etc (few steps but active)
MCC	Female		Employee	13019	Achieved	Improved my activity level and felt healthier for taking part in the study
MCC	Female	31	Employee	7253	Failed	the study initially motivated me to do 10000 steps which I found much harder than expected, after the battery fell out of the pedometer and then forgetting to put the pedometer on, my motivation to achieve the steps fell
MCC	Female	39		7721	Failed	the wording of some of the questions is confusing e.g. 8000 steps per day or 8000 steps over a period of time
MCC	Female	27	Employee	6488	Failed	Thank you. It has been interesting to see how many steps a day (on average) that I take. Some days I have surprised myself. I would like to maintain a healthy lifestyle, and therefore try and increase my daily steps.
MCC	Female	53	Employee	11303	Failed	I enjoyed the study and think that had the amount of steps necessary increased, I would have attempted to increase the no of steps necessary. I would prefer a watch or something less awkward than the clip on, as the clip falls off!
MCC	Female	56	Employee	9328	Achieved	I enjoyed seeing the number of steps I walked each day and liked having the incentive of a goal. I am thinking of getting a pedometer so I can monitor my steps regularly

MCC	Female	43	Employee	7163		The pedometer fell off several times - would have been great if it had a belt loop on it, or some other way to more securely attach
MCC+FI	Female	33	Employee	5890	Achieved	Has really pushed me to increase the number of steps I do per day and actually get into a healthier routine, will miss my pedometer
MCC+FI	Female	61	Employee	6191	Achieved	I found it hard at first to stick to my goal of 10000 but enjoyed monitoring my progress
MCC+FI	Female	49	Employee	8896	Failed	Easy to do and during first week was amazed at how many steps I do a day
MCC+FI	Female	45	Employee	5284	Failed	The pedometer was not easy to use, too bulky, kept coming off waste band depending on what i was wearing. The idea of the study was good, just circumstances meant that I couldn't reach the goals this time.
MCC+FI	Female	27	Student	2986	Failed	I have never used a pedometer before, so this was a first and it was quite a fun experience. Initially it was tough to figure out where to clip the pedometer on if I am not wearing pants but I managed to figure it out in the end. Using vouchers was a great incentive, though money would be a good incentive as well
MCC+FI	Female	30	Employee	13949	Achieved	I found some questions difficult to answer. unclear whether Q5 refers to steps per day or steps per week. Difficult to answer the

						questions about increasing my daily steps as the average given to me after the first 7 days showed I wouldn't be increasing my daily steps
MCC+FI	Female	55	Employee	6172	Achieved	I was aware I wasn't active enough before the study. It was difficult at first to increase my steps but I found more convenient ways to do so during the study e.g. getting off the bus before my stop and taking a short walk during lunch breaks. Thank you for helping
MCC+FI	Female	26	Employee	10468	Achieved	I thought it was a good way to workout current steps and it motivated me to walk more. I liked the challenge of doing it and it’s helped me improve my fitness and stamina and lose weight. I enjoyed the gamble
CG	Female		Employee	7515	Failed	It acted as an encouragement to walk more thank you
CG	Male	42	Employee	4687	Failed	Pedometer count was lower than individuals Fitbit recordings
CG	Female	21	Student	14609	Achieved	I enjoyed participating in the study as I never knew how many steps i took per day. The only thing that was a bit confusing was the aspect of goal setting for number of steps, which could have been explained a bit clearer
CG	Male	25	Student	4833	Achieved	Was fun
CG	Female	28	Employee	7713		I just thought it was great, I am really not sure why I haven't used a pedometer before. Regarding the study itself, I thought

						everything was well done
CG	Female	51	Employee	2686	Failed	Unfortunately my pedometer was faulty so inaccurate steps recorded. This has encouraged me though and have now bought my own pedometer to wear
CG	Female	18	Student	14518	Failed	The study does not take into account other reasons for which people's daily activity might change: type of work, location of the house, financial issues. The number of steps takes everyday does not have to be motivated by the willingness to improve one’s health but by more practical reasons
CG	Female	32	Student	7079	Failed	At the beginning it was difficult for me to remember the pedometer. However it became something to achieve everyday and I like that. the problem was also the timing. I could achieve more but right now I am in my dissertation time, so I couldn't spend time doing exercise or walking more.
CG	Female	24	Student	4774	Achieved	When I started increasing my daily steps I felt like the main reason I wanted to do that was taking responsibility for my own health. However, being an extremely busy person, I realised over the past 14 days that most of the times I would struggle to reach my daily goal mostly because I was asked to do so for the purpose of this research.
CG	Female	28	Employee	9521	Failed	The target I was given was difficult to achieve due to time

						constraints and also it left me a lot more tired - a more realistic target would have been better as I would have been less exhausted.
CG	Female	23	Student	11045	Failed	I enjoyed the study once I got into it, became quite competitive with myself and I was quite disappointed on days when I didn't reach my goal. I am considering buying my own pedometer to continue monitoring my steps.
CWOG	Female	30	Student	7306	Failed	I think pedometers need more functions it just checks only my steps. maybe have some more information about steps or heath
CWOG	Male	32	Student	6941	Failed	It was eye opening to see how many steps i had done particularly on days where I didn't go out as much. Helpful in raising awareness. Though sometimes it can be difficult to get the steps in if you're at a desk all day
CWOG	Female	55	Employee	13094		The amount I walked per day was not changed due to the pedometer. It was more to do with weather conditions, how I felt, what tasks/activities I had to do. I found it interesting to know how far I was walking, but it hasn't changed my walking habits
CWOG	Male	47	Employee	9529		Good fun to monitor the amount of steps per day. The results can be quite surprising
CWOG	Female	59	Employee	9256	Failed	Not sure my pedometer was recording my steps accurately. I

						checked it against my own pedometer, my phone and friend 'Fitbit'. Some days counted more steps than others, even though very similar activity
CWOG	Female	30	Employee	13510	Achieved	The pedometer did not track steps climbing/Pilates/yoga etc. so intensity of exercise was overlooked therefore distorting the outcome re: sense of achievement
CWOG	Female	29	Student	6436	Failed	Reminder to put pedometer on
CWOG	Female	47	Employee	5002	Failed	By wearing the pedometer made you think more about actually walking

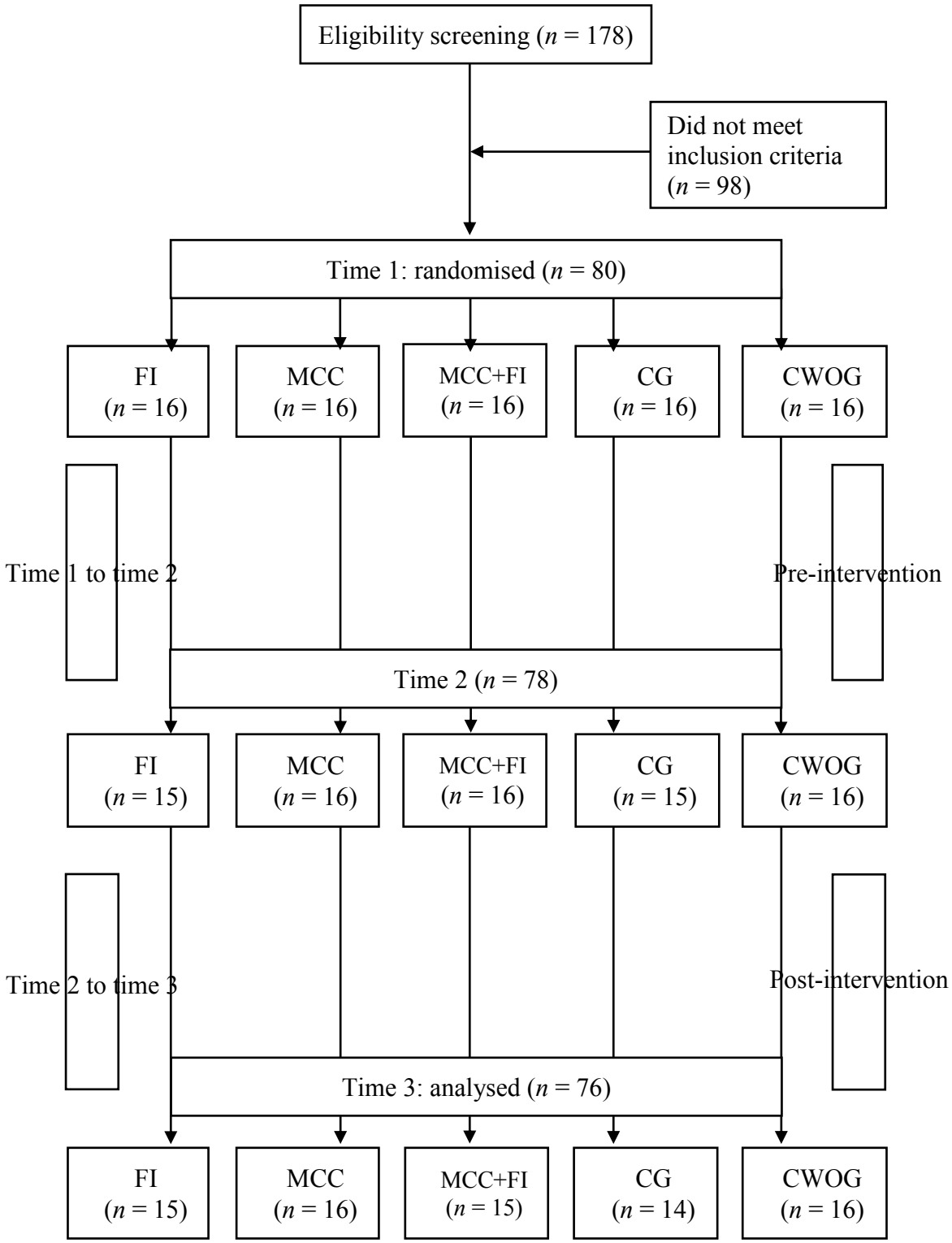

Figure S1. Diagram of participant flow through the study. FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal.

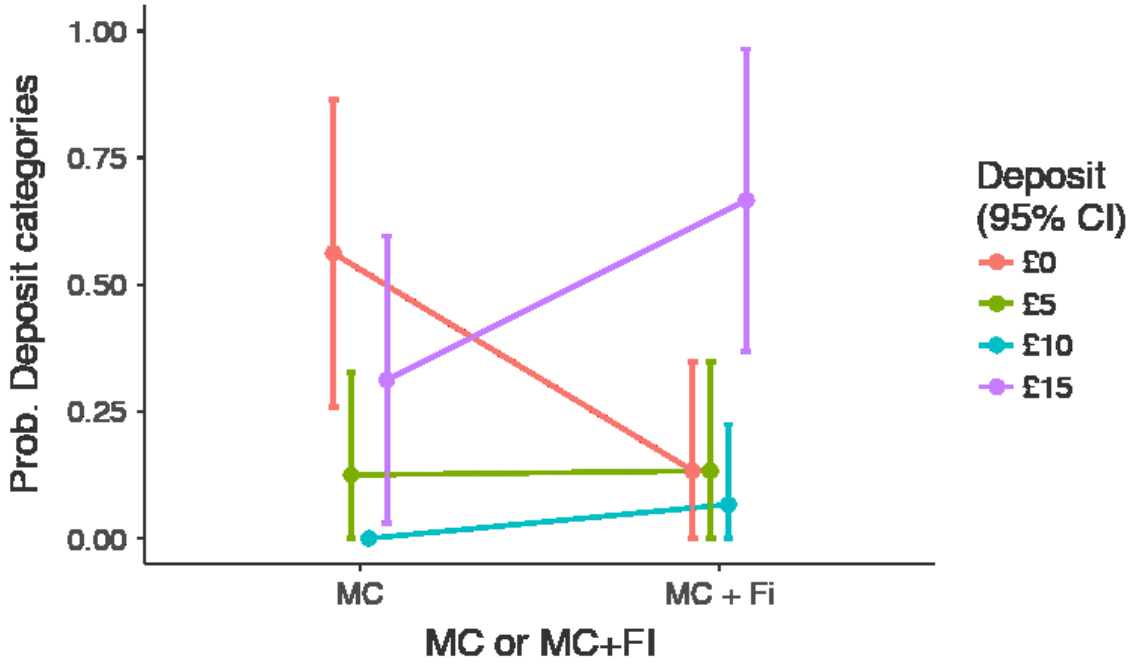

Figure S2. Probability of different deposit amounts between monetary contingency conditions. Error bars (95% CI) = 95% confidence intervals. MC = monetary contingency contract, MC+FI = monetary contingency + standard financial incentive.

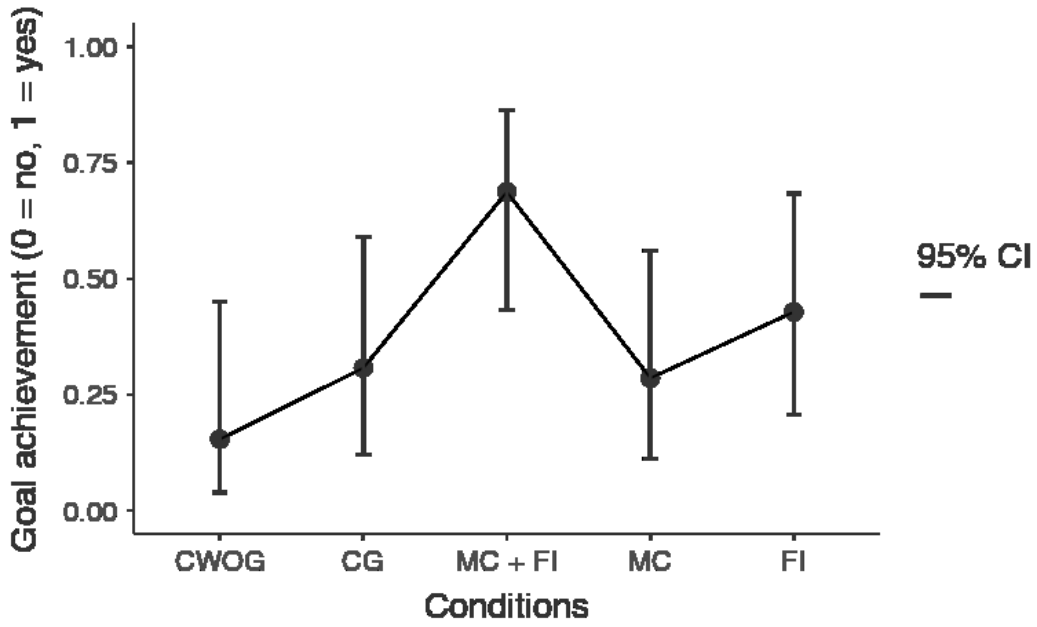

Figure S3. Proportion of goal achievement between conditions. Error bars (95% CI) = 95% confidence intervals. FI = standard financial incentive; MCC = monetary contingency contract; MCC+FI = monetary contingency contract + standard financial incentive; CG = control with a set goal; CWOG = control without a set goal.