Moving the Goalpost Closer: Do Flexible Targets Improve the Behavioral Impact of Incentives?

October 2019

Haijing Huang[1] Sebastian Linnemayr[2]

Abstract

Incentives are increasingly used to motivate behavior change. An unexplored yet critical question in incentive design is how to set the eligibility condition for receiving the reward; should everyone be subject to a uniform, "fixed" level of the targeted behavior, or does "flexible" targets that may vary across individuals, lead to greater improvements in outcomes? Participants at an HIV clinic in Uganda were randomly assigned to either receive incentives with "Fixed" or "Flexible" targets for improving medication adherence over nine months. Setting a high "Fixed" target did not affect adherence while allowing participants to adjust their target relative to their baseline performance improved adherence, with particularly large improvements among those with low levels of adherence at baseline. Findings suggest commitment to a goal and subsequent effort to achieve it may depend critically on perceiving the goal to be attainable.

Keywords: Behavioral economics, incentives, health behaviors, HIV, medication adherence, goal-setting

[1] Corresponding author. Affiliation: IDinsight. Email: crystalhuang33@gmail.com Present address: 785 Market Street, Suite 200, San Francisco, CA 94103
[2] Affiliation: RAND Corporation. Email: sebastian_linnemayr@rand.org

1 Introduction

Personal financial or in-kind incentives are increasingly used to promote healthy behaviors, but questions remain around how they can be designed to best serve their function of motivating sustained behavior change. This paper explores a key feature in the design of incentives which moderates their impact - the condition or requirement for receiving the incentive, which is often based on attaining some pre-determined level of the targeted behavior. We investigate how this eligibility condition should be determined, and whether the same target should be applied to everyone.

Incentives change behaviors by alleviating either behavioral or income barriers to an action. Through the lens of standard economic theory, incentives drive behavior change if the incentive amount is large enough to lower the cost of an action relative to its value to the decision-maker. The empirical literature has shown that this transactional approach can backfire- attaching a monetary reward to an intrinsically valuable behavior can crowd out intrinsic motivation and the sense of agency to complete the task (Kamenica, 2012). By contrast, smaller-sized or in-kind incentives have been effective in motivating behavior change among individuals with preferences that deviate from assumptions of standard economic theory; an example of this is "present-biasness" where the current cost or inconvenience of an action -such as getting immunized- is disproportionately weighted over the long-term benefits of taking that action (i.e. remaining healthy)- leading to lower than rationally optimal levels of the desired health behavior (Dupas 2011). In these situations, incentives modify the trade-off between the current costs and future benefits of a behavior by altering the *psychological*, rather than monetary, value an action has in the present.

Banerjee et al. (2010), for example, found that providing a small bag of lentils increased one-time malaria immunization rates for children; this small incentive was enough to neutralize present-bias by offering immediate gratification for adopting a behavior that had only long-run payoffs. Other studies found that vouchers for small amounts of less than $3 resulted in large attendance gains in obtaining HIV test results and even the smallest amount, a tenth of daily wage, had a large effect (Thornton 2008) or that providing food vouchers increased the uptake of voluntary medical male

circumcision in Kenya (Thirumurthy et al. 2014). In a field experiment in the US, small incentives improved the take-up of a healthy snack among school children (List and Samek, 2015).

Gneezy (2011) points out that while incentives can be a potent way to motivate behavior change, their effectiveness depends on factors such as: whether the incentive may potentially crowd out intrinsic motivation (more likely for larger financial incentives) and whether they are designed in accordance with the decision-making environment. For any incentives-based intervention, several design decisions must be made that may moderate the incentive's impact. So far, some studies have explicitly varied certain design elements and found that features of an incentive- such as size (Baird et al. 2009), framing incentives as losses versus gains (Patel et al. 2016, Chokshi et al. 2018), timing/frequency (Loewenstein et al. 2013), type of payout (Evans et al. 2008, Banerjee et al. 2010) - play a role in how effective they are. Drawing on the Prospect Theory insight that people tend to overweight small probabilities, designing incentives to take the form of lotteries has also been used a lower-cost and more effective alternative to providing a fixed value amount.

In this study, we make two contributions to the literature on incentives. First, we contribute to the evidence base around the *impact* of small incentives on improving health behaviors in a chronic health context, where applications of behavioral economics have potential but thus far limited empirical application. Secondly, we add to the literature on the *design* of incentives by considering the question of how to set a key incentive parameter –the eligibility condition for receiving an incentive, and whether varying this condition can lead to improvements in impact. The specific question we investigate in this paper is whether the threshold for eligibility should be set the same for everyone , or whether incentives are more effective when taking into consideration individuals' starting point, i.e. their performance level of the incentived behavior at the outset.

For the first contribution, our study builds on the limited evidence base on the use of behavioral economic tools to improve health behaviors for chronic conditions. Failure to adhere to medication regimens, estimated at about 50 percent among patients with chronic illness (WHO 2003), can have severe public health consequences; non-adherence leads to increased morbidity and death and is estimated to incur costs of around $100 billion per year (Brown and Bussell 2011). As chronic disease rates are rising in every region of the world, a key policy question for the

management and prevention of chronic diseases is how to improve adoption of and adherence to repeated health behaviors over time. A few studies have in the US have tested the effect of incentives and found them to improve outcomes in chronic disease management including warfarin medication adherence (Kimmel et al. 2012) and weight loss (Volpp et al. 2008). In low and middle-income countries however, most "nudge" studies have focused on one-time behaviors such as purchasing preventive products, immunizations, getting tested or screened, etc. Only recently have nudge-type interventions focus on chronic health behaviors; for example, Linnemayr and Stecher (2017) find that adult HIV patients receiving small, in-kind incentives were more likely to achieve clinically optimal medication adherence compared to the control group (Linnemayr et al. 2017).

This study also speaks to the literature on the design of incentives. Incentive studies typically set targets for receiving the payment based on a "fixed" (and typically uniformly high) performance threshold (e.g. reaching proficiency or grade "A" on a test) that applies to everyone (Bettinger 2012; Leuven, Osteerbeck, and van der Klauw 2010; Patel et al. 2016). An alternative approach, which we test in this paper is to introduce flexibility to the target, i.e. to allow individuals to choose their own, variable thresholds. To our knowledge, this is the first study to experimentally vary the eligibility threshold and test the relative effectiveness of conditioning the incentive on a flexible (participant-chosen) threshold versus a high, fixed level.

We varied the conditionality for receiving rewards among 216 study participants presenting at a publicly-funded HIV clinic in Kampala, Uganda. Participants were randomized to two treatment arms that allowed them to enter a prize drawing for a small incentive if their medication adherence, defined as taking *x* percent of prescribed pills, met or exceeded a target. In the "Fixed Target" treatment arm, a participant had to reach the fixed, clinically recommended target of 90 percent adherence (i.e. they must take 90 percent of pills prescribed) to be eligible to participate in the prize drawings for an incentive. A participant randomized to the "Flexible Target" arm had to meet an adherence target which they could select themselves to be closer to their initial performance for the same incentive amount. Eligibility to receive an incentive (a prize drawing with expected value of $1.40 in mobile airtime) during routine clinic visits was conditional on achieving electronically-measured adherence above the target. If the target was met and participants drew a non-zero number, study coordinators directly disbursed the incentive (mobile airtime of the amount

indicated on the card they drew) to the participant's phone. Medication adherence is the primary outcome of interest, with on-time clinic visits (that typically coincide with pharmacy refills) and successful viral suppression as secondary outcomes.

Conceptually, we draw on several strands in the psychology and behavioral economics literature. For example, the seminal work by psychologist Edwin Locke suggests that goals mediate the effect of incentives on performance (Locke 1982), yet the incentives literature to date has largely sidestepped the question of how goals or targets in the context of an incentive policy should be set. In the psychology literature, making goals *participative* - that is, allowing individuals to participate in determining a goal for themselves- and creating "subgoals" have been shown to improve performance via boosting self-efficacy. Participative goals result in higher self-efficacy, goal commitment and performance, compared to assigned goals (Erez et al. 1985, Latham et al. 1994, Sue-Chan and Ong 2002). Sub-goaling, or setting smaller achievable goals, leading up to an overall target performance, boosts initial self-efficacy perceptions, self-satisfaction with performance, and subsequent task persistence (Stock et al. 1990). The concept of subgoaling has also been conceptualized in a Prospect Theory framework in economics; Heath et al. (1999) present a parsimonious framework, based on the idea that goals serve as reference points and alter outcomes in a manner consistent with the value function in Prospect Theory (Kahneman and Tversky 1979). A key behavioral implication of this conceptualization is the "goal gradient" effect: the closer an individual gets to a goal, the more motivated s/he is to achieve it. Conversely, someone far away from a target is less likely to exert effort to move toward it than someone who is initially closer and indeed may be demotivated by the goal (Heath et al. 1999).

The "Flexible" incentive treatment arm in our experiment is based on this strand of the literature and increases goal ownership through making the goal participative and simultaneously allows for individuals to create subgoals - relative to the "Fixed" arm. To assess whether the mechanism of subgoaling may be driving results, we use the framework set forth in Heath et al. 1999 and assess differential treatment impact by individual's starting point performance. We hypothesize that low initial performers would experience a faster improvement in adherence in the "Flexible" treatment arm, as they are able to set more manageable targets closer to their baseline performance, compared to similarly low performers in the "Fixed" arm who may be demotivated by a high, uniform target.

Several insights emerge from the experiment: first, when pooling the two incentive arms, incentives improve mean adherence by 7.7 percentage points (pp) compared to the control group at borderline levels of conventional statistical significance (p-value=.08). However, when comparing each treatment arm separately to the control group, we find that this average pooled incentive effect is driven by the Flexible Target treatment arm. Compared with the control group, those in the Flexible Target arm improve adherence by 11.4 pp (statistically significant at the 5 percent level) compared to a small and not statistically significant difference when comparing Fixed Target to Control.

Second, in an analysis of treatment heterogeneity by baseline levels of adherence, we find that, consistent with our conceptual model, the difference in effect size between Flexible and Fixed Target groups is largely driven by those who struggled with adherence at baseline ("low" adherers, defined as having pre-intervention adherence below 60 percent). Among this group, Flexible incentives improve adherence by 17.3 percentage points relative to Control (p-value<.05), compared to a 5.1-percentage point improvement between Fixed Target and Control that is not statistically significant. This finding is in line with the idea that high goals may exert a demotivating influence on those performing far below them, and that allowing for personalized, lower standards may be more effective at improving outcomes. On the other hand, for those who are initially closer to the goal of 90 percent, the Fixed Target is more effective. While the Fixed Target treatment arm did not show a statistically significant treatment impact on average, for those performing just below the 90 percent threshold at baseline adherence improved by 17 percentage points. Due to small cell sizes, these sub-analyses should be interpreted with some caution. However, as they test pre-specified hypotheses based on a clearly defined conceptual framework, the results provide a useful groundwork for future research in testing a more granular implementation of a goal-gradient approach (i.e. incrementally increasing goal targets) versus a fixed goal.

A third insight is that participants in the Flexible Target group choose adherence targets that suggest intrinsic motivation as a driver to reach clinically meaningful adherence when setting goals. In each game, over 70 percent of participants in this group choose a goal higher than the

allowed minimum of 80 percent (which we set in order not to encourage clinically suboptimal adherence), with the modal target increasing with baseline adherence level, suggesting that many participants view the prize drawing as an opportunity to hold themselves to a higher adherence standard rather than maximizing expected prize monies. This is consistent with the idea that small, nudge-type incentives target behavioral, rather than income barriers. Study participants' response to choosing higher-than-minimum goals suggest they view the targets as a type of commitment device, risking a financial loss in order to meet a goal, such as was found in a study targeting smoking cessation in the Philippines (Giné et al. 2010).

For secondary outcomes, we find that participants in the incentive groups are less likely to miss clinic visits that typically coincide with pharmacy refills on their ART prescriptions. Incentives increase the percentage of timely clinic attendance by 11 percentage points across both treatment groups, or roughly 1 additional timely visit out of 7 visits per 9 months on average. This result has implications for clinics that may wish to improve clinical efficiency, in addition to corroborating the finding that incentives improve adherence, since the two outcomes are linked: timely drug refills are required to maintain consistent pill-taking behaviour. We do not find a statistically significant effect for the fraction of participants showing viral suppression, potentially due to high baseline levels and a relatively short intervention period of nine months that may not have sufficed to bring about changes in viral load despite the improvements in adherence behaviors we reported above.

2. Context and Conceptual Framework

2.1 Adherence to ART and clinic attendance

In Sub-Saharan Africa (SSA), 25.6 million people are living with HIV, accounting for 71 percent of the global total (WHO 2016). ART has transformed HIV from a deadly infectious disease to a chronic condition, but achieving long-term health and reduced likelihood of transmission requires lifelong, high medication adherence and usage of health services. Many patients have been found to experience treatment fatigue, i.e. a "decreased desire and motivation to maintain vigilance in

adhering to a treatment regimen among patients prescribed in long-term protocols" resulting in adherence levels that are below what is clinically recommended (Claborn et al. 2015).

We focus on improving chronic health decisions in the context of HIV/AIDS, where the actions of those at risk or living with HIV have a particularly large effect on disease containment. Increases in international funding for ART and generic competition led to the mass disbursement of ART in many SSA countries, and patients in treatment adhering to their medications achieve nearly normal life expectancies (Egger et al. 2002, Coetzee et al. 2004, Ivers et al. 2005, Laurent et al. 2005) with improvements to quality of life as well as employment and mental health (Bor et al. 2012, Baranov et al. 2015). HIV-infected individuals are thought to have to take at least 90 percent of pills prescribed in order to suppress viral replication and prevent transmitting the virus to others (Gross et al. 2001, Hogg et al. 2002) – yet mean ART adherence typically ranges from 60 to 80 percent when measured objectively with the help of electronic monitoring devices, and only 30 to 60 percent of patients achieve at least 85 percent adherence (Byakika-Tusiime et al. 2005, Mills et al. 2006, Ortego et al. 2011). From a public health perspective, this means that ensuring one-time treatment take-up is insufficient to reap the full benefits of ART; maintaining adherence is an equally important policy goal.

Adherence or non-adherence is a series of daily behaviors by the patient that is generally unobservable to the clinician; as such, its definition and measurement warrants some discussion. Medication adherence, as defined in this chapter, is the use of antiretrovirals (ARTs) at the prescribed dosing frequency, i.e. the percentage of doses taken out of total doses prescribed. Regularly missing doses or dropping out of programs increases the number of resistant strains of the virus, causing effectiveness to wear off over time. In low-income countries like Uganda, suboptimal adherence is exacerbated by the limited accessibility and often prohibitively expensive costs of second and third line compared to first-line drugs.

Few studies can observe patient pill-taking behaviors over an extended period; hence surrogate measures of adherence are often used in research studies. These measures fall into three main categories: (1) subjective measures of adherence based on self-report, or others' report; (2) pharmacologic measures such as pill count, pharmacy refill records, and electronic drug

monitoring (EDM) devices; and (3) physiological methods or indicators, such as plasma HIV RNA levels. Gill et al. (2005) reviewed papers using various forms of the first two measures, comparing them against the objective health outcome, i.e. undetectable viral load. The authors concluded that physician assessment and self-report are the least accurate, pill counts as intermediate, and EDM as the most accurate adherence marker (Gill et al. 2005).

There are few studies investigating ART adherence among youth, the target population in our study, but those available point to particularly severe adherence problems: a review of adherence studies among HIV-infected youth from the United States finds that youth are likely to face greater adherence problems than adults (Reisner et al. 2009). A study from South Africa using pharmacy refills as an adherence measure found that adolescents aged 11-19 were approximately 50 percent less likely than adults to maintain perfect adherence at all time points and 70–75 percent less likely to be virologically suppressed (≤400 copies/mL) at 1 and 2 years after ART initiation (Nachega et al. 2009). In our own paper using an AYA sample in the same care center, we found that less than 30 percent were adherent at the clinically optimal rate (Linnemayr et al. 2017).

Though adherence is the primary outcome variable of interest, we also evaluate timely clinic attendance as a secondary but important process outcome. Timely visits correlate with adherence. In similar HIV settings on-time pharmacy refills are associated with a better CD4 count response (Anoje et al. 2017). A delay in obtaining drug refills during scheduled clinic visits can mean going a few days without pills; over time, patients who are frequently late or have an irregular care-seeking pattern, would not be able to adhere well and are at greater risk of repeatedly missing care or dropping out entirely (Kimeu et al. 2016). The two outcomes must complement one another to ensure good adherence behavior: clinic attendance measures whether a participant is getting their pills on time, whereas adherence measures whether they are indeed taking the pills. From the clinics' standpoint, timely visits by patients are an important outcome as they improve clinic efficiency and planning; unannounced or last-minute reschedules often lead to longer wait-times and uncertainties in staffing.

We also collected viral load data at endline to construct a measure of viral suppression, as another secondary outcome. Since the focus of this study is health behaviors, we consider viral suppression asecondary outcome due to heterogeneity in its relationship with adherence. While long-term

optimal adherence is correlated with being virally suppressed, other factors such as genetics, time since initiating ART, type of ART, length of time maintaining optimal adherence, may influence viral load, introducing more variation in the link between adherence and viral load across individiuals. Recent meta-analyses confirm that the odds of virologic failure is higher for those with suboptimal adherence; however, there was considerable heterogeneity in the relationship between adherence and virologic outcomes based on the type of ART, with newer antiviral drugs requiring a lower level of adherence (Behabze et al. 2016, Kahana et al. 2013). All primary and secondary outcomes are pre-specified in our trial registry on ClinicalTrials.gov, number NCT02918838.

2.2 Conceptual Framework

In this study, we conceptualize goals as reference points that motivate people by creating a negative discrepancy between a person's desired state and their actual state. (Heath et al. 1999, Bonezzi et al. 2011). Initial position affects motivation towards reaching the goal. Those performing far below a certain goal experience the "starting problem" -- inaction resulting from the belief that the goal is unattainable (Louro et al. 2007, Huang et al. 2012). Etkin and Ratner (2011) for example, found that people seek confirmation of the goal's attainability before investing further effort into the pursuit (Etkin and Ratner 2011). A solution to this is "subgoaling": creating two (or more) distinct reference points in addition to the single reference point of the ultimate objective. Bandura and Schunk (1981) found that a proximal reference point created from sub-goaling increases motivation and performance because it provides an immediate and achievable benchmark, whereas a distal goal is ineffective in mobilizing or directing effort (Bandura and Schunk 1981). Subgoaling is most useful when people are doubtful about reaching a goal or are performing far below their goal (Latham 1990, Brunstein 1993, Soman and Shi 2003).

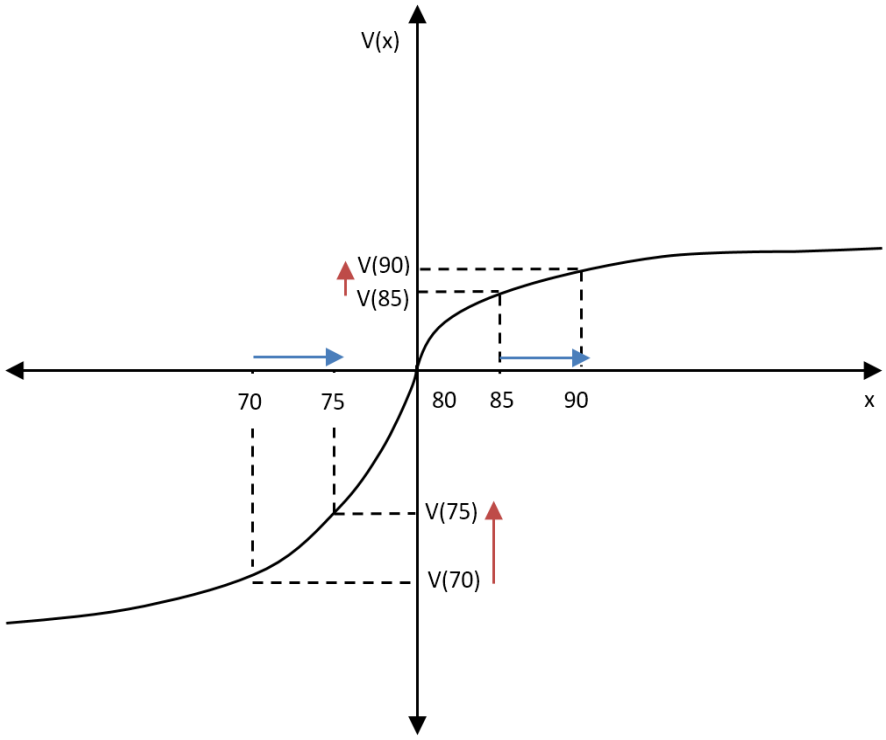
Heath et al. 1999 and Wu et al. 2008 show that the properties of the value function from Prospect Theory is a parsimonious way to conceptualize these empirical findings in the goal literature on effort, persistence and performance: the first insight from the application of Prospect Theory to goal-setting is that goals operate as reference points in that people tend to categorize achievement outcomes as a "success" (gain) or a "failure" (loss) relative to the goal marker. The utility of

outcomes is associated with a psychological value: negative if one falls in the space of "losses" or "failures" where an outcome is below a reference point and positive if one is in the region of "gains" or "successes", where the outcome meets or exceeds the reference point. This means that one's utility or happiness does not depend on absolute but rather on *relative* performance. Consider two individuals A and B who typically adhere to their medication at 80 percent. Person A decides she will try to meet a goal of 90; she ends up scoring 87. Person B decides she will try to meet a goal of 85. She scores 85. Prospect theory predicts that even though Person A outperformed Person B, Person A will feel worse about her performance.

A second property of the value function that applies to goal-setting is "loss aversion" or the idea that losses are more painful than gains are pleasurable ($v(x)<|v(-x)|$ in Figure 1). Loss aversion implies that people who are below their goal by 5 units will perceive their performance as a loss relative to their goal; they will work harder to increase their performance by a given increment than people who are above their goal by 5 units. Because of the shape of the value function, there is a "goal gradient" – the closer one is to their goal, the faster utility increases with an additional unit increase in outcome/performance – and hence the more effort will be exerted to reach the goalpost. Conversely, those who have reached their goal will not exert as much effort for the next unit. This can be seen in Figure 1, which shows movements in adherence outcomes relative to the goal of 80 percent. When the goal is 80, a person moving from 70 to 75 percent adherence will experience a larger utility increase [$v(75) – v(70)$] compared to a person moving from 85 to 90 percent adherence [$v(90) - v(85)$].

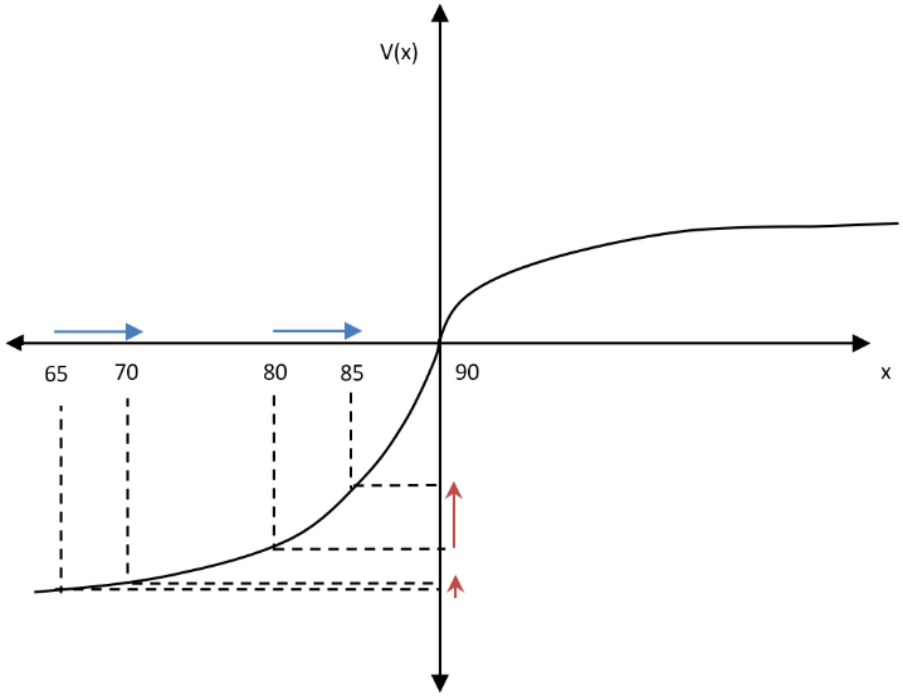
Empirical studies of risky choice and riskless choice have presented converging evidence that losses are weighted approximately twice as much as equivalent gains (common values for the "coefficient of loss aversion" fall between 2 and 4), meaning that people are willing to work twice as hard when they are approaching their goal than after they have exceeded it. Pope and Schweitzer (2011) show that even the behavior of highly experienced professionals in high stakes settings seems to follow predictions of prospect theory: professional golfers for instance, exert more effort when attempting a putt for par than when attempting a putt for scores other than a par, a finding that is indicative of loss aversion relative to the par reference point. (Pope and Schweitzer 2011)

Figure 1. Value Function and Utility and Loss aversion



The S-shape of the value function also shows the property of diminishing sensitivity – that is, the additional loss or gain will impact utility less than the one before it ($v''(x) < 0$ *when* $x > 0$ *and* $v''(x) > 0$ *when* $x < 0$). People who are very far from a goal will not be as motivated as those close to it. Figure 2 shows two people who are both adhering below the goal – one is far away (starting at adherence 65) and the other is closer (adherence 80). The value from moving from 65-70 is much smaller than the value of moving from 80 to 85. The implication from this insight is that high goals could be demotivating for people far away from them. Because of diminishing sensitivity, the marginal unit of effort exerted by someone far away from their goal generates less utility gain than if they were closer – hence progress becomes harder to discern and motivation decreases. This is illustrative of the "starting problem" where it becomes difficult to motivate oneself to start a task when the goalpost is far away.

Figure 2. Value Function and Diminishing Sensitivity



The key takeaway from this conceptual framework is that goals can be viewed as reference points and people's position relative to that point determines how they respond to it. Allowing individuals to set different targets based on varying initial starting values may improve effectiveness by taking advantage of the properties of the value function and maximizing sensitivity right below a goal, which further can improve targeting and therefore the overall impact of interventions.

Using the two features of the value function – loss aversion and diminishing sensitivity - we can make predictions about the two incentive treatment groups and how they may impact participants with differing levels of baseline performance. We begin with grouping participants by their starting point adherence at baseline into "low", "medium", and "high" adherence categories. "High" adherers are those who, at baseline, adhere above the clinical standard of 90 percent. We define "low" adherers to be those adhering below 60 percent at baseline; and "medium" adherers as those adhering between 60 and 90 percent at baseline. Since we do not know the exact shape of the underlying value function, the groupings for "low" and "middle" are somewhat arbitrary- a point

addressed further in section 5.6. For now, some broad predictions can be made for each adherence category:

- *Low-adherers:* Low baseline individuals experience diminishing sensitivity by being far from the traditional, fixed 90 percent target. "Low" adherence individuals are able to select closer targets in the Flexible Target group, and hence are likelier to do better compared to those who similarly struggle with adherence in the Fixed Target group.
- *Medium-adherers: "*Medium" adherence individuals stand to benefit most from the "goal gradient" around the 90 percent adherence target, as their performance is within the range below the reference point where the value function slope is the steepest. Fixed Target medium adherers will improve relative to the Control group. Predictions for medium-adherers in the Flexible Target group depends on how participants select the targets. Someone setting lower targets for themselves, they may improve but with a smaller magnitude of improvement compared to Fixed Targets, who are all subject to the 90 percent reference point.
- *High-adherers-* "High" adherence individuals are already at 90 or past it at baseline. The benefits of the reference point of 90 are lower for this group. Since everyone in the Fixed Goal group faces the same reference point of 90, but some people in the Flexible Target group may adjust upwards and set higher goals for themselves (up to 100%, as seen in Table 3), high adherers in the Flexible Target group may do better than their Fixed counterparts. Since 90 is widely known by patients as a clinical reference point and there is less room for further improvement, the difference between groups is likely to be small.

3. Experiment Design

3.1 Intervention and timeline

This study was registered on ClinicalTrials.gov, number NCT02918838. The experimental design and study population closely relates to the intervention that preceded it. Details from this earlier study (Study 1) can be found in Linnemayr et al. 2017; we describe key details below before describing the main experiment which is the focus of this chapter (Study 2).

Study 1: SMS Intervention

From 2014-16, 332 youth and young adults living with HIV between ages 15-22 participated in a Short Messaging Service (SMS) intervention, a randomized controlled trial which tested the effect of reminder messages on adherence. At the time of recruitment, participants were in HIV care and on ART or prophylaxis regimens at two health facilities in Kampala, Uganda: Mildmay and Infectious Diseases Institute. Both are non-profit organizations that provide ART and other services free of charge to the general population in and around Kampala, serving the general, typically low-income population. Overall the SMS interventions did not have a statistically significant effect on adherence outcomes.

Study 2: Incentives intervention

As Study 1 ended in July 2016, the remaining 229 participants from Study 1 presenting at Mildmay were randomized to a control group or one of two treatment groups in the study discussed here.[3] The main experiment evaluated in this paper is the provision of incentives in the form of prize drawings or "games", as study participants and coordinators called it during the study. Participants randomized into one of the two intervention arms were eligible to receive small incentives of either 0; 5,000 USH ($1.4); or 10,000 USH ($3) of mobile airtime during regular clinic visits conditional on achieving ART adherence above a certain target and making an "on-time" clinic visit (defined as visiting the clinic within 5 business days of their scheduled appointment). In the "Fixed Target" group, participants had to reach the externally imposed and clinically meaningful adherence of 90 percent and make an on-time clinic visit to be eligible for a game. In the "Flexible Target" group, participants were required to reach a target of their own choosing over 80 percent in five-point intervals: 80, 85, 90, 95, or 100 percent that they were free to re-adjust after each visit[4]. If the participant met his/her target, the expected payment of winnings in each game was 5,000 USH, for a total of up to 30,000 USH (approximately $5.58) over up to six games during the 9-month study period; most participants visited the clinic about every two months.

[3] The second clinic (Infectious Disease Institute) was not included in Study 2 because of delays in Institutional Review Board (IRB) approval at that institution.

[4] For ethical reasons we did not want not incentivize people to perform to lower than a medically beneficial target, and hence the target selected had to be above 80%.

The SMS intervention described as part of Study 1 above continued to be implemented in both the control and the two treatment groups. All three groups (including control) therefore continued to receive a weekly text message plus airtime of 1,000 USh (~$0.27) for responding to the message. SMS messages continued primarily for reasons of fairness to the control group who did not receive an intervention in Study 1. The three study groups are summarized in Table 1.

Table 1. Experimental Design

<table>
<tr><th>Fixed Target</th><th>Flexible Target</th><th>Control</th></tr>
<tr><td>Participant draws from 3 cards with expected value ~$1.4 in mobile airtime if fixed adherence target of 90% is met, and participant made a timely clinic visit

Can play up to 6 games during clinic visits over 9 months</td><td>Participant draws from 3 cards with expected value ~$1.4 in mobile airtime if flexible target of participant's own choosing is met, and participant made a timely clinic visit

Can play up to 6 games during clinic visits over 9 months</td><td>No incentives</td></tr>
<tr><td colspan="3">Weekly SMS motivational message + airtime top-up of 1,000 USh (0. 27$) if participant responds to the message</td></tr>
</table>

3.2 Randomization

Prior to recruiting participants for Study 2, we randomized the 229 study participants at Mildmay clinic remaining at the end of Study 1 to one of the three study groups. To maximize the statistical power from this sample, we performed a block randomization based on four strata of baseline adherence (Bruhn and McKenzie 2008). As statistical efficiency is greatest when block variables are highly predictive of follow-up outcomes, baseline adherence is the preferred variable to use. However, we were unable to use an exact definition of "baseline" adherence at the time of randomization due to a lag in when electronic bottle caps were collected and extracted, as well as the rolling recruitment. Study participants were recruited for Study 1 over 6 months so at any given time there was variation in how many months each person had been in the study. Electronic adherence data was also extracted about once every 1-3 months during clinic visits; hence at the time of randomization (3 months before Study 2 started), the latest period of available adherence data for most of the sample was around 6 months prior to the planned Study 2 start date. We used

the three-month average adherence in this period[5]. The four strata of baseline adherence used for randomization were determined by percentile cutoffs of the continuous mean adherence variable. The lowest strata (25th percentile) had average adherence 20 percent, second strata (25-50th percentile) was 66 percent, 3rd strata (50-75th percentile) was 88 percent, and highest strata (75th and up) was 98 percent. The results of the stratified randomization on the original sample (mean adherence in each of the three groups, by strata) are shown in Appendix Table A1.

3.3 Recruitment

Participants were approached for rolling recruitment for Study 2 from August 2016 to January 2017. Eligibility for participation in Study 2 follows from those in Study 1 and means that participants were aged 17-24 years, had daily access to a mobile phone, and were familiar with SMS messaging. Individuals who did not own mobile phones were eligible if they had shared phone access for at least five days per week. Mobile phones or phone airtime were not provided; at baseline about 72 percent of the sample owned a mobile phone. Exclusion criteria included attending boarding school or expecting to attend one, since mobile phone use is commonly prohibited in these institutions.

All participants were using electronic drug monitoring (EDM) bottle caps since the start of Study 1, and simply continued to use these for the new study. These electronic monitors resemble regular bottle caps and contain a tracker (unseen to the patient) that counts each opening and the time of opening. Medication was transferred to the bottle with this cap under the supervision of the study coordinator, and the time and date of each opening was electronically recorded. Participants were not informed exactly how the caps worked hence were less likely to open the bottle strategically, a concern addressed in Section 6.

During recruitment, all treatment participants played a trial game that was not contingent on adherence. This was done both to enhance understanding of how the game works, including its probabilistic aspects (i.e. the reward amount is based on chance) as well as to increase participant "buy-in" by making the game salient and fun from the outset.

[5] The period 9 months to 6 months prior to study start date.

3.4 Clinic and game procedures

Participants came to the clinic for their regular clinic visits on average every 2 months. During each such scheduled clinic visit, study coordinators updated the phone number and entered next dates of appointment into a study-specific electronic tracking database. Participants were not eligible for a game if they missed their appointment dates by a margin of 5 weekdays. For all participants, information from the EDM caps was uploaded to a computer during routine clinic visits, which was easy to implement. The computer software that reads in the data has a front-end interface showing the exact date and times of opening in calendar format including a summary measure of adherence. Appendix Figure A1 shows the chart of opening times for a patient adhering with a twice-daily regimen. Because the spacing between clinic visits varied by participants, we standardized the period of adherence that was checked in order to be eligible to draw a prize; participants were told their adherence would be checked for the last 30 days before their visit. This has the additional advantage of allowing for a robustness check, further detailed in Section 6.

For each participant in the Fixed Target or Flexible Target treatment groups, the game procedures during a visit were the following:

1. The study coordinator verified whether a timely clinic visit was made, meaning the visit corresponded to the scheduled appointment date within +/- 5 days. If so, the study coordinator checked whether the adherence target was met over the last 30 days before the visit. The program showed them the percentage of doses taken out of prescribed doses during this time frame.
2. If the participant's adherence was below their target the study coordinator told them, "This past month you have not reached your target. Your target was [for example] 95 percent, so we cannot play the game today. Next time if your adherence exceeds your target, you can qualify for a draw." If the participant was in the Flexible Target group, the study coordinator would additionally ask: "Your target was [for example] 85 for this visit. Would you like to set a new target or keep the same target for next time?" The new target would then be added to the tracking database for the next game.
3. If the participant scored at or above the target, s/he qualified for a draw and was told "Congratulations, your adherence this past month exceeded your target of [for example]

90 percent. You are eligible to play the game!" Participants then drew once from three face-down cards with the amounts of 0; 5000 USh; and 10,000 USh.

4. When drawing a positive amount, study coordinators congratulated them and disbursed the prize winnings. If the participant had their own phone, study coordinators sent them mobile airtime on the spot via SMS and made sure the participant received it. When the participant did not have a device but had shared access to one, participants were given the choice of receiving airtime on the shared mobile device, or getting a mobile airtime voucher (scratch card).

4. Data and descriptive statistics

4.1 Randomization success and sample characteristics

Of the participants at Mildmay clinic at the end of Study 1, 216 participants were recruited for Study 2. Only 4 patients who were approached declined to participate, and the remaining 9 could not be reached for enrolment. From the starting sample of 216, 209 participants had at least one follow-up clinic visit during the study period wherein data from their EDM caps were extracted. These comprise our final sample, with 67 in Control, 72 in Fixed Target, and 70 in Flexible Target. An endline survey was done at the end of Study 1, which we use to construct baseline variables for Study 2. At baseline, the average age of participants was 19 years, and roughly half (54 percent) were female. Almost all participants had completed primary education and the majority (63 to 77 percent) could read and write a simple sentence easily. Median self-reported weekly income was about 10,000 USh (~$3 USD).

Table 2 shows the balancing table of pre-intervention covariates across study groups. We compare socioeconomic and demographic variables, as well as several measures of adherence, and report the p-value from a chi-square test of equivalence between the three groups, and t-test comparisons of variables between each of the three groups. Randomization appears to have been successful at balancing the majority of baseline variables across groups. None of the socio-demographic variables differ significantly across groups in the three-group comparison. Most baseline characteristics are also well-balanced when comparing each treatment group with the control group.

Stratifying randomization by baseline adherence levels should have ensured that adherence is well-balanced across the groups. However, both adherence at the randomization and adherence 1-month pre-intervention are slightly higher among the Fixed Target group than the Control or Flexible Target group. Adherence the month before the intervention starts follows a similar pattern and is higher among Fixed Target compared to both Control and Flexible Target (significant at 5% when comparing Fixed Target vs. Control, col 5). This discrepancy is due to the change in composition of participants from the initial 229 at randomization and the final sample of 209. Though we could not foresee the selective attrition between studies, we account for this pre-treatment difference in the estimation of treatment impact via the difference-in-differences analysis, which removes any potential biases that result from permanent (i.e. non time-varying) differences between groups observed at baseline.

Table 2. Participant summary statistics

	(1) Control Mean (n=67)	(2) Fixed Target Mean (n=72)	(3) Flexible Target Mean (n=70)	(4) Three-group comparison (p-value)	(5) Fixed vs. Control (p-value)	(6) Flexible vs. Control (p-value)
Female (%)	50.70	53.52	57.97	0.57	0.62	0.29
Age (years)	19.04	19.55	19.25	0.39	0.19	0.30
Married (%)	25.35	29.58	40.58	0.40	0.52	0.24
Literacy (%)						
read easily	69.01	69.01	63.77	0.77	0.96	0.55
write easily	77.46	77.46	65.22	0.20	0.96	0.14
Education (%)						
Completed primary	100.00	100.00	98.55	0.38	.	0.33
Completed secondary	77.46	77.46	65.22	0.23	0.92	0.13
Housing (%)						
Self-rated house as "poor"	43.66	49.30	37.68	0.35	0.55	0.39
Has electricity	78.87	88.73	81.16	0.26	0.10	0.68
Has piped water	53.52	59.15	50.72	0.55	0.51	0.67
Weekly income (USh)						
Mean	20,478.00	19,492.96	17,405.88	0.82	0.89	0.56
income>50k USh (%)	41.43	45.07	42.03	0.85	0.60	0.65
income>75k USh (%)	20.00	23.94	21.74	0.77	0.48	0.64
Owns phone (%)	0.71	0.73	0.71	0.94	0.77	0.98
Adherence(%)						
self-reported	83.85	80.21	85.77	0.18	0.21	0.55
at randomization	0.70	0.77	0.69	0.12	0.11	0.68
1-month pre-intervention	0.63	0.75	0.61	0.04	0.04	0.75

NOTE: This table presents pre-intervention summary statistics for the final sample of 209 participants who had their electronic drug monitoring bottle caps extracted at least once during the intervention. Demographic and socio-economic variables are from the baseline survey. Self-reported adherence is the share of doses taken as prescribed in the past month, as reported by the participant also during the baseline survey. Adherence at randomization refers to the average electronically-measured three-month adherence taken three months before intervention started, when we did the randomization. At an exchange rate of 1 USD to 3,328 USh around the time of recruitment in August 2016, the mean self-reported weekly income is about 6 USD. Col (4) shows the p-values from the F statistic of mean comparisons between all three groups. Cols (5) and (6) shows the p-values of t-tests between Fixed vs. Control, and Flexible vs. Control, respectively.

4.2 Game descriptives

Participants in the intervention groups had an average of 4.5 clinic visits corresponding to a potential prize drawing over 9 months. Across all game-visits, 43 percent of participants were eligible for a prize drawing and played a game. Of those who were ineligible, having a 30-day EDM adherence lower than the target for that visit was the main reason (50 percent), followed by having missed the appointment date (14 percent).

Table 3 shows the adherence targets chosen by Flexible Target participants in each game. Overall, the targets selected suggest that participants were not solely motivated by maximizing their chances of winning a prize (in which case they would select the lowest allowed threshold of 80 percent). In each game only 15-20 percent of participants chose that bottom target. The modal choice was 90 percent; in all games, the majority of participants (over 60 percent in each game) chose a target of 90 or higher suggesting that participants were willing to incur a chance of financial loss in order to commit themselves to a higher standard.

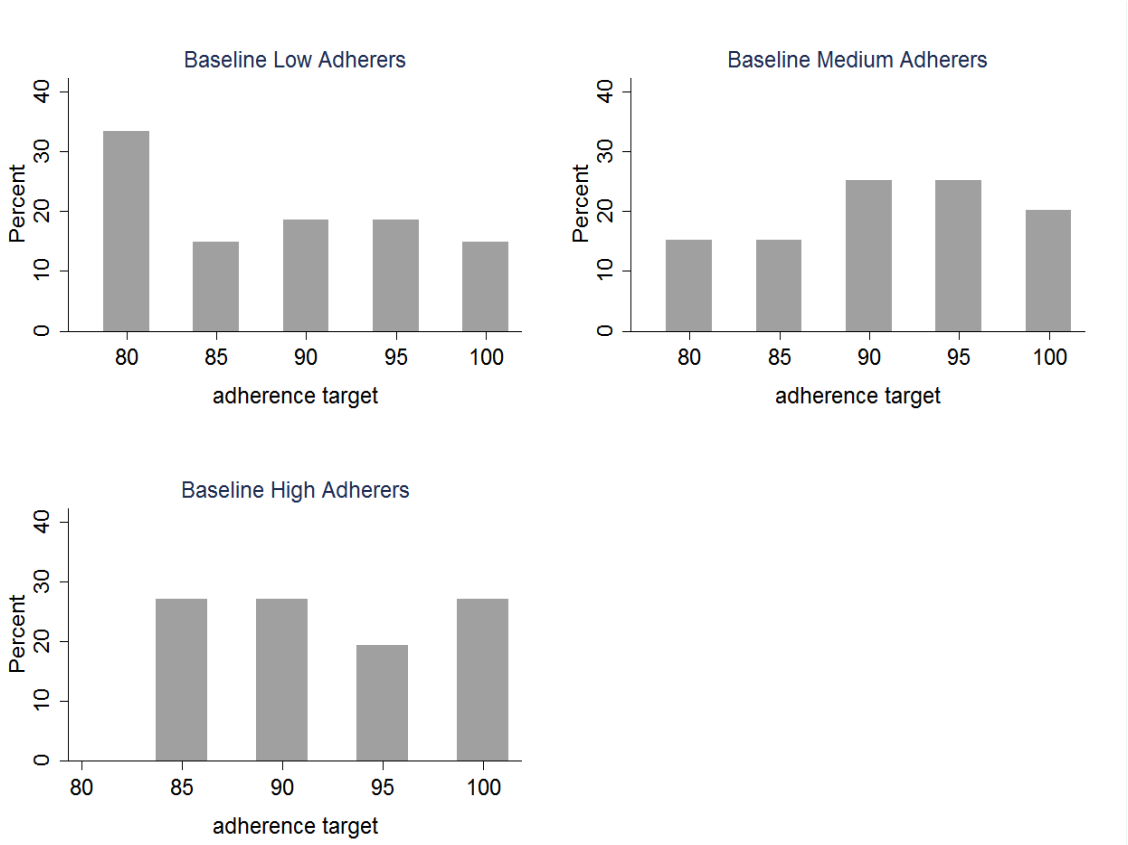
Table 3. Adherence targets chosen by Flexible Target participants

target	Game 1		Game 2		Game 3		Game 4		Game 5	
	n	%	n	%	n	%	n	%	n	%
80	12	17	10	15	11	16	12	20	4	11
85	13	19	13	20	12	18	11	18	10	26
90	17	24	17	26	23	33	19	31	12	32
95	14	20	13	20	13	19	11	18	9	24
100	14	20	13	20	9	13	8	13	3	8
Total	70	100	66	100	68	100	61	100	38	100

NOTE: Percentages for each target may not sum to 100 due to rounding.

Figure 3 shows the adherence target chosen in the first game among the Flexible Target group by baseline adherence category (low, medium or high). Though such an analysis is limited by the small cell sizes in each category, the graph is suggestive that participants were aware of and account for their own adherence abilities at baseline and selected a target commensurate with their adherence level: among low adherers at baseline, the modal target is 80 percent; for medium adherers it is 90 and 95 percent. Among high adherers, none chose the lowest target of 80. The levels of 85, 90 and 100 were the most common selections.

Figure 3. Adherence target among Flexible Target participants by baseline adherence level, Game 1



NOTE: These figures show the distribution of adherence targets chosen by Flexible Target participants, for game 1. "Low", "Medium" and "High" refer to baseline/pre-intervention adherence levels, where Low= 60 percent or less, Medium= greater than 60 and less than 90, High= 90 and higher.

5. Empirical specification and experimental results

5.1 Construction of outcome variables

Adherence (primary outcome)

The primary outcome variable of mean adherence is calculated as the number of bottle openings observed by EDM cap divided by the number of prescribed pills. To prevent inflating adherence by extra openings on a given day we set the maximum number of daily openings to one for patients on once-daily regimens, and equal to two for patients on twice-daily regimens. To prevent counting

erroneous openings, we also discard extra openings within an hour. For those who changed ART regimens over the course of the study, we use pharmacy patient records and patient self-report at each clinic visit to create adherence measures that account for these regimen changes. In such cases, adherence under the new regimen is calculated using the date of change as the first day of the new regimen. Appendix Figures A1 and A2 show examples of EDM readings for a participant with a consistent twice-daily regimen (taking one dose around 8:30 am and the other at 8:30 pm), and another participant with a more sporadic once-daily regimen (taking their pill around 9 am – 1 pm). The second adherence variable "adherence over 90 percent" is a binary variable equal to 1 if the proportion of pills taken over pills prescribed was equal to or exceeded 90% in each month.

Timely clinic visit (secondary outcome)

The secondary outcome of interest is whether patients kept their appointment for a scheduled clinic visit within the week. Clinic visit data is collected from our study patient tracking database that study coordinators compare and update with the clinics' own tracking system. We define "on-time within 7 days" to account for the fact that participants in the two incentive groups were eligible to play a game if they made it to their scheduled visit within a five-business day grace period (or seven including weekends), provided other requirements were met.

Viral load (secondary outcome)

While this study focuses on behavior, we also collect endline viral load data as a secondary outcome, as prespecified in the trial registry. Viral suppression is defined as having plasma HIV RNA of less than 400 copies/ml as is standard in the literature. We also do a sensitivity test applying a more stringent cutoff of 40 copies/ml. Viral load was collected at the 9-month follow-up among 198 study participants who could be reached at the time.

5.2 Empirical specification

While treatment assignment was random, we employ a difference-in-differences (DID) framework to account for baseline adherence differences between Fixed and Control group as discussed above as well as increase the precision of estimates. We analyze monthly adherence over the 9 month

study and 2 months pre-intervention. The availability of pre-intervention adherence allows for differencing out time-invariant individual-specific effects, using the following main specification:

$$Y_{it} = \propto + \mu Treat + \tau Post_t + \beta(Post_t * Treat) + \varepsilon_{it} \quad (1)$$

Where *i* and *t* index individual and month, respectively. Y_{it} is the adherence outcome for individual *i* in month *t*. $Treat_i$ is the treatment assignment of each individual, either pooled (=1 if individual is in assigned to any incentives treatment, 0 if Control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control)[6]. $Post_t$ is a dummy equal to one after the intervention started. ε_{it} is the idiosyncratic error term. All standard errors are clustered at the individual level. The coefficient of interest is β, the estimate of the effect of treatment on adherence. For adherence over 90 percent and timely clinic visit outcomes, we use a linear probability model.

To account for any omitted time-invariant individual effects and individual-invariant time effects, we also run a secondary specification to the above model with both time and individual fixed effects:

$$Y_{it} = \propto + \theta(Post_t * Treat) + \varphi_i + \pi_t + \varepsilon_{it} \quad (2)$$

Where $\varphi_i$are individual effects, π_t are month fixed effects and ε_{it} is the remaining idiosyncratic error.

Equation 1 estimates the treatment effect over 9 months, where β captures a constant treatment effect over time. To examine whether the intervention effect varies over the 9-month study period, we modify equation 1 and estimate the monthly treatment effect with the following specification:

$$Y_{it} = \propto + \mu Treat + \tau Post_t + \sum_{t=0}^{9} \beta_t Month_t * Treat + \varepsilon_{it} \quad , \quad (3)$$

[6] For clarity in the analysis tables below we present results from separate regressions comparing Pooled vs. Control, Fixed vs. Control and Flexible vs. Control. These results are almost identical in magnitude and significance to estimating the joint model $Y_{it} = \propto + \mu T1 + \gamma T2 + \tau Post_t + \theta(Post_t * T1) + \beta(Post_t * T2) + \varepsilon_{it}$.

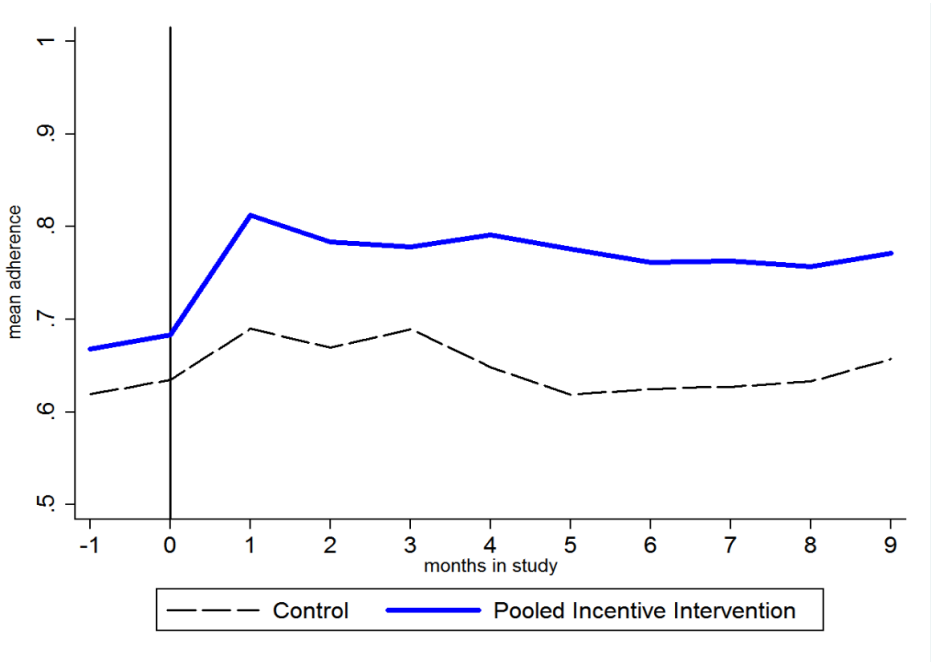
where $Month_t * Treat$ is an interaction between an indicator for each month and indicator for treatment (again, either Pooled, Fixed Target or Flexible Target). The coefficients β_t capture the effect of adherence in month *t* relative to the reference category (two months pre-intervention). For clinic visit outcomes, linear probability models[7] are used, in the standard DID regression and a fixed effect framework following equations (1) and (2). For this analysis, the data are at the individual level by clinic visit where the *t* subscript denotes scheduled appointment. For viral suppression, lacking baseline data, a simple t-test of proportions is used, as well as linear probability regression analysis adjusting for baseline adherence.

Impact of incentives on adherence

First, we estimate the effect of small mobile airtime incentives on adherence. Figure 4 shows mean monthly adherence comparing the pooled intervention arms and control. The pooled incentives intervention group has slightly higher adherence than the control group in the pre-period (months -1 and 0). In the first month of the intervention, there is a sharp increase in adherence in the pooled intervention group that remains between 75 and 80 percent for the duration of the 9 months; the time trend for the control group is much less pronounced- a gradual increase at month 1, which then tapers off to around pre-baseline levels of between 60 and 70 percent.

[7] A linear probability model is preferable to logit/probit model for this analysis, as the DID coefficient is readily interpretable as a marginal effect, which is not the case for interaction terms in nonlinear models. See Ai, C. and E. C. Norton (2003). "Interaction terms in logit and probit models." *Economics letters* 80(1): 123-129.

Figure 4. Mean monthly adherence, Pooled treatment vs. Control



NOTE: This figure shows mean electronically monitored adherence comparing the pooled intervention groups and the control group, from 2 months before the intervention through the 9-month study period. The vertical line at month 0 denotes when the intervention started.

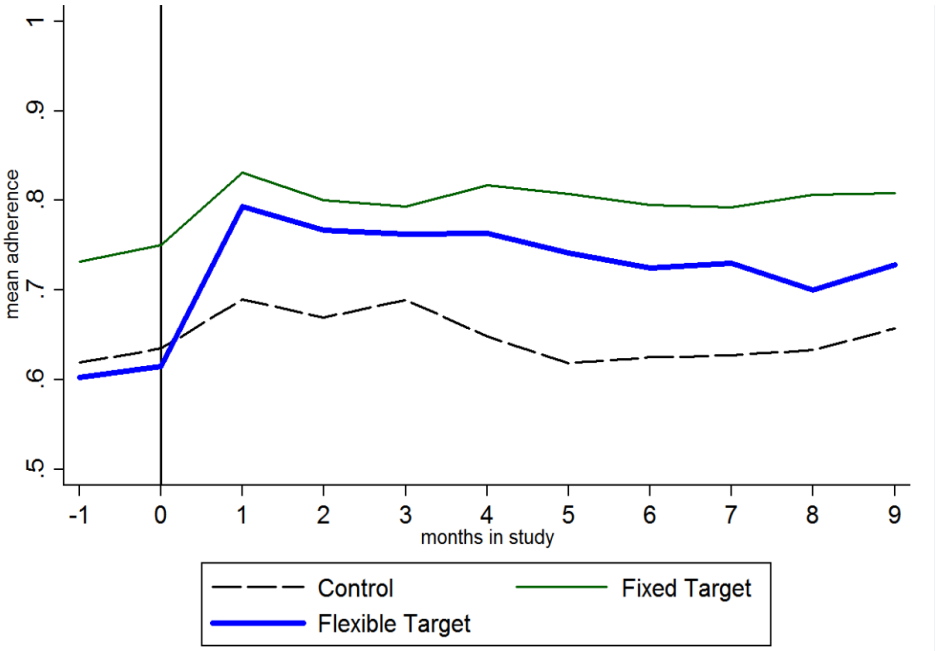
Regression results are presented in Table 4. Among control group participants, average adherence for the two months pre-intervention was 62.7 percent. Overall the pooled incentives intervention has an effect of similar magnitude and significance across specification in equations 1 and 2 (columns 1 and 2 of Table 5). In both specifications, the coefficient for the pooled effect of incentives on mean adherence is 7.7 percentage points, significant at the 10 percent level.

Considering the two treatment arms separately, we find that the pooled intervention effect is mainly driven by the Flexible Target arm, as shown in Figure 5, which plots mean monthly adherence in each of the three groups. In the period before the intervention, mean adherence among those in the Fixed Target group was higher than those in either the Flexible Target or the Control group by over 10 percentage points. Over the course of the study, the between-group difference between Fixed Target and Control participants remained relatively constant, suggesting a relatively small effect from the Fixed Target incentive treatment. However, adherence in the Flexible Target group increased substantially from baseline to month 1 and remained between 70 to 80 percent while Control hovered between 60 and 70 percent. The effects of the Flexible Target

incentives intervention also appear to be sustained throughout the 9-month period. Estimation results in Table 4 columns 3-6 corroborate this picture: when separately comparing Fixed vs. Control and Flexible vs. Control, we find that the pooled intervention effect is largely driven by the Flexible Target group. The treatment estimates are similar between our main specification (eq. 2) and when using a fixed effects model (eq. 3). In our main specification comparing Flexible Target versus Control, mean adherence increased by 11 percentage points on average over 9 months, with p-value<.05. Comparing Fixed Target versus Control, mean adherence increases by just 4 percentage points and is not statistically significant.

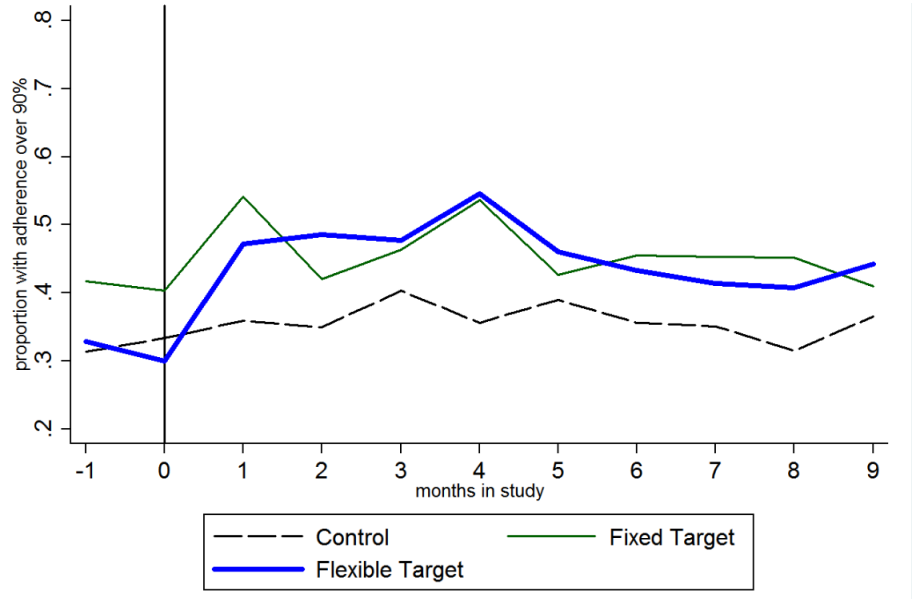
Figure 7 shows our second adherence outcome, the proportion of participants taking at least 90 percent of their scheduled medication, by intervention and control groups. Table 5 shows results from the linear probability regressions. To provide a point of reference, the proportion of control group participants adhering over 90 percent is only 32.3 percent in the period two months pre-intervention. The pooled incentives DID coefficient shows an increase of 7.2 and 6.8 percent in the proportion of those adhering over 90 percent, but is not statistically significant. Among Fixed Target participants, the proportion adhering over 90 percent is slightly higher than control (1.6 percentage points) and not statistically significant. For Flexible Target, the proportion of those adhering over 90 percent are 11.0 and 12.0 percentage points higher than that of the Control group, with borderline statistical significance at conventional levels, p-value<0.1.

Figure 5. Mean monthly adherence, three-group comparison



NOTE: This figure shows mean electronically monitored adherence comparing two treatment groups and the control group, from 2 months before the intervention through the 9-month study period. The vertical line at month 0 denotes when the intervention started.

Figure 6. Proportion adhering over 90 percent, three-group comparison



NOTE: This figure shows the proportion of participants adhering over 90 percent (pills taken over pills prescribed using electronically monitored adherence data), comparing two treatment groups and the control group, from 2 months before the intervention through the 9-month study period. The black line at month 0 denotes when the intervention started.

Table 4. Impact of small incentives on mean adherence, group comparisons

	Pooled vs. Control		Fixed vs. Control		Flexible vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat x Post	0.077*	0.077*	0.040	0.036	0.114**	0.119**
	(0.039)	(0.039)	(0.041)	(0.040)	(0.048)	(0.049)
Treat	0.049		0.114**		-0.018	
	(0.051)		(0.055)		(0.061)	
Post	0.025		0.025		0.025	
	(0.033)		(0.033)		(0.033)	
Control mean (pre-intervention)	0.627	0.627	0.627	0.627	0.627	0.627
Individual + time fixed effects	No	Yes	No	Yes	No	Yes
Observations	2,104	2,104	1,406	1,406	1,360	1,360
Individuals	209	209	139	139	137	137
R-squared	0.038	0.716	0.060	0.734	0.026	0.722

NOTE: The outcome variable is mean adherence (percentage of pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. Standard errors (in parenthesis) are clustered at individual level. *** *p<0.01, ** p<0.05, * p<0.1*

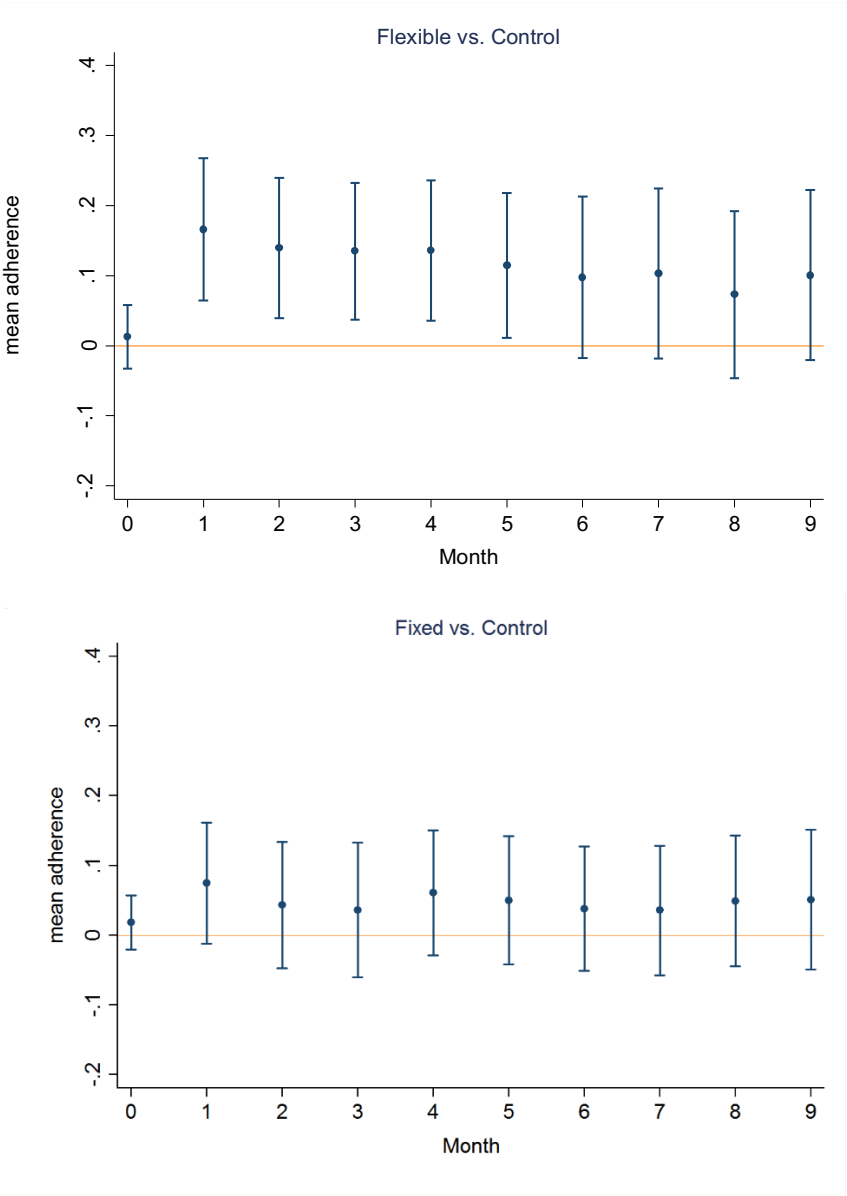
Table 5. Impact of small incentives on probability of adhering over 90 percent, group comparisons

	Pooled vs. Control		Fixed vs. Control		Flexible vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat x Post	0.062	0.070	0.016	0.021	0.110*	0.121*
	(0.049)	(0.051)	(0.055)	(0.057)	(0.060)	(0.063)
Treat	0.039		0.086		-0.009	
	(0.063)		(0.073)		(0.072)	
Post	0.038		0.038		0.038	
	(0.039)		(0.039)		(0.039)	
Control mean (pre-intervention)	0.323	0.323	0.323	0.323	0.323	0.323
Individual + time fixed effects	No	Yes	No	Yes	No	Yes
Observations	2,104	2,104	1,406	1,406	1,360	1,360
Individuals	209	209	139	139	137	137

NOTE: The outcome variable is an indicator=1 if participant has adherence over 90 percent (percentage of pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. Standard errors (in parenthesis) are clustered at individual level. *** *p<0.01, ** p<0.05, * p<0.1*

The impact of incentives on mean adherence for each month of the 9-month study period is shown in Figure 7. The coefficients β_t and standard errors are from regressions following equation (3), comparing each treatment arm to control separately. Appendix Table A2 reports the full regression estimates from this analysis. The effect of Flexible Target incentives on mean adherence is generally persistent over time in terms of size of the coefficient, though the point estimates for the last two months of the study are not statistically significant. The widening standard errors in later months are most likely due to fewer observations because of missing data as the study progressed (participants' attrition, leaving clinic without returning the EDM caps, losing their caps, caps breaking, among other reasons). The effects of Fixed Target incentives in each month also does not deviate substantially from the overall average effect over time, but none of the coefficients are significant at conventional levels. Like Flexible Target participants, there was a positive treatment effect when the study started; Fixed Target participants increased adherence by 7.5 pp in month 1 relative to the Control group, at 10 percent statistical significance. The treatment effect then drops to about 5 pp and is not statistically significant throughout the remainder of the study.

Figure 7. Dynamic treatment effects of incentives on mean adherence



NOTE: These figures show the point estimates and 95% confidence intervals for the effect of Flexible Target and Fixed Target incentives (separately) on mean adherence, in each month of the study. Coefficients are from regressions following equation (3). All coefficients are relative to the Control group, an interaction of a baseline time dummy (=1 if two months pre-intervention) and the treatment indicator. Standard errors are clustered at the individual level.

5.3 Heterogeneous effects of incentives on adherence, by participants' baseline adherence

The conceptual framework in Section 3 suggests that the amount of improvement from participants' increased motivation to act towards a goal may depend on the (initial) distance to that goal. To explore potential treatment heterogeneity by baseline adherence, we first descriptively assess 9-month aggregate adherence as well as treatment impact following the DID regression framework, by study group and participant adherence level at baseline: low-adherers (<60), medium adherers (greater than 60, less than 90) and high adherers (>90). There are 70 low adherers, 65 medium-adherers, and 74 high-adherers in the sample of 209 with measured data.[8]
To illustrate the distribution of adherence across the three study groups, we plot their k-density curves of 9-month aggregate adherence by baseline adherence level (Figure 8). Since this aggregate outcome measure does not account for baseline differences between groups, we also run regressions of treatment impact for Pooled vs. Control, Fixed Target vs. Control, and Flexible Target vs. Control, conditional on each level of baseline adherence, following the DID framework in equation 1.

The first hypothesis based on the conceptual framework is that for low-adherers, those in the Flexible Target group would outperform their Fixed Target counterparts as they have the option to select a more personally attainable target and may obtain greater psychological value and motivation from movements towards it. We find some evidence for this. In Figure 8, low adherers seem to benefit the most in the Flexible Target group - in this group, there is a rightward shift in the distribution for both incentive groups compared to control, but a pronouncedly larger shift for the flexible group. In the Control group the distribution for baseline low-adherers after 9 months in the study has bi-modal peaks around 25 and 45 percent; for the Fixed Target group it peaks around 60 percent; for the Flexible Target Group it is around 75 percent.

[8] We used different cut-offs as robustness checks and observed results that are in line with predictions. We always maintained the 90 percent cut-off as this is a widely agreed-on measure of clinical importance, but varied the cut-off for low adherence from 50-70 percent and settled on 60 percent as it gives us the largest relative sample sizes in each group.

Figure 9 shows mean adherence over time by baseline level of adherence. Among baseline low-adherers, there is a sizeable increase immediately after the intervention started among Flexible Target participants but not Fixed Target participants. This descriptive assessment is corroborated by the regression results. Appendix Table A3 shows estimates from the heterogeneous analysis and Figure 10 summarizes the results. Among low-baseline adherers, Flexible group participants increased adherence by 17.3 pp (p-value<.05). By contrast, the coefficient among low-adherers in the Fixed vs. Control regression is 5.1 pp and not statistically significant. The difference between these coefficients of 12.2 pp is large in terms of magnitude, but only borders on 10 percent significance (p-value= 0.115) likely because of the insufficient power due to the small sub-sample in these regressions (the study was not powered to detect subgroup differences).

In addition to improving mean adherence, low baseline adherers in the Flexible Target group also are more likely to achieve clinically optimal adherence. Figure 11 summarizes the results from the heterogeneous analysis of adherence over 90 percent (full regression estimates in Appendix Table A4). Compared to a Control-group pre-intervention mean of only 2 percent adhering at above 90 percent among low-baseline adherers, the proportion of low-baseline adherers who reached clinically optimal adherence increased by 15.5 pp for Flexible Target group (p-value <.05) compared to only a 0.3 pp, not statistically significant increase among the Fixed Target group. The difference between these two coefficients is statistically significant at the 10 percent level (Appendix Table A4, column 11).

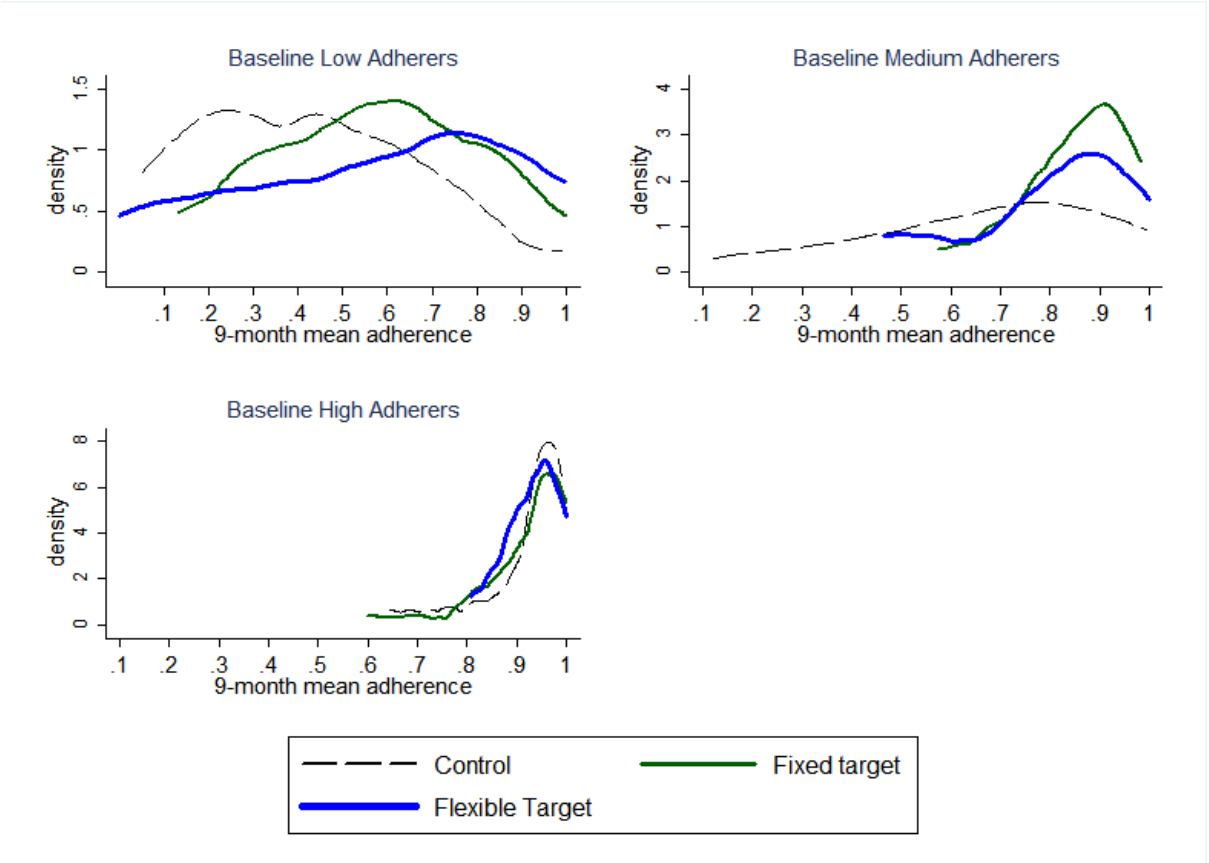
The second hypothesis arising from the conceptual framework is that medium adherers will show a pronounced increase in adherence in the Fixed Target group that may be even larger than in the Flexible group, in which medium-adherers have the possibility of opting for a (suboptimally low) adherence target. From the k-density curves for 9-month mean adherence in Figure 8, the distribution shift towards higher adherence values is slightly greater for Fixed Target than for Flexible Target; compared to Flexible Target participants, there is a higher density of Fixed Target participants around 90 percent. In a regression framework, we find that the average null effect for the Fixed Target group may be masking meaningful heterogeneity. The coefficient on *treatment x post* is statistically significant among those with baseline medium adherence, but not for high or low adherers in the Fixed Target group. Among this group, the Fixed Target arm improved

adherence by 17.3 percentage points compared to medium-adherers in the control group (p-value <0.01), where the control group had pre-intervention mean adherence of 78.4 percent in the medium adherence category. The corresponding treatment effect among medium-adherers comparing Flexible vs. Control is lower at 12.7 percentage points with significance at the 5 percent level. The two coefficients (Fixed and Flexible group impact among medium adherers) are not statistically different (col 12, Appendix Table A3). When looking at the proportion adhering over 90 percent in this baseline adherence category, we find that compared to a Control-group pre-intervention mean of only 12 percent adhering at above 90 percent, the proportion of medium-baseline adherers who reached clinically optimal adherence increased by 15.2 pp in the Flexible Target group, 12.2 in the Fixed Target group, though neither is statistically significant (Figure 11).

The third hypothesis is that high adherers may do slightly better in the Flexible Target group. However, as adherence of 90 percent is widely considered a clinically satisfactory standard and there is less room for further improvement up to 100, any differences are likely too small to be detected as significantly different with such small cell sizes. In Figure 9, we find that among baseline high-adherers, all three groups follow a similar distribution peaking at about 95 percent, and the Control group has a higher peak than the other groups. From the estimation results, the coefficient on *treat x time* for baseline high-adherers shows an adherence improvement of 3.5 pp for Flexible Target vs. Control compared to a smaller change of 0.3 pp among Fixed Target vs. Control, though none of these are statistically significant (cols 7 and 10, Appendix Table A3). The proportion adhering over 90 percent is also higher among Flexible Target participants compared to Control – an increase of 6.5 pp – compared to a 0.7 pp increase among Fixed Target participants, though neither is statistically significant (Figure 11).

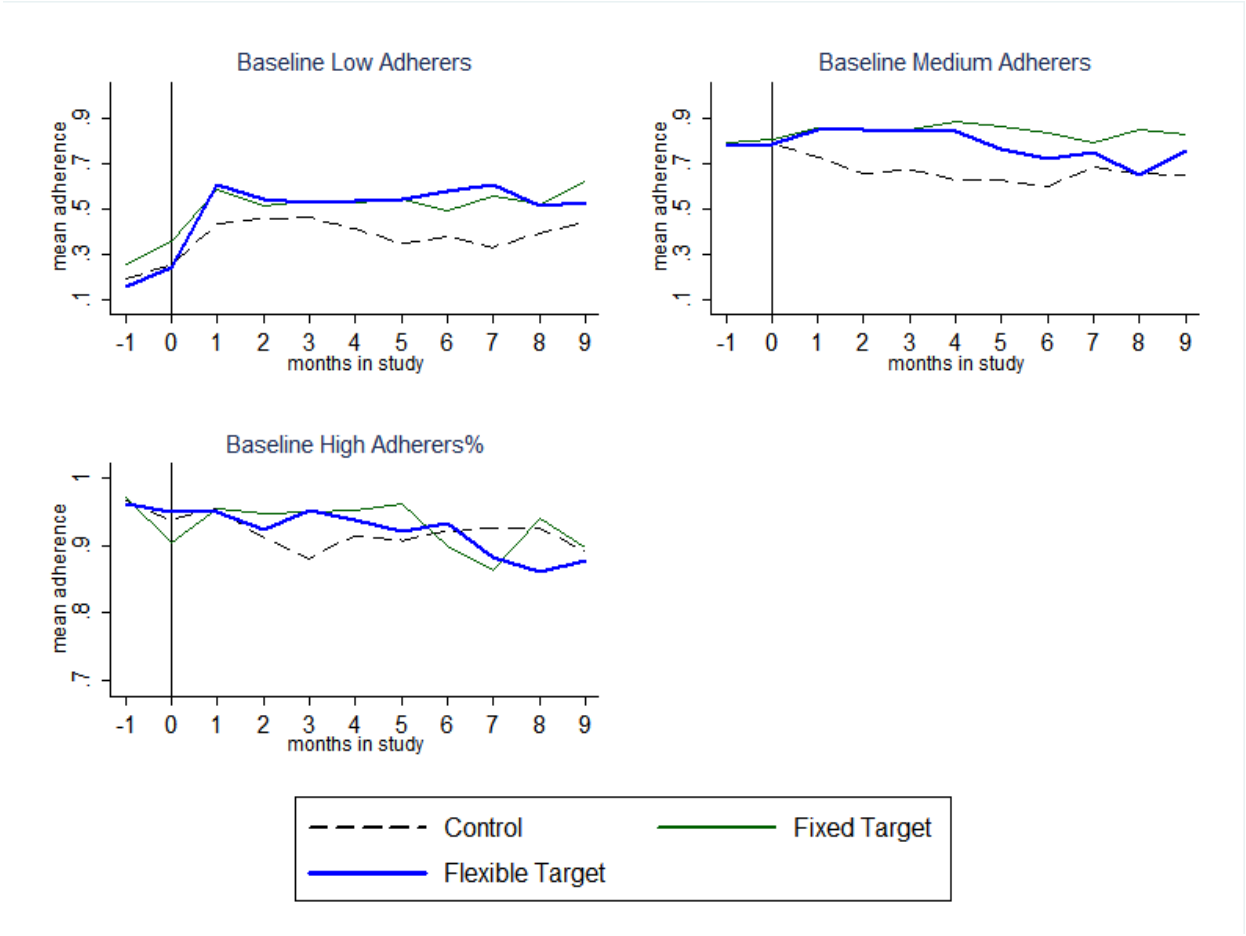
Sensitivity analyses using a different definition of "baseline" adherence to construct participant classifications are presented in Appendix Tables A7 and A8, along with an explanation of what was done.

Figure 8. K-density distributions of 9-month mean adherence



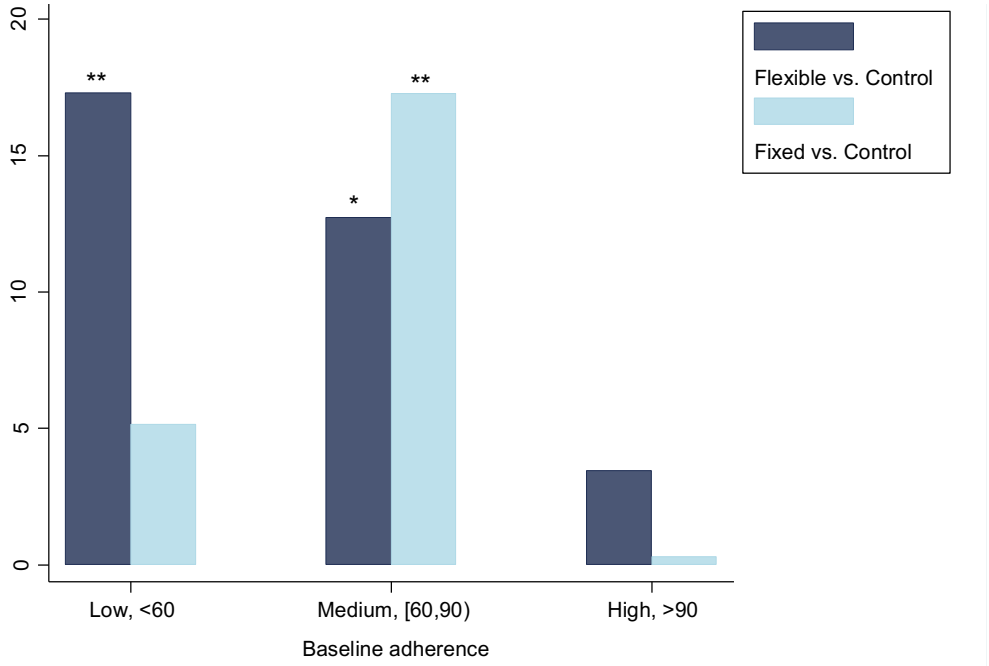
NOTE: These figures show the k-density distributions (epanechnikov kernel and equal bandwidths) of mean adherence averaged over 9 months. "Low", "Medium" and "High" refer to baseline/pre-intervention adherence levels, where Low= 60 percent or less, Medium= greater than 60 and less than 90, High= 90 and higher. The study sample consists of 209 participants with electronically monitored adherence data (Low- n=70, Medium- n=74, High- n=65).

Figure 9. Mean monthly adherence, by baseline adherence level



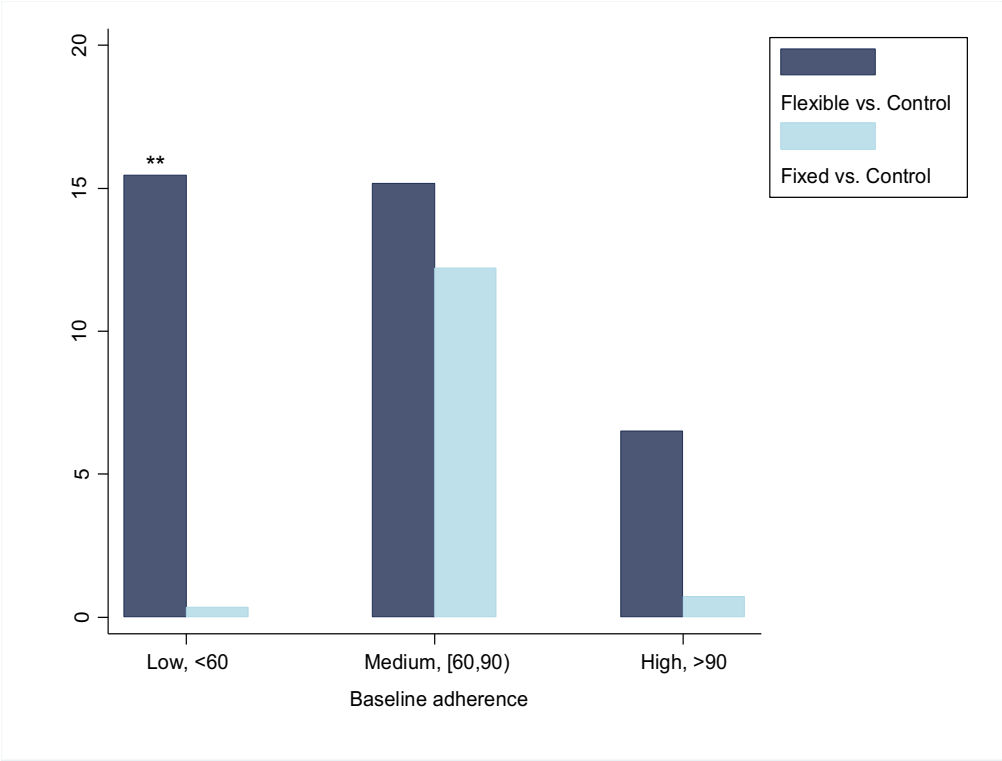
NOTE: These figures show mean electronically monitored adherence comparing two treatment groups and the control group, from 2 months before the intervention through the 9 month study period. The black line at month 0 denotes when the intervention started. "Low", "Medium" and "High" refer to baseline/pre-intervention adherence levels, where Low= 60 percent or less, Medium= greater than 60 and less than 90, High= 90 and higher. The study sample consists of 209 participants with electronically monitored adherence data (Low- n=70, Medium- n=74, High- n=65).

Figure 10. Summary of treatment effect of incentives on mean adherence, by baseline level of adherence



NOTE: This figure summarizes results from the analysis of heterogeneous treatment effects on mean adherence by baseline level of adherence. Dark blue bars show the magnitude of the effect of Flexible Target incentives on mean adherence; light blue bars show the magnitude of the effect of Fixed Target incentives on mean adherence. "Low", "Medium" and "High" refer to baseline/pre-intervention adherence levels, where Low= 60 percent or less, Medium= greater than 60 and less than 90, High= 90 and higher. The study sample consists of 209 participants with electronically monitored adherence data (Low- n=70, Medium- n=74, High- n=65). **p-value<0.1 **p-value<0.05.*

Figure 11. Summary of treatment effect of incentives on mean adherence, by baseline level of adherence



NOTE: This figure summarizes results from the analysis of heterogeneous treatment effects on adhering over 90%, by baseline level of adherence. Dark blue bars show the magnitude of the effect of Flexible Target incentives on proportion adhering over 90%; light blue bars show the magnitude of the effect of Fixed Target incentives on the same. "Low", "Medium" and "High" refer to baseline/pre-intervention adherence levels, where Low= 60 percent or less, Medium= greater than 60 and less than 90, High= 90 and higher. The study sample consists of 209 participants with electronically monitored adherence data (Low- n=70, Medium- n=74, High- n=65). **p-value<0.1 **p-value<0.05*

5.4 Impact on timely clinic attendance and viral suppression

Table 6 shows the results from linear probability regressions for the outcome "on-time within 7 days". Both Flexible and Fixed Target incentives increased the probability of making a timely clinic visit by this measure. Participants in the pooled incentive treatment groups had a higher rate of making an on-time visit within 7 days than the control group- an increase of 11.3 pp, compared to the Control group mean of 84.9 percent. In the specification including month and individual fixed effects, this decreases slightly to a 9.9 pp increase. The magnitude of the effect of Fixed and

Flexible Target incentives on a timely visit within 7 days are similar: a 10.3 –12.1 pp increase among Fixed Target participants relative to Control (p-value<.01) and an increase of 9.3 – 10.4 pp among Flexible Target participants (p-value<.05). For the "average" participant who has 7 scheduled appointments in 9 months and typically makes 6 of these, this effect translates to roughly making 1 extra timely appointment.

Table 6. Impact of small incentives on making an on-time clinic visit within 7 days

	Pooled vs. Control		Fixed vs. Control		Flexible vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat x Post	0.113***	0.099***	0.121***	0.103***	0.104**	0.093**
	(0.041)	(0.028)	(0.046)	(0.036)	(0.050)	(0.043)
Treat	-0.054		-0.038		-0.070*	
	(0.036)		(0.043)		(0.042)	
Post	-0.023		-0.023		-0.023	
	(0.033)		(0.033)		(0.033)	
Control mean (pre-intervention)	0.849	0.849	0.849	0.849	0.849	0.849
Individual + time fixed effects	No	Yes	No	Yes	No	Yes
Observations	1,606	1,606	1,067	1,067	1,066	1,066
Individuals	209	209	139	139	137	137
R-squared	0.010	0.196	0.014	0.184	0.005	0.193

NOTE: The outcome variable is a binary variable =1 if participants went to the clinic for a visit within 7 days of the scheduled appointment, and =0 if otherwise. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. Standard errors (in parenthesis) are clustered at individual level.
**** p<0.01, ** p<0.05, * p<0.1*

Because baseline adherence was higher among Fixed Target participants relative to control, we cannot assume that post-intervention differences in viral suppression between these groups reflect true impact, particularly lacking baseline viral load data. Estimating endline differences in viral suppression between Flexible and Control group was not subject to this issue, as the two groups were balanced at baseline. While in our analysis of viral load we control for baseline adherence, the comparisons between Fixed and Control should still be interpreted with caution.

198 participants were reached for viral load tests at the end of the study (89 percent in Control, 97 percent in Fixed Target, and 89 percent in Flexible Target). Viral suppression was high across all three groups: 83 percent in the control group, 85 percent in the Flexible Target group, and 91

percent in the Fixed Target group had HIV RNA less than 400 copies per mL. There were no statistically significant differences in viral suppression between either treatment group compared to control (Table 7), although it is worth noting that 8 pp more participants in the Fixed Target group were virally suppressed compared to Control, likely reflecting the higher adherence rates among this group even at baseline. Subgroup analysis by baseline adherence level did not reveal statistically significant differences in rates of viral suppression between the three groups. Analyses adjusting for baseline adherence and subgroup analyses are in Appendix Tables A5 and A6. While there is a strong link attested in the literature between adherence and viral load, we do not find a result on viral suppression in our analysis despite strong evidence for behavioral change. This could be due to high (unobserved) baseline levels of viral suppression and/or a relatively short intervention period that may not have sufficed for behavioral change to translate into biological outcomes. Viral suppression, while the clinical outcome of interest in HIV care, is also not fully under the control of the individual as other, in this study unobserved inputs such as food intake and genetic variables can also influence suppression.

Table 7. Impact of small incentives on viral suppression

	Proportion virally suppressed			Unadjusted difference		Unadjusted p-value		Adjusted difference		Adjusted p-value	
	Control	Fixed Target	Flexible Target	Fixed - Control	Flexible-Control	Fixed vs. Control	Flexible vs. Control	Fixed-Control	Flexible-Control	Fixed- Control	Flexible vs. Control
<400 copies/ml	0.83 (0.74 - 0.92)	0.91 (0.84 - 0.98)	0.85 (0.76 - 0.94)	0.08 (-0.03 - 0.20)	0.02 (-0.11 - 0.14)	0.16	0.81	0.03 (-0.08 - 0.14)	0.02 (-0.10 - 0.14)	0.58	0.75
<40 copies/ml	0.65 (0.53 - 0.77)	0.78 (0.68 - 0.88)	0.66 (0.54 - 0.78)	0.13 (-0.02 - 0.29)	0.02 (-0.15 - 0.18)	0.09	0.86	0.07 (-0.08 - 0.21)	0.04 (-0.13 - 0.20)	0.36	0.66

NOTE: Adjusted analysis controls for baseline adherence; 95% confidence interval shown in parenthesis

6. Internal validity and robustness checks

In this section we address potential internal validity concerns and present some robustness and sensitivity analyses.

6.1 Internal validity

In the primary analysis above, we use all available EDM adherence data in each month observed. Because EDM data is extracted during every clinic visit, there may be missing data in some months either due to the participant forgetting to bring their cap during that visit or requiring a replacement cap due to malfunction or loss. Hence, a possible threat to the internal validity of the results would be if the missing observations in each month resulted in a non-random selection of treatment versus control participants.

We assess both *differential attrition* (if missingness is different across treatment and control groups) and *selective attrition* (if the mean of observable baseline characteristics differ across treatment and control responders). For the first, we regress an indicator for missingness on treatment status for all study months and include month fixed effects (Appendix Table A9). Missingness is higher among the control group when compared to Fixed Target (9.8pp higher than control group average of 9.8 percent and significant at 5 percent level). The difference between missing rate for Flexible Target and Control is comparable (8.0 pp vs 5.9 pp) and not statistically significant (p-value= 0.683). For the second test of differential attrition, we compare the two treatment groups to the control group along baseline variables listed in Table 1 (including baseline adherence) for every month of the study to determine whether the differential attrition rates noted above among Fixed vs. Control group may be driven by selection on observables. This analysis checks whether the composition of the sample changed due to attrition, arguably a more telling indication of missingness bias than differential attrition. In other words, differential attrition may not be as concerning if missingness is plausibly unrelated to the outcome of interest.

This comparison for Fixed vs. Control and Flexible vs. Control over 9 months yields a total of

18 balancing tables. As a parsimonious way of displaying these results, we show graphs of ranked p-values associated with the mean tests in each month for the three comparisons (Appendix Figures A4-6). In Appendix Figures A4-6, the horizontal red line marks p-value at 0.05. For each of the three sets of comparisons, none of the baseline balancing variables fall below this line, suggesting that treatment and control groups are balanced along these observables in each period of the study.

These results suggest missing data in each period are not likely causing time-varying selection bias in the sample. If differential attrition based on time-*invariant* unobservable characteristics remain, these are controlled for in our analyses, which include individual fixed effects.

6.2 Robustness check: strategic bottle openings

Another concern is that because game eligibility is contingent on EDM adherence as measured by bottle openings, participants may open their pill bottle without taking out medication. There is good reason to believe this is not a major issue within our study sample. First, study coordinators did not explain the exact mechanisms of the EDM cap, and qualitative evidence suggests that most participants were only vaguely aware how their adherence is calculated, apart from knowing that the cap somehow observes their pill-taking and that the more pills they take the higher their adherence will be; to extract the data, the study coordinators simply placed the bottle upside-down against a flat reader. Participants were not necessarily aware that each opening counted as one dose. Second, if forced openings were common study coordinators would see multiple sequential openings on the EDM charts (Appendix Figure A1 and A2 shows an example of a EDM chart) every time they checked the participants' data. Such events were however not reported, and study coordinators expressed a strong opinion that participants were not trying to game the system, or simply lacked sufficient detail of how adherence was measured to do so. Third, if indeed a participant did open their cap multiple times in a day this would be discounted in our analysis as we top-code daily doses at the number of doses in a daily regimen (i.e. we do not count doses beyond 2 for those on a twice-daily regimen, or above 1 for those on a single-pill regimen).

In addition to these reasons against EDM data not accurately reflecting actual pill-taking behavior, we can explicitly test for evidence of gaming the system in the following way: as discussed, participants were told that prize drawings were conditional on adherence *within the last month* before their game visit. In the EDM extraction software, the study coordinators can select the exact temporal window to check adherence and hence if a clinic visit corresponding to a game occurs for example on March 20, they would select February 20 – March 19 as the window. They reinforced this by informing participants, "*in the last month, you did [not] reach your target*." This allows for a robustness check: we discard the last four weeks before clinic visits and only examine those weeks in which adherence was not explicitly incentivized. Considering that there may be a level effect across all groups resulting from a final push to adhere well before a clinic visit (the so-called white coat effect) we also discard such observations among the control group. This robustness check ensures that any potential strategic bottle openings wherein a pill is not actually taken out is not driving the intervention effect. If results are robust to this analysis, then it also suggests that participants were changing their behavior even when there was no direct, extrinsic incentive to do so.

For this robustness analysis, we use adherence at the week level to minimize discarding too much information outside the 4-week period before a visit. For instance, under the "monthly" adherence approach if a clinic visit occurred in month 2.5 since recruitment then both month 2 and month 1 data will be discarded; using weekly adherence allows us to keep weeks 1 and 2. Note that some participant's appointments are monthly and will be dropped from this analysis entirely, hence estimates will generally be noisier than the full sample analysis.

Appendix Table A10 shows the results of this analysis. Excluding the 4 weeks before each visit does not alter the results presented above in any significant way as there remains a statistically significant effect comparing Flexible Target versus Control and no effect for the Fixed Target group. The coefficient in column 3, an increase in mean adherence of 10.9 percentage points is of similar magnitude as that found in the main analysis using monthly adherence and the full sample. The pooled effect is also of similar magnitude – 6.6 percentage points, significant at 10%. Overall, results are robust to this sub-sample analysis.

7. Discussion and limitations

Our study investigates a question that needs to be decided in almost any intervention using conditional rewards: how should the conditionality threshold be set? Using data from clients in HIV care in Uganda, we find that flexible goals decided on by participants improve medication adherence more than setting a fixed, relatively high threshold for participants equally. Another finding is that the starting point relative to a goal mediates treatment impact as predicted by a conceptual framework based on Prospect Theory. Participants who have low adherence at baseline see large and statistically significant increases in adherence when they can choose their own target, compared to no effect when given a fixed target that is relatively distant to their starting position.

While we find that incentives in the form of small mobile airtime lottery rewards resulted in significant improvements in medication adherence and timely clinic attendance rates among adolescents and young adults in HIV care in Uganda, these improvements in behavior did not translate to meaningful differences in viral suppression at 9 months post-intervention. The lack of impact on viral outcomes is disappointing, but could be due to several factors. First, the follow-up period may not represent a sufficient time frame for behavior change to translate to changes in health markers. Second, initial levels of viral load suppression was already quite high even among the control group, making it difficult to detect changes over time. High viral suppression compared to the lower average adherence at baseline also suggest that the ART regimens taken by patients at the study clinic were relatively forgiving- meaning that viral suppression was attainable with a wider range of adherence than previously thought, either due to the greater potency of newer regimens or the heterogeneity in their effect.

We also collected qualitative data on potential pathways through which the intervention may have worked: at the end of our study Flexible Target participants were asked the open-ended question "During your game visits, how did you decide on an adherence level to choose?". Frequent responses were that they considered their starting point and subsequent achievements in making their choice, with several suggesting that reaching a proximal goal reinforced their motivation to aim for a subsequent higher one. This is in line with prior research documenting that individuals actively monitor their progress (Carver and Scheier 1990) and adjust efforts accordingly (Kivetz

et al. 2006, Nunes and Drèze 2006). One interpretation of this behavior is that the flexible target level acted as a commitment device, as when the level was not reached the participant knew that s/he would be ineligible for the reward, which could also trigger regret aversion. For example, one individual reported "The first time when I had chosen 80, I was glad to know that I had achieved it, and I set another goal, higher than that because I wanted to see if I can test myself if I could meet it."

Responses among some high-baseline adherence participants suggest that they anchored their goal choice to their pre-intervention behavior (e.g. "I wanted it to be 100 because I love my life and I have no problem taking my medicine.") Other responses suggested that participants had a strong intrinsic motivation to adhere well; these participants saw the Flexible Target intervention as a way to further boost their motivation: "I realized that taking medicine well is okay as you can avoid falling sick from malaria and such other illnesses. So, when this program was introduced, I liked it so much and I made sure I continue taking my medicines well." MacCarthy et al. (2018) reports findings from our full qualitative analyses.

This study has several limitations. While we applied the value function framework as a useful way to guide the analysis of heterogeneity, an important caveat is that the literature on Prospect Theory and goal-setting is largely silent on how goals should be assigned: whether they should be participatory or externally imposed. In this study, the goals selected by participants themselves demonstrated commitment and motivation to reach targets that were, on average, higher than participants' pre-intervention adherence. This is not surprising considering the previous literature on the key role of intrinsic motivation in interventions using incentives. Gneezy's review of the literature surrounding incentives suggests that incentives work best at motivating behavior change when they are tied to controllable outcomes and the incentives are aligned with intrinsic motivation, reinforcing what individuals already want to do (Gneezy et al. 2011). The literature on commitment contracts also demonstrates that the individual's objective is to maximize the chances of reaching a desired behavior, rather than (solely) maximize income. This is consistent with the idea that small incentives operate through a behavioral, rather than an income channel – providing a psychological reward to a behavior rather than one that changes budget constraints. However since in this study the goal choice was endogenous we are unable to fully isolate the effect of "sub-

goaling". A strict test of "sub-goaling" against fixed goals requires the implementer to set a series of externally imposed, individualized goals that are adjusted in each game depending on previous performance. This is an important avenue for future research.

Since the goal choice was endogenous, another mechanism that can drive participant motivation is goal ownership. Giving participants the option to choose their own targets in the Flexible Target arm may instill a sense of goal ownership, which is linked to self-efficacy- a person's self-belief in his or her ability to perform specific tasks - and motivation to improve (Sue-Chan and Ong 2002). The same objective target of 90 percent could elicit greater sense of self-efficacy in the participant who actively selected it compared to one to whom it was assigned. We assumed this to be a constant, which would be the case if greater intrinsic motivation exerts an intercept shift across all participants. However, if low adherers derive greater self-efficacy from goal ownership than middle and high adherers, then the relative effects are not disentangled and warrant further inquiry.

In the psychology literature, both sub-goaling and participative goal-setting are believed to improve performance by boosting self-efficacy. Stock et al. (1990) found that setting subgoals boosted initial self-efficacy perceptions, self-satisfaction with performance, and subsequent task persistence; this change in self-efficacy mediated the effects of goal attainment. In a related strand of literature participative goals also result in higher self-efficacy, goal commitment and performance, compared to assigned goals (Erez et al. 1985, Latham et al. 1994, Sue-Chan and Ong 2002). Although the relative impact of participative versus strict sub-goaling on performance, and their interaction with an individuals' starting point, is yet unclear, a feature common to both interventions is the allowance (either assigned or chosen by the participant) of proximal goals, which can provide an initial boost in self-efficacy to low-performers facing the 'starting problem'. The results from this study suggest this to be a plausible mechanism driving our results in the Flexible Target arm.

Another limitation of our study is that the analysis is not powered for sub-sample regressions in our heterogeneous analysis. The magnitudes of coefficients are often in line with our predictions and robust to different specifications; however, due to the insufficient power the differences

between coefficients are often not statistically significant. A future research study with a larger sample would establish whether our subgroup findings hold, with greater precision.

A policy implication of our findings is that for patients struggling with adherence problems, a graduated approach may be more effective than placing a strict emphasis on 90 or 100 percent, the typical approach in current counselling practices, at least as an initial step followed by subsequently higher goals. The design of incentive interventions could also benefit from accounting for the relative positions of participants, particularly in contexts where goals are not a discrete, binary outcome, but exist along a continuum and require repeated behaviors over time (adherence for chronic diseases, test scores, etc.).

Conclusion

Incentives are increasingly used to encourage healthy behaviors, and there is a growing literature investigating different dimensions influencing the effectiveness of incentives when designing such interventions. This study shows that an important yet to date neglected consideration is the participant's starting-point performance. If motivation and subsequent effort are starting-point dependent, then setting an absolute, fixed threshold may not have a uniformly positive effect. We find that for those with initially low adherence levels setting the conditionality threshold in a way that allows participants to choose their own, nearer targets resulted in substantially larger adherence improvements than imposing a high, fixed target; for medium baseline performers, a high, fixed target was more effective. A plausible mechanism to explain this is that for low adherers, commitment to a goal and subsequent effort to achieve it depends critically on the perception that the goal is attainable. While this study directly informs the use of incentives to help the specific problem of ART adherence among youth, the conceptual framework discussed and our empirical findings have broader applicability to the design of incentive interventions in general and point towards a deserving avenue for future research.

Competing interests statement - Haijing Huang

This research was supported by National Institues of Health grant R01HD074925-04S1.
Haijing Huang has no conflicts of interest to disclose.

Competing interests statement - Sebastian Linnemayr

This research was supported by National Institues of Health grant R01HD074925-04S1.
Sebastian Linnemayr has no conflicts of interest to disclose.

Appendix

Appendix Table A1. Stratified Randomization, by Baseline Adherence Quartiles

	Control	Fixed Target	Flexible Target	Total
Overall	67.30%	66.80%	68.20%	67.50%
N	75	76	78	229
Strata 1 [0,p25]	19.90%	15.20%	24.00%	19.80%
N	19	19	20	58
Strata 2 (p25-p50]	65.10%	66.60%	65.50%	65.70%
N	19	19	19	57
Strata 3 (p50, p75]	88.40%	87.80%	86.80%	87.60%
N	19	20	20	59
Strata 4 (p75,1]	97.40%	98.20%	97.90%	97.90%
N	18	18	19	55

Appendix Table A2. Treatment effects over time

	(1) Pooled	(2) Fixed vs. Control	(3) Flexible vs. Control
preperiod1 x treat	0.016	0.018	0.013
	(0.015)	(0.020)	(0.023)
month 1 x treat	0.120***	0.075*	0.166***
	(0.041)	(0.044)	(0.052)
month 2 x treat	0.091**	0.043	0.139***
	(0.042)	(0.046)	(0.051)
month 3 x treat	0.085**	0.036	0.135***
	(0.042)	(0.049)	(0.050)
month 4 x treat	0.098**	0.061	0.136***
	(0.041)	(0.046)	(0.051)
month 5 x treat	0.083*	0.050	0.115**
	(0.042)	(0.047)	(0.053)
month 6 x treat	0.069	0.038	0.098*
	(0.044)	(0.045)	(0.059)
month 7 x treat	0.070	0.036	0.103*
	(0.045)	(0.048)	(0.062)
month 8 x treat	0.064	0.049	0.073
	(0.045)	(0.048)	(0.061)
month 9 x treat	0.078*	0.051	0.101

	(0.046)	(0.051)	(0.062)
post	0.025	0.025	0.025
	(0.033)	(0.033)	(0.033)
treat	0.041	0.105*	-0.025
	(0.052)	(0.056)	(0.063)
Observations	2,104	1,406	1,360
R-squared	0.039	0.060	0.029

NOTE: The outcome variable is mean adherence (percentage of pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. Coefficients shown are from regressions following equation (3) in the main text. All month x treat coefficients are relative to the reference category, preperiod2 x treat (interaction of an indicator for two periods before intervention started and the treatment indicator). Standard errors are clustered at the individual level. **** p<0.01, ** p<0.05, * p<0.1*

Appendix Table A3. Heterogeneous treatment effects on mean adherence, by baseline adherence level

	Pooled vs. Control			Fixed vs. Control			Flexible vs. Control			Difference (Flexible - Fixed)		
	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low (col 8-5)	Medium (col 9-6)	High (col 10-7)
Treat x Post	0.124	0.153***	0.017	0.051	0.173***	0.003	0.173**	0.127*	0.035	0.122	-0.045	0.032
	(0.076)	(0.054)	(0.028)	(0.089)	(0.054)	(0.031)	(0.073)	(0.064)	(0.030)	(0.077)	(0.044)	(0.025)
Treat	0.019	0.009	-0.010	0.086	0.017	-0.004	-0.026	-0.000	-0.019			
	(0.053)	(0.027)	(0.013)	(0.067)	(0.030)	(0.013)	(0.065)	(0.032)	(0.019)			
Post	0.184***	-0.125**	-0.040	0.184***	-0.125**	-0.040	0.184***	-0.125**	-0.040			
	(0.059)	(0.050)	(0.025)	(0.059)	(0.050)	(0.025)	(0.052)	(0.050)	(0.025)			
Control mean	0.225	0.784	0.956	0.225	0.784	0.956	0.225	0.784	0.956			
Observations	681	667	756	418	462	526	505	412	443			
Individuals	70	65	74	43	45	51	52	41	44			
R-squared	0.130	0.088	0.007	0.109	0.140	0.011	0.134	0.062	0.009			

NOTE: The outcome variable is mean adherence (percentage of pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. "Low" is baseline adherence <60%, "Medium" is baseline adherence between 60 and 90% and "High" is baseline adherence >90%. Constant refers to the mean adherence in the pre-intervention months, for the control group. Standard errors in parenthesis are clustered at the individual level. **** p<0.01, ** p<0.05, * p<0.1*

Appendix Table A4. Heterogeneous treatment effects on proportion adhering over 90, by baseline adherence level

	Pooled vs. Control			Fixed vs. Control			Flexible vs. Control			Difference (Flexible - Fixed)		
	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low (col 8-5)	Medium (col 9-6)	High (col 10-7)
Treat x Post	0.094	0.135*	0.032	0.003	0.122	0.007	0.155**	0.152	0.065	0.151*	0.030	0.058
	(0.074)	(0.075)	(0.091)	(0.085)	(0.083)	(0.097)	(0.075)	(0.104)	(0.099)	(0.078)	(0.109)	(0.070)
Treat	-0.020	0.017	-0.044	-0.020	0.027	-0.036	-0.020	0.006	-0.055			
	(0.020)	(0.058)	(0.054)	(0.020)	(0.067)	(0.060)	(0.067)	(0.068)	(0.066)			
Post	0.089	0.123**	-0.147*	0.089	0.123**	-0.147*	0.089*	0.123**	-0.147*			
	(0.056)	(0.054)	(0.084)	(0.056)	(0.054)	(0.084)	(0.054)	(0.054)	(0.084)			
Control mean	0.020	0.119	0.902	0.020	0.119	0.902	0.020	0.119	0.902			
Observations	681	667	756	418	462	526	505	412	443			
Individuals	70	65	74	43	45	51	52	41	44			
R-squared	0.042	0.054	0.014	0.019	0.050	0.019	0.067	0.056	0.013			

NOTE: The outcome variable is a binary indicator =1 if in adherence is at or above 90 percent (pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. "Low" is baseline adherence <60%, "Medium" is baseline adherence between 60 and 90% and "High" is baseline adherence >90%. Control mean refers to the mean adherence in the pre-intervention months, for the control group. Standard errors in parenthesis are clustered at the individual level. *** *p<0.01, ** p<0.05, * p<0.1*

Appendix Table A5. Impact of small incentives on viral suppression (vl<40 copies/ml)

	Proportion virally suppressed			Unadjusted difference		Unadjusted p-value		Adjusted difference		Unadjusted p-value	
	Control	Fixed Target	Flexible Target	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs. Control
Complete case analysis	0.65 (0.53 - 0.77)	0.78 (0.68 - 0.88)	0.66 (0.54 - 0.78)	0.13 (-0.02 - 0.29)	0.02 (-0.15 - 0.18)	0.09	0.86	0.07 (-0.08 - 0.21)	0.04 (-0.13 - 0.20)	0.36	0.66
Subgroup analyses											
Low adherers (n=67)	0.38 (0.17 - 0.58)	0.50 (0.24 - 0.76)	0.64 (0.44 - 0.84)	0.12 (-0.19 - 0.44)	0.27 (-0.02 - 0.55)	0.43	0.07				
Medium adherers (n=59)	0.70 (0.48 - 0.92)	0.81 (0.63 - 0.99)	0.61 (0.36 - 0.86)	0.11 (-0.17 - 0.39)	-0.09 (-0.41 - 0.23)	0.43	0.58				
High adherers (n=72)	0.90 (0.77 - 1.04)	0.93 (0.83 - 1.03)	0.73 (0.53 - 0.93)	0.03 (-0.13 - 0.19)	-0.18 (-0.42 - 0.06)	0.74	0.14				

NOTE: Adjusted analysis controls for baseline adherence

Appendix Table A6. Impact of small incentives on viral suppression (vl<40 copies/ml)

	Proportion virally suppressed			Unadjusted difference		Unadjusted p-value		Adjusted difference		Unadjusted p-value	
	Control	Fixed Target	Flexible Target	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs Control	Fixed vs Control	Flexible vs. Control
Complete case analysis	0.65 (0.53 - 0.77)	0.78 (0.68 - 0.88)	0.66 (0.54 - 0.78)	0.13 (-0.02 - 0.29)	0.02 (-0.15 - 0.18)	0.09	0.86	0.07 (-0.08 - 0.21)	0.04 (-0.13 - 0.20)	0.36	0.66
Subgroup analyses											
Low adherers (n=67)	0.38 (0.17 - 0.58)	0.50 (0.24 - 0.76)	0.64 (0.44 - 0.84)	0.12 (-0.19 - 0.44)	0.27 (-0.02 - 0.55)	0.43	0.07				
Medium adherers (n=59)	0.70 (0.48 - 0.92)	0.81 (0.63 - 0.99)	0.61 (0.36 - 0.86)	0.11 (-0.17 - 0.39)	-0.09 (-0.41 - 0.23)	0.43	0.58				
High adherers (n=72)	0.90 (0.77 - 1.04)	0.93 (0.83 - 1.03)	0.73 (0.53 - 0.93)	0.03 (-0.13 - 0.19)	-0.18 (-0.42 - 0.06)	0.74	0.14				

NOTE: Adjusted analysis controls for baseline adherence

Appendix Table A7. Heterogeneous treatment effects on mean adherence, by baseline adherence level – sensitivity analysis

	Pooled vs. Control			Fixed vs. Control			Flexible vs. Control			Difference (Flexible - Fixed)		
	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low (col 8-5)	Medium (col 9-6)	High (col 10-7)
Treat x Post	0.130*	0.139**	0.051	0.079	0.112*	0.044	0.167**	0.167**	0.059	0.088	0.056	0.014
	(0.074)	(0.066)	(0.036)	(0.082)	(0.064)	(0.048)	(0.075)	(0.081)	(0.035)	(0.080)	(0.060)	(0.047)
Treat	0.004	-0.009	-0.032	0.086	0.017	-0.004	-0.061	-0.062	-0.041			
	(0.060)	(0.052)	(0.025)	(0.067)	(0.030)	(0.013)	(0.066)	(0.066)	(0.031)			
Post	0.163***	-0.083	-0.049*	0.184***	-0.125**	-0.040	0.163***	-0.083	-0.049*			
	(0.055)	(0.059)	(0.026)	(0.059)	(0.050)	(0.025)	(0.051)	(0.059)	(0.026)			
Control mean	0.247	0.752	0.951	0.247	0.752	0.951	0.247	0.752	0.951			
Observations	597	796	711	392	534	480	435	493	432			
Individuals	61	80	68	40	53	46	45	51	41			
R-squared	0.122	0.040	0.005	0.131	0.067	0.008	0.109	0.028	0.008			

NOTE: This table shows results from a sensitivity analysis where we define “baseline” adherence using average adherence in the 6-month period before the study started. The outcome variable is mean adherence (percentage of pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. "Low" is baseline adherence <60%, "Medium" is baseline adherence between 60 and 90% and "High" is baseline adherence >90%. Constant refers to the mean adherence in the pre-intervention months, for the control group. Standard errors in parenthesis are clustered at the individual level. *** *p<0.01, ** p<0.05, * p<0.1*

Appendix Table A8. Heterogeneous treatment effects on proportion adhering over 90, by baseline adherence level – sensitivity analysis

	Pooled vs. Control			Fixed vs. Control			Flexible vs. Control			Difference (Flexible - Fixed)		
	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	Low	Medium	High	Low	Medium	High	Low	Medium	High	Low (col 8-5)	Medium (col 9-6)	High (col 10-7)
Treat x Post	0.102*	0.077	0.053	0.055	0.071	-0.012	0.139**	0.085	0.133	0.084	0.014	0.146
	(0.061)	(0.085)	(0.105)	(0.077)	(0.094)	(0.111)	(0.069)	(0.105)	(0.123)	(0.075)	(0.102)	(0.099)
Treat	-0.021	-0.023	-0.004	-0.021	-0.031	0.062	-0.021	-0.015	-0.085			
	(0.021)	(0.091)	(0.085)	(0.021)	(0.097)	(0.090)	(0.062)	(0.106)	(0.109)			
Post	0.045	0.093	-0.084	0.045	0.093	-0.084	0.045	0.093	-0.084			
	(0.038)	(0.069)	(0.093)	(0.038)	(0.069)	(0.094)	(0.047)	(0.069)	(0.094)			
Control mean	0.021	0.255	0.789	0.021	0.255	0.789	0.021	0.255	0.789			
Observations	597	796	711	392	534	480	435	493	432			
Individuals	61	80	68	40	53	46	45	51	41			
R-squared	0.037	0.017	0.004	0.016	0.014	0.011	0.054	0.018	0.004			

NOTE: This table shows results from a sensitivity analysis where we define “baseline” adherence using average adherence in the 6-month period before the study started. The outcome variable is a binary indicator =1 if in adherence is at or above 90 percent (pills taken out of pills prescribed) in each month. Treat is either pooled (=1 if individual is in assigned to any incentives treatment, 0 if control) or separated by treatment (=1 if Fixed, =0 if Control; =1 if Flexible, 0 if Control). Post =1 after the intervention started. "Low" is baseline adherence <60%, "Medium" is baseline adherence between 60 and 90% and "High" is baseline adherence >90%. Control mean refers to the mean adherence in the pre-intervention months, for the control group. Standard errors in parenthesis are clustered at the individual level. **** p<0.01, ** p<0.05, * p<0.1*

Appendix Table A9. Differential attrition analysis

	Flexible vs. Control	Fixed vs. Control
Treatment	-0.021	-0.098***
	(0.020)	(0.018)
Month 2	0.021	0.028
	(0.043)	(0.038)
Month 3	0.035	0.035
	(0.043)	(0.038)
Month 4	0.062	0.056
	(0.043)	(0.038)
Month 5	0.083*	0.063*
	(0.043)	(0.038)
Month 6	0.104**	0.077**
	(0.043)	(0.038)
Month 7	0.132***	0.105***
	(0.043)	(0.038)
Month 8	0.181***	0.140***
	(0.043)	(0.038)
Month 9	0.208***	0.161***
	(0.043)	(0.038)
Observations	1,296	1,287
R-squared	0.034	0.045

NOTE: Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
"Treatment" refers to a binary variable =1 if participant belongs to Flexible group (col 1) or Fixed group (col 2) and =0 if they are in the Control group.

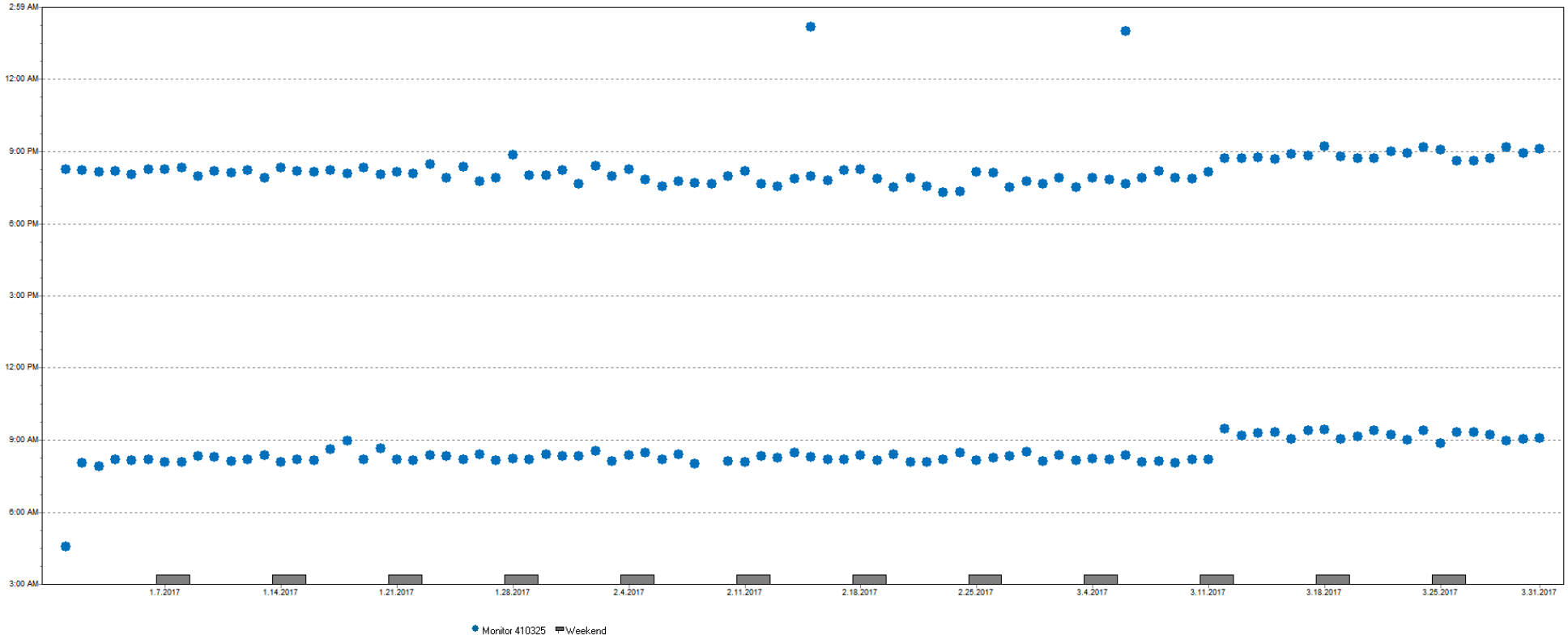
Appendix Table A10. Intervention Treatment Effect on Mean Adherence over 36 weeks, Robustness Check

	Pooled vs. Control		Fixed vs. Control		Flexible vs. Control	
	(1)	(2)	(3)	(4)	(5)	(6)
Treat x Post	0.066*	0.076**	0.022	0.035	0.109**	0.119**
	(0.040)	(0.037)	(0.041)	(0.038)	(0.049)	(0.047)
Treat	0.057		0.125**		-0.012	
	(0.052)		(0.056)		(0.063)	
Post	0.055*		0.055*		0.055*	
	(0.033)		(0.033)		(0.033)	
Individual + Time FE	No	Yes	No	Yes	No	Yes
Observations	5,784	5,784	3,841	3,841	3,680	3,680
Individuals	209	209	139	139	137	137
R-squared	0.041	0.655	0.056	0.670	0.034	0.670

NOTE: Robust standard errors in parentheses

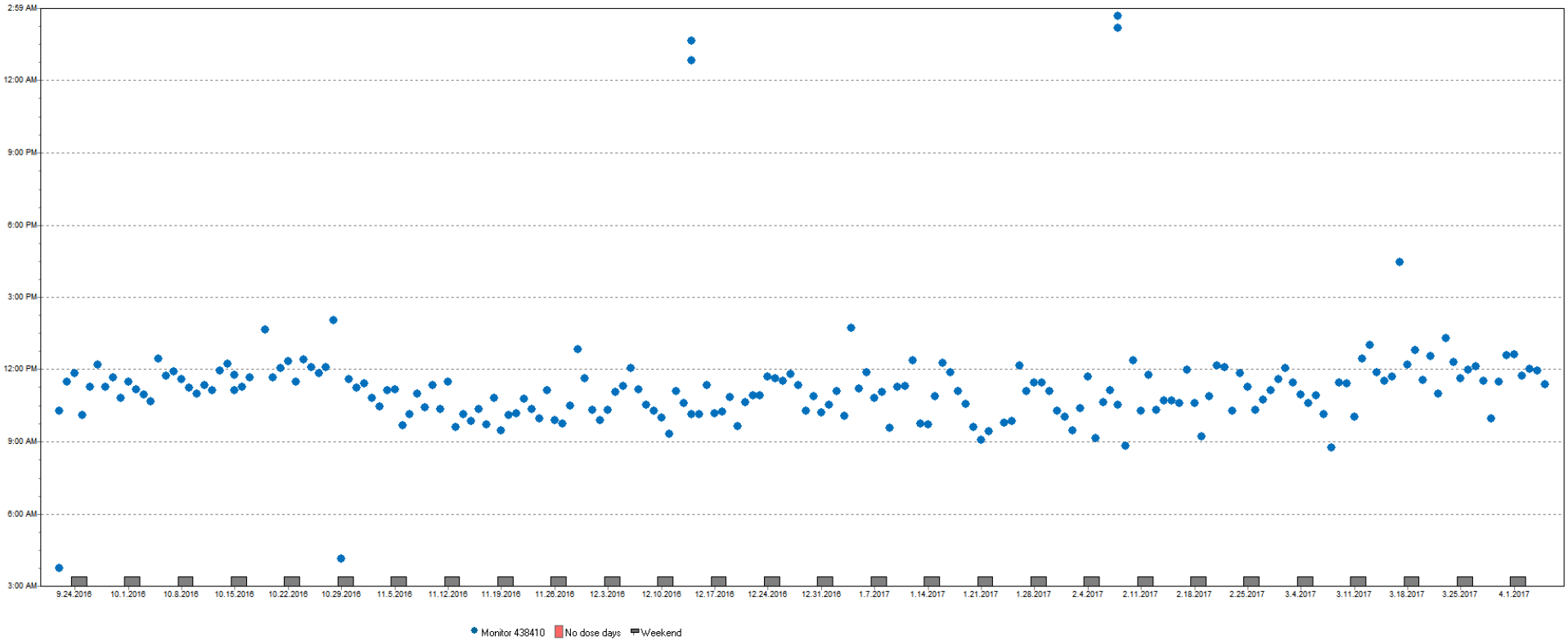
**** p<0.01, ** p<0.05, * p<0.1*

Appendix Figure A1. EDM Adherence from Jan 1 to April 1 2017, twice-daily regimen



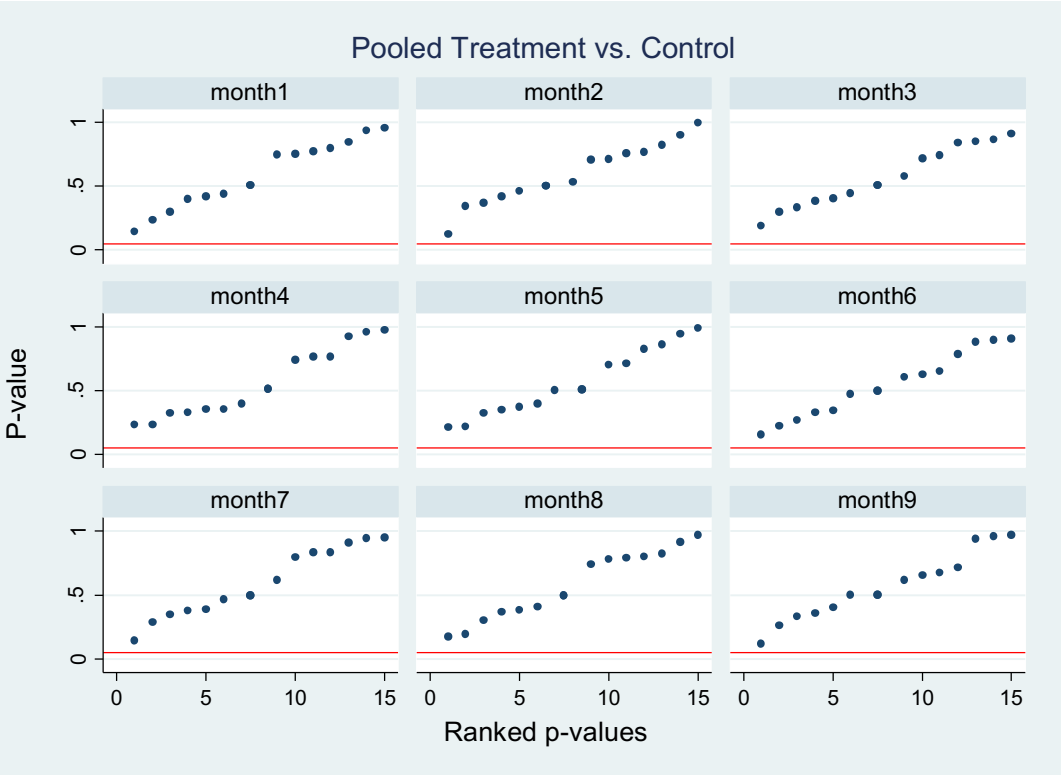
NOTE: This figure shows the time and date of each pill-taking event based on this participants' electronically monitored adherence.

Appendix Figure A2. EDM Adherence from September 2016 to April 1 2017, once-daily regimen



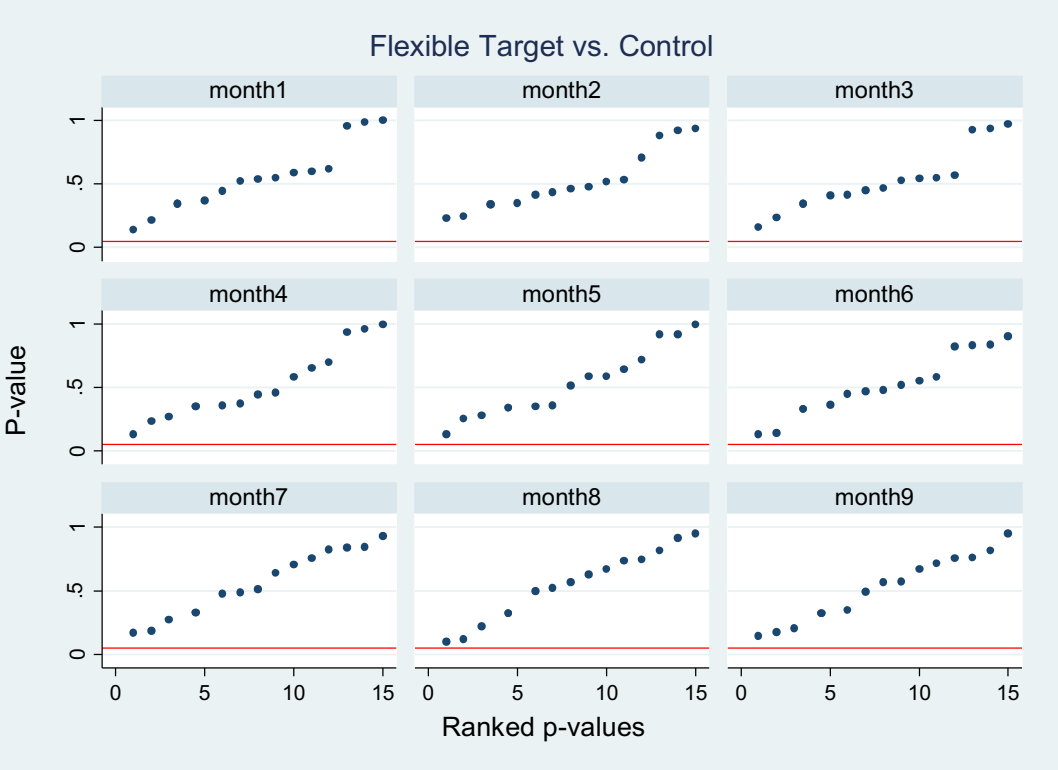
NOTE: This figure shows the time and date of each pill-taking event based on this participants' electronically monitored adherence.

Appendix Figure A4. Comparing Pooled Treatment vs. Control along Baseline Variables in Each Study Month



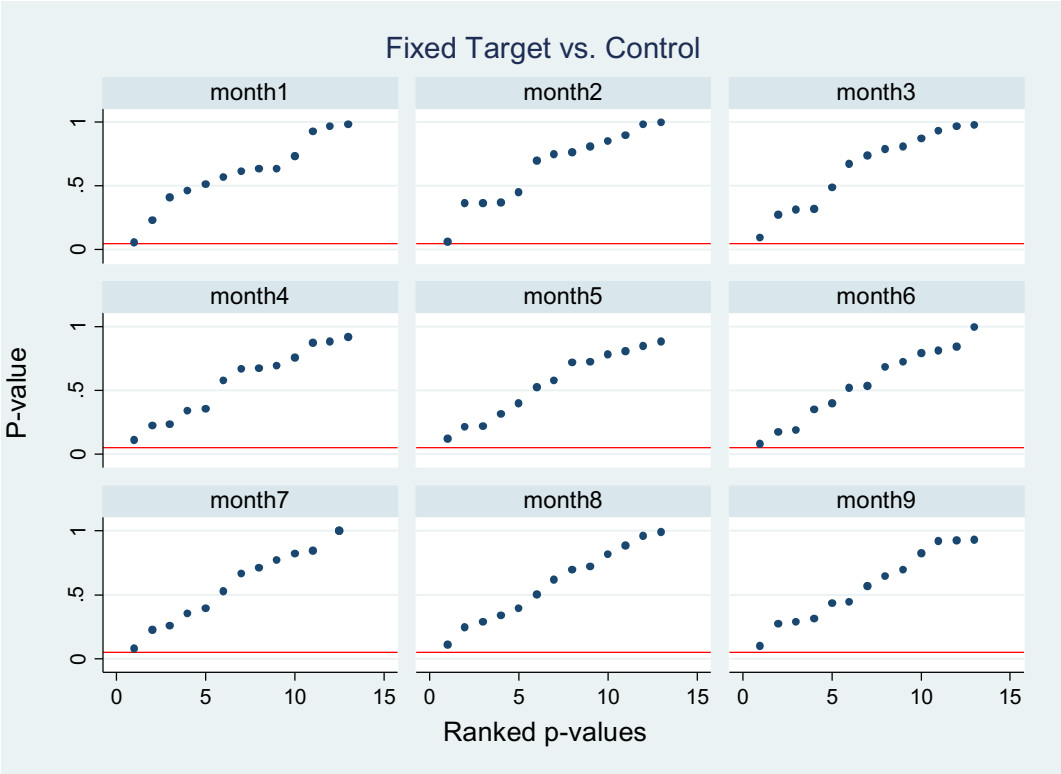
NOTE: To check that missing observations in each month does generate an unbalanced sample, ranked p-values are shown from a means test for all baseline balancing variables comparing Pooled treatment versus Control for the non-missing observations in each month included in the analysis. Red line marks 5 percent significance.

Appendix Figure A5. Comparing Flexible Target vs. Control along Baseline Variables in Each Study Month



NOTE: To check that missing observations in each month does generate an unbalanced sample, ranked p-values are shown from a means test for all baseline balancing variables comparing Flexible Target treatment versus Control for the non-missing observations in each month included in the analysis. Red line marks 5 percent significance.

Appendix Figure A6. Comparing Fixed Target vs. Control along Baseline Variables in Each Study Month



NOTE: To check that missing observations in each month does generate an unbalanced sample, ranked p-values are shown from a means test for all baseline balancing variables comparing Fixed Target treatment versus Control for the non-missing observations in each month included in the analysis. Red line marks 5 percent significance.

References

Ai, C. and Norton, E.C., 2003. Interaction terms in logit and probit models. *Economics letters*, *80*(1), pp.123-129.

Anoje, C., Agu, K.A., Oladele, E.A., Badru, T., Adedokun, O., Oqua, D., Khamofu, H., Adebayo, O., Torpey, K. and Chabikuli, O.N., 2017. Adherence to on-time ART drug pick-up and its association with CD4 changes and clinical outcomes amongst HIV infected adults on first-line antiretroviral therapy in Nigerian hospitals. *AIDS and Behavior*, *21*(2), pp.386-392.

Baird, S., McIntosh, C. and Özler, B., 2009. Designing cost-effective cash transfer programs to boost schooling in sub-Saharan Africa. *World Bank Policy Research Working Paper*, *5090*.

Banerjee, A.V., Duflo, E., Glennerster, R. and Kothari, D., 2010. Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. BMJ, 340, p.c2220.

Baranov, V., Bennett, D. and Kohler, H.P., 2015. The indirect impact of antiretroviral therapy: mortality risk, mental health, and HIV-negative labor supply. *Journal of health economics, 44*, pp.195-211.

Bastard, M., Pinoges, L., Balkan, S., Szumilin, E., Ferreyra, C. and Pujades-Rodriguez, M., 2012. Timeliness of clinic attendance is a good predictor of virological response and resistance to antiretroviral drugs in HIV-infected patients. *PLoS One*, 7(11), p.e49091.

Bettinger, E.P., 2012. "Paying to learn: The effect of financial incentives on elementary school test scores." *Review of Economics and Statistics*, *94*(3), pp.686-698.

Bezabhe, W.M., Chalmers, L., Bereznicki, L.R. and Peterson, G.M., 2016. Adherence to antiretroviral therapy and virologic failure: a meta-analysis. *Medicine*, *95*(15).

Bonezzi, A., Brendl, C.M. and De Angelis, M., 2011. Stuck in the middle: The psychophysics of goal pursuit. *Psychological science, 22(5),* pp.607-612.

Bor, J., Tanser, F., Newell, M.L. and Bärnighausen, T., 2012. In a study of a population cohort in South Africa, HIV patients on antiretrovirals had nearly full recovery of employment. *Health affairs, 31(7),* pp.1459-1469.

Bruhn, M. and McKenzie, D., 2009. In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, *1*(4), pp.200-232.

Brunstein, J. C. 1993. Personal goals and subjective well-being: A longitudinal study. *Journal of personality and social psychology 65(5)*: 1061.

Byakika-Tusiime, J., Oyugi, J.H., Tumwikirize, W.A., Katabira, E.T., Mugyenyi, P.N. and Bangsberg, D.R., 2005. Adherence to HIV antiretroviral therapy in HIV+ Ugandan patients purchasing therapy. *International journal of STD & AIDS*, *16*(1), pp.38-41.

Carver, C. S. and M. Scheier., 1990. Principles of self-regulation: Action and emotion, Guilford Press.

Claborn, K.R., Meier, E., Miller, M.B. and Leffingwell, T.R., 2015. A systematic review of treatment fatigue among HIV-infected patients prescribed antiretroviral therapy. *Psychology, health & medicine, 20(3)*, pp.255-265.

Datta, S. and S. Mullainathan., 2014. Behavioral design: a new approach to development policy. *Review of Income and Wealth 60(1):* 7-35.

Dupas, P., 2011. Health behavior in developing countries. *Annu. Rev. Econ., 3(1),* pp.425-449.

Egger, M., May, M., Chêne, G., Phillips, A.N., Ledergerber, B., Dabis, F., Costagliola, D., Monforte, A.D.A., De Wolf, F., Reiss, P. and Lundgren, J.D., 2002. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *The Lancet*, *360*(9327), pp.119-129.

Erez, M., Earley, P.C. and Hulin, C.L., 1985. The impact of participation on goal acceptance and performance: A two-step model. *Academy of Management journal, 28(1)*, pp.50-66.

Etkin, J. and R. K. Ratner (2011). "The dynamic impact of variety among means on motivation." Journal of Consumer Research 38(6): 1076-1092.

Evans, D., Kremer, M. and Ngatia, M., 2008. The impact of distributing school uniforms on children's education in Kenya.

Gill, C.J., Hamer, D.H., Simon, J.L., Thea, D.M. and Sabin, L.L., 2005. No room for complacency about adherence to antiretroviral therapy in sub-Saharan Africa. *Aids, 19(12),* pp.1243-1249.

Giné, X., Karlan, D. and Zinman, J., 2010. Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics, 2(4),* pp.213-35.

Gneezy, U., Meier, S. and Rey-Biel, P., 2011. When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives, 25(4),* pp.191-210.

Gross, R., et al. (2001). "Effect of adherence to newly initiated antiretroviral therapy on plasma viral load." Aids 15(16): 2109-2117.

Heath, C., Larrick, R.P. and Wu, G., 1999. Goals as reference points. *Cognitive psychology, 38(1),* pp.79-109.

Hogg, R.S., Heath, K., Bangsberg, D., Yip, B., Press, N., O'Shaughnessy, M.V. and Montaner, J.S., 2002. Intermittent use of triple-combination therapy is predictive of mortality at baseline and after 1 year of follow-up. *Aids, 16(7)*, pp.1051-1058.

Huang, S.C., Zhang, Y. and Broniarczyk, S.M., 2012. So near and yet so far: The mental representation of goal progress. *Journal of Personality and Social Psychology, 103(2)*, p.225.

John, L.K., Loewenstein, G., Troxel, A.B., Norton, L., Fassbender, J.E. and Volpp, K.G., 2011. Financial incentives for extended weight loss: a randomized, controlled trial. *Journal of general internal medicine*, *26*(6), pp.621-626.

Just, D. R. and J. Price (2013). "Using incentives to encourage healthy eating in children." Journal of Human Resources 48(4): 855-872.

Kahana, S.Y., Rohan, J., Allison, S., Frazier, T.W. and Drotar, D., 2013. A meta-analysis of adherence to antiretroviral therapy and virologic responses in HIV-infected children, adolescents, and young adults. *AIDS and Behavior*, *17*(1), pp.41-60.

Kahneman, D. and A. Tversky., 1979. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society:* 263-291.

Kamenica, Emir., 2012. Behavioral economics and psychology of incentives. *Annu. Rev. Econ.* 4, no. 1: 427-452.

Kessler, J. B. and C. Y. Zhang., 2014. Behavioral Economics and Health. *Paper for Oxford Textbook of Public Health.* Available at: http://assets. wharton. upenn. edu/~ czhan/KesslerZhang_BehavioralEconom icsHealth. pdf.

Kimeu, M., Burmen, B., Audi, B., Adega, A., Owuor, K., Arodi, S., Bii, D. and Zielinski-Gutiérrez, E., 2016. The relationship between adherence to clinic appointments and year-one mortality for newly enrolled HIV infected patients at a regional referral hospital in Western Kenya, January 2011–December 2012. *AIDS care*, *28*(4), pp.409-415.

Kimmel, S.E., Troxel, A.B., Loewenstein, G., Brensinger, C.M., Jaskowiak, J., Doshi, J.A., Laskin, M. and Volpp, K., 2012. Randomized trial of lottery-based incentives to improve warfarin adherence. *American heart journal, 164(2),* pp.268-274.

Kivetz, R., Urminsky, O. and Zheng, Y., 2006. The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research, 43(1)*, pp.39-58.

Latham, G.P., Winters, D.C. and Locke, E.A., 1994. Cognitive and motivational effects of participation: A mediator study. *Journal of Organizational Behavior, 15(1),* pp.49-63.

Latham, L. a., 1990. A Theory of Goal Setting and Task Performance. Englewood Cliffs.

Laurent, C., Gueye, N.F.N., Ndour, C.T., Gueye, P.M., Diouf, M., Diakhaté, N., Kane, N.C.T., Laniece, I., Ndir, A., Vergne, L. and Ndoye, I., 2005. Long-term benefits of highly active antiretroviral therapy in Senegalese HIV-1-infected adults. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *38*(1), pp.14-17.

Lawless, L., Drichoutis, A.C. and Nayga, R.M., 2013. Time preferences and health behaviour: a review. *Agricultural and Food Economics*, *1*(1), p.17.

Leuven, E., Oosterbeek, H. and Van der Klaauw, B., 2010. The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, *8*(6), pp.1243-1265.

Levitt, S.D., List, J.A., Neckermann, S. and Sadoff, S., 2016. The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, *8*(4), pp.183-219.

List, J. A., & Samek, A. S., 2015. The behavioralist as nutritionist: Leveraging behavioral economics to improve child food choice and consumption. *Journal of health economics*, *39*, 135-146.

Linnemayr, S., Huang, H., Luoto, J., Kambugu, A., Thirumurthy, H., Haberer, J.E., Wagner, G. and Mukasa, B., 2017. Text messaging for improving antiretroviral therapy adherence: no effects after 1 year in a randomized controlled trial among adolescents and young adults. *American journal of public health, 107(12),* pp.1944-1950.

Linnemayr, S. and C. Stecher., 2015. Behavioral economics matters for HIV research: the impact of behavioral biases on adherence to antiretrovirals (ARVs). *AIDS and Behavior 19(11):* 2069-2075.

Linnemayr, S., Stecher, C. and Mukasa, B., 2017. Behavioral economic incentives to improve adherence to antiretroviral medication. *AIDS (London, England), 31(5)*, p.719.

Loewenstein, G., Asch, D.A. and Volpp, K.G., 2013. Behavioral economics holds potential to deliver better results for patients, insurers, and employers. *Health Affairs*, *32*(7), pp.1244-1250.

Louro, M.J., Pieters, R. and Zeelenberg, M., 2007. Dynamics of multiple-goal pursuit. *Journal of personality and social psychology*, *93*(2), p.174.

MacCarthy, S., Mendoza-Graf, A., Huang, H., Mukasa, B. and Linnemayr, S., 2019. Supporting Adolescents to Adhere (SATA): Lessons learned from an intervention to achieve medication adherence targets among youth living with HIV in Uganda. *Children and Youth Services Review*, *102*, pp.56-62.

Mani, A., Mullainathan, S., Shafir, E. and Zhao, J., 2013. Poverty impedes cognitive function. *science*, *341*(6149), pp.976-980.

Mills, E.J., Nachega, J.B., Buchan, I., Orbinski, J., Attaran, A., Singh, S., Rachlis, B., Wu, P., Cooper, C., Thabane, L. and Wilson, K., 2006. Adherence to antiretroviral therapy in sub-Saharan Africa and North America: a meta-analysis. *Jama*, *296*(6), pp.679-690.

Nabukeera-Barungi, N., Elyanu, P., Asire, B., Katureebe, C., Lukabwe, I., Namusoke, E., Musinguzi, J., Atuyambe, L. and Tumwesigye, N., 2015. Adherence to antiretroviral therapy and retention in care for adolescents living with HIV from 10 districts in Uganda. *BMC infectious diseases, 15(1)*, p.520.

Nachega, J.B., Hislop, M., Nguyen, H., Dowdy, D.W., Chaisson, R.E., Regensberg, L., Cotton, M. and Maartens, G., 2009. Antiretroviral therapy adherence, virologic and immunologic outcomes in adolescents compared with adults in southern Africa. *Journal of acquired immune deficiency syndromes (1999), 51(1),* p.65.

Nunes, J. C. and X. Drèze., 2006. The endowed progress effect: How artificial advancement increases effort. *Journal of Consumer Research 32(4):* 504-512.

Ortego, C., Huedo-Medina, T.B., Llorca, J., Sevilla, L., Santos, P., Rodríguez, E., Warren, M.R. and Vejo, J., 2011. Adherence to highly active antiretroviral therapy (HAART): a meta-analysis. *AIDS and Behavior*, *15*(7), pp.1381-1396.

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. and Wesby, L., 2016. Framing financial incentives to increase physical activity among overweight and obese adults: a randomized, controlled trial. *Annals of internal medicine, 164(6)*, pp.385-394.

Pope, D.G. and Schweitzer, M.E., 2011. Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, *101*(1), pp.129-57.

Reisner, M.S.L., Mimiaga, M.J., Skeer, M.M., Perkovich, M.B., Johnson, M.C.V. and Safren, S.A., 2009. A review of HIV antiretroviral adherence and intervention studies among HIV–infected youth. *Topics in HIV medicine: a publication of the International AIDS Society, USA, 17(1)*, p.14.

Soman, D. and M. Shi (2003). Virtual progress: The effect of path characteristics on perceptions of progress and choice. *Management Science 49(9):* 1229-1250.

Sue-Chan, C. and M. Ong., 2002. Goal assignment and performance: Assessing the mediating roles of goal commitment and self-efficacy and the moderating role of power distance. *Organizational Behavior and Human Decision Processes 89(2):* 1140-1161.

Thirumurthy, H., Masters, S.H., Rao, S., Bronson, M.A., Lanham, M., Omanga, E., Evens, E. and Agot, K., 2014. Effect of providing conditional economic compensation on uptake of voluntary medical male circumcision in Kenya: a randomized clinical trial. *Jama*, *312*(7), pp.703-711.

Thornton, R.L., 2008. The demand for, and impact of, learning HIV status. *American Economic Review, 98(5)*, pp.1829-63.

Volpp, K.G., John, L.K., Troxel, A.B., Norton, L., Fassbender, J. and Loewenstein, G., 2008. Financial incentive–based approaches for weight loss: a randomized trial. *Jama, 300(22),* pp.2631-2637.

World Health Organization. 2003. Adherence to long-term therapies: evidence for action.

World Health Organization. 2006. From access to adherence: the challenges of antiretroviral treatment (studies from Botswana, Tanzania and Uganda).

Wu, G., Heath, C. and Larrick, R., 2008. A prospect theory model of goal behavior. *Unpublished manuscript*.