A Spoonful of Sugar: The Effect of Small Lottery Incentives on Medication Adherence among HIV-positive Youth in Uganda

Haijing Huang

Sebastian Linnemayr

This draft: October 2017

Abstract

Low adoption of healthy behaviors and products is a critical issue in developing countries. In the context of HIV, achieving long-term health and thereby reducing the likelihood of transmission to others requires a life-time commitment to medication adherence and usage of health services. Despite scale-up of access to antiretroviral therapy, low adherence and treatment fatigue threaten disease containment. We experimentally investigate the impact of small lottery incentives in the form of mobile airtime on medication adherence among HIV-positive youth in Uganda. An unexplored issue in the design of incentives is how conditionality thresholds should be set—whether they should be relatively high and target clinically meaningful thresholds, or take into consideration individuals’ current behavior to avoid potential demotivation, particularly for those with low adherence observed at baseline. Drawing on insights from psychology and Prospect Theory as applied to goal-setting, we assess the relative effectiveness of conditioning incentives on a flexible target (the participants choose their adherence target at 80 percent or higher, which can subsequently be adjusted) versus fixed target at the clinically recommended rate of 90 percent. We find that incentives overall have a sizeable, positive impact, and that this pooled effect is largely driven by the Flexible Target group. Furthermore, the difference in treatment impact is due to heterogeneous effects: only those with low observed baseline adherence in the Flexible Target group have a large and positive treatment effect relative to control (17 percentage point increase) compared to no statistically significant effect in the Fixed Target group. These findings provide preliminary evidence to suggest that incorporating a “goal gradient” may be useful in the design of incentive interventions, particularly for those with low levels of the incentivized behavior.

1 Corresponding author, Pardee RAND Graduate School, RAND Corporation, email: chuang1@rand.org. This research was funded by the Anne and James Rothenberg Dissertation Award 2015 & 2016, the Pardee Global Human Progress Award 2017, and the National Institutes of Child Health and Development administrative supplement grant no. R01HD074925-04S1 to Sebastian Linnemayr’s R01 no. R01HD074925.

2 RAND Corporation
1. Introduction

The global burden of disease is concentrated in poor countries. Malaria, waterborne infections from unclean drinking water, tuberculosis (TB), and HIV/AIDS - to name but a few - cause more than 15 million deaths each year (WHO, 2012). For many of these ailments effective treatment or prevention products exist and are often widely available at low cost, freely provided or subsidized through public sector programs, yet usage of health services and maintenance of required health behaviors is often problematic (Dupas 2011). Even when treatment is freely available in many low-income countries as for TB and HIV low uptake, improper usage of and adherence to drugs threaten disease elimination efforts. Given that many health behaviors, such as adherence for chronic illnesses must be maintained over a long period, how can policies be designed to be both sustainable and effective at motivating repetitive behaviors?

We explore these questions in the context of HIV/AIDS, where health decisions made by those at risk or living with HIV have a particularly large effect on disease containment, especially in Sub-Saharan Africa (SSA), where 25.6 million people are living with HIV in the region as of 2013, accounting for 71 percent of the global total (WHO 2016). Since HIV is both a chronic and infectious disease, achieving long-term health and reduced likelihood of transmission to others requires a lifelong commitment to medication adherence and usage of health services. Since the 2000s, increases in international funding for antiretroviral therapy (ART) and generic competition led to the mass disbursement of ART in many SSA countries through public sector treatment programs. ART does not permanently eradicate the HIV virus but substantially improves health and survival among those taking it and reduces the chance of transmitting HIV to others by interrupting viral replications, lowering viral load in the bloodstream and enabling immune system recovery. Those who remain in treatment and adhere 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).
Despite this progress in treatment access, low ART adherence remains a major economic and public health concern. Taking fewer than 90 percent of prescribed medications is linked to incomplete suppression of viral replication (Gross et al. 2001), progression to AIDS or death (Hogg et al. 2002), as well as the development and spread of antiretroviral drug-resistant HIV (Hecht et al. 1998, Pillay 2000). 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). As ART is a life-long prescription, many patients 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.” (Claborn et al. 2015) This means that ensuring one-time treatment take-up is insufficient to reap the full benefits of ART; adherence and health behaviors over the long term are equally important policy goals.

There is a growing body of research showing that decision-making errors are an important factor in health decision-making. Dupas (2011) reviews the literature on barriers to health behaviors in low-income countries and finds there are a substantial number of deviations from the neoclassical model: lack of and limited ability to process information, and deviations from rationality where people exhibit time-inconsistent preferences and myopia when it comes to their health (Dupas 2011). Systematic decision making errors or ‘behavioral biases’ such as present biasness – disproportionally favoring current consumption or pleasure over long-term benefits - are found to be negatively correlated with health decisions (Lawless et al. 2013, Kessler and Zhang 2014); for HIV, Linnemayr and Stecher (2015) find that present-bias, information salience, and over-optimism are linked to sub-optimal adherence among Ugandan adult HIV patients (Linnemayr and Stecher 2015). The field of behavioral economics also suggests that offers descriptive insights into the nature of poverty; Mani et al. (2013) find that poverty is as much about “psychological and cognitive scarcity as about financial and material deprivation.” (Mani et al. 2013, Datta and Mullainathan 2014) Those contending with time and monetary scarcity have fewer cognitive resources to devote to other cognitively-intensive tasks – such as keeping a strict pill-taking regimen.
Small material or financial incentives used to counter such biases can have a disproportionately large impact on behaviors and outcomes. Banerjee et. al. (2010), for example, found that providing a small bag of lentils to those presenting for malaria immunization was enough to neutralize present-bias by offering immediate gratification for adopting a behavior that had only long-run payoffs (Banerjee et al. 2010). Variable rewards (i.e. those allocated in a non-deterministic manner) were successful in improving outcomes in chronic disease management including warfarin medication adherence (Kimmel et al. 2012) and weight loss (Volpp et al. 2008). The small number of studies that have used small, ‘nudge’-incentives in the context of HIV have shown promising results, but are primarily focused on testing and prevention. Thornton (2008) 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). Nyqvist et. al. (2016) evaluated a lottery program with small payoffs in rural Lesotho among young adults (Bjorkman Nyqvist et al. 2016) and find that the pooled lottery intervention decreased HIV incidence rate over the two-year trial period, compared to the control group. Linnemayr and Stecher (2017) find in a pilot study that participants receiving small, in-kind incentives were 23.7 percentage points more likely to achieve 90% adherence compared to the control group (Linnemayr et al. 2017).

In this paper, we experimentally evaluate the impact of small incentives on medication adherence among adolescents and young adults (henceforth “AYA”) using a randomized experiment in Uganda. Our study makes several contributions to the existing literature:

First, we inform the design of incentives by evaluating the relative effectiveness of two approaches to setting incentive eligibility thresholds. In the “Fixed Target” treatment arm, participants must reach the fixed, clinically recommended rate of 90 percent mean adherence to be eligible to participate in a prize drawing. In the “Flexible Target” arm, participants can choose a target of their own liking - anywhere above 80 percent mean adherence, which can be adjusted after each drawing. The theoretical literature on goal-setting has shown that a target set too high can exert a demotivating influence on those far away from it (Wu et al. 2008); this is particularly relevant for our sample, with a baseline average adherence of only 65 percent and with less than 30 percent adhering at the 90 percent mark. Giving participants the option to set flexible targets allows them to set more personally manageable goals, which may be particularly
crucial among those who are initially far from the clinically optimal threshold. Heath et al. (1999) creates a useful and parsimonious conceptual framework to understand this relationship between demotivation and distance to a goal, 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 is the “goal gradient” effect: the closer an individual gets to a goal, the more motivated s/he would be to achieve it. Conversely, someone far away from a target goal is less likely to exert effort to move towards the goal than someone who is initially closer (Heath et al. 1999). We use this framework to better understand potential mechanisms behind the differences in treatment impact between Fixed and Flexible targets. Given that distance to a goal has a potentially large impact on motivation and subsequent performance, there has been surprisingly little done in the way of incorporating this insight into the design and analysis of incentive interventions. As far as we are aware, this is the first paper exploring the relative effectiveness of fixed, high targets versus variable targets, for health behaviors.

Second, our study contributes to the HIV literature by focusing on an understudied population with relatively large adherence barriers. Adolescents account for a large proportion of new HIV cases and they are the only age group where HIV-related mortality is going up (Nabukeera-Barungi et al. 2015). AYA also face greater barriers to optimal adherence and retention compared to adults; a longitudinal study from South Africa found that adolescents aged 11-19 were 70–75 percent less likely to be virologically suppressed (≤400 copies/mL) at 1 and 2 years after ART initiation (Nachega et al. 2009). Recent research in neuroscience suggest that AYA have greater self-regulatory challenges and discount the future at higher rates than adults (Steinberg 2008, Steinberg et al. 2009), yet few studies have applied interventions based on behavioral economic principles to this group. Our participants, who have been presenting at the clinic for an average of 6 years have been on ART for at least 2 years, are also representative of a group who are likeliest to experience treatment fatigue and for whom the need for effective intervention is greatest.

We recruited 216 AYAs from a publicly-funded HIV clinic in Uganda and present results from the first 6 months of the randomized trial. Study participants were randomized into either the Flexible Target arm, Fixed Target arm, or a comparison group. Eligibility to participate in the
prize drawing with expected value of $1.40 during routine clinic visits was conditional on achieving electronically-measured adherence above the target. If the target was met, study coordinators directly disbursed mobile airtime to the participant’s phone. Since this study uses the same sample and continues from a previous SMS messaging study, all groups continued to receive weekly SMS reminder messages; given that we found no impact of SMS messaging on adherence (Linnemayr et al 2017, forthcoming in American Journal of Public Health), the comparison SMS-only group can be considered our “de-facto” control group in this new study.

Several insights emerge from the experiment. First, pooling the two prize drawing incentive arms, the main experiment improved mean adherence by 7.2 percentage points compared to the control group at borderline statistical significance (p-value=.08). However, when comparing each treatment arm separately to control, we find that this average incentive effect size from pooling the groups is driven by the Flexible target treatment arm. Compared with control, those in the Flexible arm improved adherence by 12.0 percentage points (significant at 1 percent) compared to a small and not statistically significant difference comparing Fixed target to Control. Flexible target treatment impact remains even after discarding bottle openings a month leading up to each game – the period where adherence was explicitly checked against the target (and participants knew this). While we have no reason to believe participants are forcing bottle openings to “game” the system, this explicitly checks against this possibility in addition to suggesting that there is intrinsic motivation to reach a goal participants have set, even when participants know their good behavior will not affect their chances of getting an airtime reward.

Second, participants in the Flexible Target group chose adherence targets that suggest sophistication in goal-setting. In each game, over 70 percent of participants in this group chose a goal that is higher than the lowest possible selection (80), with the modal target increasing with baseline adherence level, suggesting that participants viewed the prize drawing as an opportunity to hold themselves to a higher adherence standard rather than maximize expected prize monies. This suggests that participants perceive an intrinsic motivation to improve their adherence. John et. al. (2011) tested financial incentives for weight loss and also found that people chose goals that were more stringent than the minimum that would be economically rational. (John et al. 2011)
Third, in an analysis of treatment heterogeneity by baseline levels of adherence, we find that, consistent with our predictions, 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 baseline adherence below 60 percent). Among this group, Flexible incentives improve adherence by 17 percentage points relative to Control (p-value <.05), compared to a 0.2-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 be demotivating to those far away from them, and that allowing for lower, personal standards may be more effective at improving performance among this group. 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 have a statistically significant average treatment impact, for medium baseline adherers (defined as between 60 and 90 percent at baseline) adherence improved by 19 percentage points. Due to small sample sizes for these sub-analyses these additional analyses should be interpreted with some caution; however, as they test pre-defined hypotheses based on a clearly defined theoretical framework, the results provide some useful insights for future research in testing a more granular implementation of sub-goaling versus fixed goals.

2. Context and Conceptual Framework

2.1 ART adherence

Adherence or non-adherence is a series of daily behaviors by the patient that is generally unobservable to the clinician; as such, we briefly discuss its definition and measurement in the context of SSA. Medication adherence in our case is the use of 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 exorbitant costs of second and third line compared to first-line drugs. Clinically optimal adherence is generally considered to be around 90 to 95 percent (Gill et al. 2005).
Few studies can observe patient pill-taking behaviors over an extended period; hence surrogate measures of adherence are used. 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 a variety of the first two measures, comparing them against the ultimate outcome of interest, 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).

A 2006 WHO report on challenges of ART reviewed several studies in African settings from pre-2005 and found that optimal adherence rates ranged from 54 percent to 98 percent: Botswana (Weiser et al., 2003: 54 percent); Nigeria (Daniels, 2004: 79 percent); South Africa (Ferris et al., 2004: 77 percent); Uganda (Byakika-Tusime, 2005: 67 percent; Munganzi, 2004: 98 percent); and Rwanda (Omes, 2004: 85 percent) Most of these, with the exception of Omes, used self-reported measures(WHO 2006).

These studies may have overestimated adherence due to relying on self-reported data or excluding patients lost to follow-up from their calculations. They are also not likely to be an accurate portrayal of ART adherence today; they were conducted while ART was newly introduced to the region, and there is evidence from longitudinal studies of declining adherence over time. Senegal, Laurent et. al. noted that over 95 percent of patients had adherence exceeding 80 percent after 1 month on therapy, but 18 months later only 80 percent of patients remained above that level. The proportion of patients with undetectable viral loads also fell, correspondingly, from 79.6 to 59.3 percent. A more recent 2012 paper from Kenya also found a 12-percentage point discrepancy between mean self-reported adherence of (98.7 percent) and electronically monitored adherence (86 percent), which grew over time as EDM adherence declined. Results from another 2011 study in Kenya, found that in the context of a randomized control trial, those in the control group saw declines in EDM-measured adherence from 60 percent achieving at least 90 percent adherence, to 40 percent within about a year (48 weeks).

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 forthcoming 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).

2.2 Conceptual Framework

In our study design, we were motivated by the consideration that in our sample average adherence is 65 percent, with 30 percent adhering at over 90 percent and more than half below 70 percent. If our sample is representative of the clinic at large, then a substantial number of young patients are struggling with adherence and far from the clinical benchmark. What is the best way to motivate their adherence behavior?

There is a vast literature in psychology on goal-setting and motivation, with several relevant findings. The first is that goals function as reference points and 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 difficult 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 (2012) 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 from sub-goaling increases motivation and performance because they provide immediate and achievable benchmarks, whereas a distal goal is ineffective in mobilizing or directing effort(Bandura and Schunk 1981). Subgoaling is most useful when people are most doubtful about reaching a goal or farther from their goal (Latham 1990, Brunstein 1993, Soman and Shi 2003).
Heath et al. 1999 develop a parsimonious framework using Prospect Theory to summarize this literature. Heath et al. 1999 and Wu et al. 2008 show that the properties of the value function can parsimoniously explain previous empirical results in the goal literature on effort, persistence and performance.

The first insight is that goals operate as reference points in that people tend to categorize achievement outcomes as a “success” or a “failure” 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 outcome is below a reference point and positive if one is in the region of “gains” or “successes”, where outcome meets or exceeds the reference point. This means that one’s utility or happiness does not depend on absolute but relative performance. Consider two persons 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, she will feel worse about her performance.

A second property of the value function is that the steeper slope in the ‘loss’ region compared to the gain region of the value function illustrates “loss aversion” – 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 adherence will experience a larger utility increases [$v(75) - v(70)$] compared to a person moving from 85 to 90 adherence [$v(90) - v(85)$].

Empirical studies of risky choice and riskless choice have presented converging evidence that losses are weighted approximately two times more than equivalent gains (common values for the “coefficient of loss aversion” fall between 2 and 4). 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 even the behavior of highly experienced professionals in high stakes settings can 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 \((v''(x) < 0 \text{ when } x > 0 \text{ and } v''(x) > 0 \text{ when } x < 0)\). People who are very far from the 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 50) and the other is closer (adherence 80). The value from moving from 50-55 is much smaller than the value of moving from 70 to 75. 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.
A solution to overcoming the starting problem is the creation of sub-goals. Separating a large task into sub-goals means taking advantage of the properties of the value function and maximizing sensitivity to small movements toward the goal. The ideal goal gradient intervention would exogenously set goals for each participant, considering previous performance. Though we were logistically unable to carry this out, we can still use these two features of the value function to make some useful predictions about the two incentives treatment groups.

Classifying participants by their initial “starting point” adherence at baseline into “low”, “medium”, and “high” adherence (“Low” – below 60, “Medium” – >60, <90, “High” – 90 and up), we can make the following predictions:

- **Low-adherers**: When two individuals are below a target, the individual closer to it will exert more effort.” Hence, “low” adherence individuals will do better in the Flexible Target group compared to “low” adherence individuals in the Fixed target group since they can select a goal closer to their baseline level.
- **Medium-adherers**: “Medium” adherence individuals will benefit most from the “goal gradient”. The Flexible Target group is comprised of those who do set a challenging but attainable target for themselves and those who don’t, compared with everyone who is subject to the “high” goal of 90 in the Fixed group; as such, Fixed Target medium adherers are more likely to do better.

- **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 (as empirically seen in Table 3), we hypothesize that high adherers in the Flexible Target group will 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, we expect the difference between groups likely to be small.

3. **Experiment Design**

3.1 **Intervention and timeline**

The experiment design and study population closely relates to the RCT that preceded it. In this section we discuss the relevant details from this earlier study (Study 1) before describing the main experiment which is the focus of this paper (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 Messing Service (SMS) intervention, an RCT 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 a generally poor clientele. Overall the SMS interventions did not have a statistically significant effect on adherence.
outcomes. We document findings in our one-year impact paper (Linnemayr et al. forthcoming in the American Journal of Public Health).

**Study 2: Incentives intervention**

As Study 1 drew to a close in July 2016, we randomized the remaining 229 participants presenting at Mildmay to a control group or one of two treatment groups in the study discussed here. The main experiment evaluated in this paper is the provision of prize drawing incentives, or “games”, as study participants and coordinators called it during the study and which we refer to in this paper. Participants randomized into two intervention arms were eligible to win 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. In the “Fixed Target” group, participants had to reach the externally imposed and clinically optimal adherence of 90 percent to be eligible to play a game. In the “Flexible Target” group, participants must reach a target of their own choosing, which must be any point over 80 percent in five-point intervals: 80, 85, 90, 95, or 100, and can be re-adjusted at each visit. If the participant met their target, the expected winning per game was 5,000 USH, amounting to 20,000 USH or $5.58 over four games which spanned roughly 9 months assuming the participant has visits about every two months. In this paper, we present results for the first six months of the intervention for which currently full data are available; results will be updated as additional information becomes available.

The SMS intervention continued to be implemented in both comparison and treatment groups. All three groups (including control) continued to receive a weekly message plus airtime of 1,000 USH or $0.27 conditional on responding to the message. Doing so guaranteed that those in the comparison group who received standard care in the preceding 24 months during Study 1 would also receive an intervention in Study 2 for reasons of fairness. Given that we found no effect of messaging on adherence in both years of the SMS study, we can consider this the de-facto “control” group. SMS messages to all groups were sent on Sundays at 9 am. Those who did not reply to the Sunday message received a follow-up message on Tuesday evening, which

---

3 We could not include the second clinic IDI in the new study because of delays due to IRB approval at that institution.

4 For ethical reasons we did not want to incentivize people to perform to lower than a medically beneficial target, and hence the target selected must be above 80.
was the same across comparison and intervention groups: “Hello, how are you? We have not heard back from you. Please reply in the next 24 hours to get 1000 Ush: 1 if well, 2 if unwell.”

The three groups are summarized in Figure 3.

![Figure 3. Experimental Set-up](image)

<table>
<thead>
<tr>
<th>Fixed Target</th>
<th>Flexible Target</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant draws from 3 cards with expected value $1.4 in mobile airtime if fixed adherence target of 90% is met</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.</td>
<td>No incentives</td>
</tr>
<tr>
<td>Can play up to 4 game during clinic visits over 9 months</td>
<td>Can play up to 4 game during clinic visits over 9 months</td>
<td></td>
</tr>
<tr>
<td>Weekly SMS motivational message + airtime top-up of 1,000 USH (27 cents) if participant responds</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 3.2 Randomization

Given our relatively small starting sample of 229, we did a block randomization based on four strata of baseline adherence. According to Bruhn and McKenzie (2008) in smaller samples, pairwise matching and stratification do better than a simple random draw. By randomizing within strata, block randomization theoretically eliminates differences between groups and increases sensitivity to detecting smaller treatment differences than would otherwise be possible (Box et al, 2005). 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 data collection and rolling recruitment. Study participants were recruited for Study 1 on a 6-month rolling basis so at any given calendar time there is variation in how many months each person has been in the study. Electronic adherence data was also extracted about once every 1-3 months on average 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 229 initial sample was around 6 months before the Study 2 start date. We used the three-month average adherence in this period. 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.

3.3 Recruitment

Participants were approached for rolling recruitment from August 2016 to January 2017. Patients presenting at the clinic were eligible for study participation if they 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 access for at least five days per week. Mobile phones or phone airtime were not provided; at baseline about 80 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 have used electronically monitored bottle caps since the start of parent study, 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 each opening was recorded as an “event”.

Participants were not informed exactly how the caps worked; for instance, they are unaware that each opening counted as an event, and hence were less likely to force false openings.

During recruitment, all treatment participants played a “free trial” game which 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. The procedure also aimed to increase the salience of regret aversion: we hypothesized that patients who had already participated in a game drawing once would strive harder to avoid being told at their next visit that they could not participate at that visit.

3.4 Clinic and game procedures

Participants came to the clinic for their regular clinic visits every 1 or 2 months. During each visit, study coordinators updated the phone number and 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 electronic bottle caps are downloaded into a computer during routine clinic visits. This is a simple and quick procedure: the bottle is placed top down on a flat reader. The program has a front-end which shows the exact date and times of opening in either a calendar format, or chart format. Appendix Figure 1 shows the chart of opening times for a patient adhering with a twice a day regimen. To standardize the period of adherence counted towards prize drawing eligibility and to build in a robustness check, participants were told their adherence would be checked for the last 30 days before their visit.

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

1. Study coordinator placed the EDM cap on the reader and checked whether adherence target is met over the last 30 days from the date of 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 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 is in the Flexible Target group, study coordinators would additionally ask: “Your target was 85 for this visit. Would you like to set a new target or keep the same target for next time?” The new target would 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 were told “Congratulations, your adherence this past month exceeded your target of 90 percent. You are eligible to play the game!” Participants drew once from three face-down cards with the amounts 0, 5000 USH, and 10,000 USH.
4. If they drew 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. If the participant did not have their own 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 insights

4.1 Randomization success and sample characteristics

Of the initial 229 starting sample at Mildmay clinic at the end of Study 1, 216 participants were enrolled in Study 2 (71 in Control, 72 in Fixed Target, and 73 in Flexible Target). Only 4 people who were approached declined to participate and the remaining 9 could not be reached for enrolment. Study 2 baseline variables are constructed using the Study 1 endline survey. The average age of study participants is 19 years, and roughly half (54 percent) are female. Almost all participants completed primary education and the majority could read and write a simple sentence easily.

Table 1 shows the balancing table of covariates across study groups. We compare socio-economic and demographic characteristics as well as several measures of adherence, and report the p-value from a chi-square test of equivalence. Randomization appears to have been successful at balancing majority of the baseline variables across groups. None of the socio-demographic variables differ statistically significantly across groups.

In terms of adherence variables, mean adherence at the time of randomization is balanced across groups, as well as self-reported adherence. However, adherence one month before intervention is no longer balanced; Fixed Target was higher than both Control and Flexible Target (Control = 65, Fixed Target=76, Flexible Target= 63). Because we randomized 229 in the initial sample but recruited 216, this difference could be explained in part by the change in composition in baseline adherence from the initial randomization. Part of it could also be due to a local spike in that month from unobserved factors. Still, we account for this pre-treatment difference in the estimation of treatment impact by doing a difference-in-differences analysis, which differences out adherence as well as other unobserved time-invariant individual characteristics.
Table 1. Demographics and Baseline Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Control (n=66)</th>
<th>Fixed Target (n=72)</th>
<th>Flexible Target (n=69)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female (%)</td>
<td>50.7</td>
<td>53.5</td>
<td>58.0</td>
<td>0.57</td>
</tr>
<tr>
<td>Age (years)</td>
<td>19.0</td>
<td>19.5</td>
<td>19.2</td>
<td>0.39</td>
</tr>
<tr>
<td>Married (%)</td>
<td>25.4</td>
<td>29.6</td>
<td>40.6</td>
<td>0.40</td>
</tr>
<tr>
<td>Literacy (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>read easily</td>
<td>69.0</td>
<td>69.0</td>
<td>63.8</td>
<td>0.77</td>
</tr>
<tr>
<td>write easily</td>
<td>77.5</td>
<td>77.5</td>
<td>65.2</td>
<td>0.77</td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Completed primary</td>
<td>100.0</td>
<td>100.0</td>
<td>98.6</td>
<td>0.38</td>
</tr>
<tr>
<td>Completed secondary</td>
<td>77.5</td>
<td>77.5</td>
<td>65.2</td>
<td>0.23</td>
</tr>
<tr>
<td>Housing (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-rated house as &quot;poor&quot;</td>
<td>43.7</td>
<td>49.3</td>
<td>37.7</td>
<td>0.35</td>
</tr>
<tr>
<td>Has electricity</td>
<td>78.9</td>
<td>88.7</td>
<td>81.2</td>
<td>0.26</td>
</tr>
<tr>
<td>Has piped water</td>
<td>53.5</td>
<td>59.2</td>
<td>50.7</td>
<td>0.55</td>
</tr>
<tr>
<td>Weekly income in Ugandan Shillings (USH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>20478</td>
<td>19493</td>
<td>17406</td>
<td>0.82</td>
</tr>
<tr>
<td>income&gt;50,000 USH (%)</td>
<td>41.4</td>
<td>45.1</td>
<td>42.0</td>
<td>0.85</td>
</tr>
<tr>
<td>income&gt;75,000 USH (%)</td>
<td>20.0</td>
<td>23.9</td>
<td>21.7</td>
<td>0.77</td>
</tr>
<tr>
<td>Adherence (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>self-reported</td>
<td>83.2</td>
<td>80.3</td>
<td>85.3</td>
<td>0.13</td>
</tr>
<tr>
<td>at randomization</td>
<td>72.1</td>
<td>76.0</td>
<td>68.2</td>
<td>0.15</td>
</tr>
<tr>
<td>1-month pre-intervention</td>
<td>65.0</td>
<td>76.0</td>
<td>62.5</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Notes: Self-reported adherence is the share of doses taken as prescribed in the past month, as reported by the participant during the baseline survey. Adherence at randomization corresponds to the 12-week period three months before intervention started. The last column shows the p-value from the F-statistic of the mean comparison between groups.

4.2 Game descriptives

Among the two intervention groups, an average of 3.6 game visits (clinic visits corresponding to a potential prize drawing) occurred over 6 months. Across all game-visits, slightly less than half were eligible for a prize drawing (46 percent) and played a game. Of those who were ineligible, having a 30-day EDM adherence lower than the target for that visit is the main reason (50 percent), followed by missing the appointment date (16 percent).
So how did participants in the Flexible Target group set their adherence targets? Overall, the targets are relatively evenly distributed: in each game only 14-16 percent chose the lowest adherence target of 80 percent (Table 2). The modal choice was 90 percent with majority choosing 90 or above consistently.

Table 2. Flexible Targets Chosen in Games 1-3

<table>
<thead>
<tr>
<th>target</th>
<th>Game 1</th>
<th></th>
<th>Game 2</th>
<th></th>
<th>Game 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>80</td>
<td>12</td>
<td>16</td>
<td>10</td>
<td>14</td>
<td>11</td>
<td>16</td>
</tr>
<tr>
<td>85</td>
<td>14</td>
<td>19</td>
<td>14</td>
<td>19</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>90</td>
<td>17</td>
<td>23</td>
<td>18</td>
<td>25</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td>95</td>
<td>15</td>
<td>21</td>
<td>13</td>
<td>18</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>100</td>
<td>15</td>
<td>21</td>
<td>14</td>
<td>19</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>100</td>
<td>69</td>
<td>100</td>
<td>70</td>
<td>100</td>
</tr>
</tbody>
</table>

Next, we evaluate whether those with lower adherence rates at baseline would choose lower targets than those with relatively higher adherence. Figure 4. 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 sample in each category, the graph is suggestive that participants are aware of and account for their own adherence abilities at baseline and select a target that is commensurate with their adherence level. Among low adherers at baseline, the modal target is 80 percent; for medium adherers it is 95 percent. For high adherers, it is 100 percent.

For repeated games, an interesting question is how participants adjust their targets over time. Table 3 shows the number and percentage of instances where a participant increased their goal after having met it in the previous game, by the previous game goal. Interestingly, as performance increases, subsequent goal ambition decreases. If a low goal of 80 was met, 63 percent increased their subsequent goal. Once the fixed reference point of 90 is reached however, this goes down to 24 percent. For those not meeting their adherence goals at each visit, majority adjusted downwards (Table 4).
Figure 4. Adherence target among Flexible Target participants, by baseline adherence category, Game 1

![Bar charts showing adherence distribution among Low Adherers, Medium Adherers, and High Adherers.]

Table 3. Frequency and percentage of next goal increases, by previous goal (if previous goal was met)

<table>
<thead>
<tr>
<th>Previous goal</th>
<th>Increased goal in next game</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>80</td>
<td>12</td>
<td>63</td>
</tr>
<tr>
<td>85</td>
<td>12</td>
<td>52</td>
</tr>
<tr>
<td>90</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>95</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes: Total n refers to game instances where a goal was met
### Table 4. Frequency and percentage of next goal decreases, by previous goal (if previous goal was not met)

<table>
<thead>
<tr>
<th>Previous goal</th>
<th>Decreased goal in next game</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>85</td>
<td>32</td>
<td>84</td>
</tr>
<tr>
<td>90</td>
<td>37</td>
<td>74</td>
</tr>
<tr>
<td>95</td>
<td>22</td>
<td>85</td>
</tr>
<tr>
<td>100</td>
<td>29</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: Total n refers to game instances where a goal was not met.

---

## 5. Empirical specification and main experimental results

### 5.1 Construction of adherence

Mean monthly adherence was calculated as the number of bottle openings divided by the number of prescribed pills for a 30 day period. 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 regimens over the course of the study, we used 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 was calculated using the date of change as the first day of the new regimen. Appendix Figures 1 and 2 show examples of EDM readings for a participant with a consistent twice-daily regimen (taking one 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). In addition to this primary outcome, we also evaluate whether the patient had monthly adherence of at least 90 percent.

### 5.2 Empirical specification

Because our adherence variable is measured daily, there are multiple levels of time we can use to construct a longitudinal adherence dataset over the course of 6 months (e.g. overall aggregate over 6 months, daily, weekly, monthly). Summarizing over the entire period of analysis would
risk losing some level of granularity, particularly since behaviour change interventions can sometimes produce short-term effects that may erode over time. On the other hand, a time series at the daily level would create 180 observations per person and is practically unwieldy with our relatively small sample size, resulting in a large correlation matrix to estimate. As a practical middle ground, we use monthly adherence. As mentioned in the previous section, to account for baseline adherence differences between Fixed and Control group, we conduct a standard difference-in-differences (DID) analysis with the following main specification. The availability of pre-intervention adherence data allows us to difference out time-invariant individual-specific effects.

\[ Y_{it} = \alpha + \mu \text{Treat} + \tau \text{Post}_t + \beta (\text{Post}_t \times \text{Treat}) + \varepsilon_{it} \]  

(1)

Where \( i \) and \( t \) index individual and month, respectively. \( Y_{it} \) is the adherence outcome for individual \( i \) in month \( t \). \( \text{Treat}_i \) is the treatment assignment of each individual. This 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)\(^5\). \( \text{Post}_t \) is a dummy equals to one after the intervention started. \( \varepsilon_{it} \) is the idiosyncratic error term. All standard errors are clustered at individual level. The coefficient of interest is \( \beta \), the estimate of the effect of treatment on adherence. For the secondary outcome of adherence over 90 percent, 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:

\[ Y_{it} = \alpha + \theta (\text{Post}_t \times \text{Treat}) + \varphi_i + \pi_t + \varepsilon_{it} \]  

(2)

Where \( \varphi_i \) are individual effects, \( \pi_t \) are month fixed effects and \( \varepsilon_{it} \) is the remaining idiosyncratic error.

\(^5\) 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 \( Y_{it} = \alpha + \mu T1 + \gamma T2 + \tau \text{Post}_t + \theta (\text{Post}_t \times T1) + \beta (\text{Post}_t \times T2) + \varepsilon_{it} \).
5.3 Main results

First we estimate the effect of mobile airtime incentives on adherence. Figure 5 shows mean monthly adherence by pooled intervention arm and control. The pooled incentives intervention group has slightly higher adherence than control in the pre-period (months -1 and 0). In the first month of the intervention, there is a sharp increase in adherence that prevails at around 80 percent for the duration of the 6 months; the trendline for control is much less pronounced— a gradual increase at month 1, which then tapers off between 65 and 70 percent. The between-group difference in adherence since the intervention started appears to be larger than the difference before, suggesting the pooled intervention had a positive impact on adherence.

Figure 5. Mean Monthly Adherence over 6 Months – Pooled Treatment vs. Control

Estimation results are presented in Table 5. Overall the pooled incentives intervention has an effect of similar magnitude and significance across both specifications. In our main DID specification, the DID estimator for the pooled effect of incentives on mean adherence is 7.2 percentage points, though it is only significant at 10 percent. Adding individual and month fixed effects, it is 6.8 percentage points (p-value<.1).
This pooled intervention effect is mainly driven by the Flexible Target arm, as shown in Figure 6, which plots mean monthly adherence by group. In the period before the intervention, mean adherence among the Fixed Target group was higher than those in both the Flexible Target and Control group by over 10 percentage points. Over the course of the intervention, the between group difference between the Fixed Target and Control remained fairly constant, suggesting no effect from the Fixed incentive arm. However, the Flexible Target increased sharply from baseline to month 1 and remained around an average of 80 percent while Control hovered between 60 and 70 percent. The effects of the Flexible Target intervention also appear to be sustained throughout the 6-month period.

Estimations results in Table 5 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. Comparing Flexible Target versus Control, mean adherence increased by 12 percentage points on average over 6 months, with p-value<.05. Comparing Fixed Target versus control, mean adherence increases by just 1.2 percent and is not statistically significant.

Figure 7 shows our secondary outcome, proportion adhering over 90 percent over time, by intervention and control groups and Table 6 shows DID estimation results from linear probability regressions. The pooled incentives DID coefficient shows an increase of 7.2 and 6.8 percent in the proportion of those adhering over 90 percent, with p-value<0.1. For Fixed Target, proportion adhering over 90 percent is slightly higher than control, but not statistically significant. For Flexible Target, the proportion of those adhering over 90 percent are 11.7 - 12.0 percent higher than that of the Control group, with p-value<0.1.
Figure 6. Mean Monthly Adherence over 6 Months – Three-Group Comparison

Table 5. Intervention Treatment Effects on Mean Adherence, Group Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Pooled vs. Control</th>
<th>Fixed vs. Control</th>
<th>Flexible vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Treat x Post</td>
<td>0.072*</td>
<td>0.068*</td>
<td>0.029</td>
</tr>
<tr>
<td></td>
<td>(0.040)</td>
<td>(0.041)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Treat</td>
<td>0.055</td>
<td>0.118**</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.054)</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.042</td>
<td>0.042</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.034)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Individual + Time FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>1,569</td>
<td>1,569</td>
<td>1,045</td>
</tr>
<tr>
<td>Individuals</td>
<td>208</td>
<td>208</td>
<td>139</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.046</td>
<td>0.734</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Notes: Standard errors are clustered at individual level
“Treat” refers to either Pooled treatment, Fixed Target, or Flexible Target
*** p<0.01, ** p<0.05, * p<0.1
Figure 7. Adherence over 90% over 6 Months – Three-Group Comparison

Table 6. Intervention Treatment Effects on Adherence over 90%, Group Comparisons

<table>
<thead>
<tr>
<th></th>
<th>Pooled vs. Control</th>
<th>Fixed vs. Control</th>
<th>Flexible vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Treat x Post</td>
<td>0.069 (0.050)</td>
<td>0.065 (0.053)</td>
<td>0.023 (0.057)</td>
</tr>
<tr>
<td>Treat</td>
<td>0.042 (0.063)</td>
<td>0.086 (0.073)</td>
<td>-0.004</td>
</tr>
<tr>
<td>Post</td>
<td>0.056 (0.039)</td>
<td>0.056 (0.039)</td>
<td>0.056</td>
</tr>
<tr>
<td>Individual + Time FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>1,569</td>
<td>1,569</td>
<td>1,045</td>
</tr>
<tr>
<td>Individuals</td>
<td>208</td>
<td>208</td>
<td>139</td>
</tr>
</tbody>
</table>

Notes: Coefficients are from a linear probability regression. Standard errors are clustered at individual level.
“Treat” refers to either Pooled treatment, Fixed Target, or Flexible Target
*** p<0.01, ** p<0.05, * p<0.1
6. Heterogeneous effects by baseline adherence

In this section, we explore potential mechanisms and heterogeneous effects to better understand why Flexible Target incentives have a greater impact than Fixed Target incentives. From our conceptual framework the degree of improvement from increased motivation to act towards a goal depends on how far they are from the goal. To explore predictions from this outlined framework in section 3, we analyse treatment impact and 6-month aggregate adherence, by participant adherence level at baseline: low-adherers (<60), medium adherers (greater than 60, less than 90) and high adherers (>90). There are 69 low adherers, 65 medium-adherers, and 74 high-adherers in the sample of 208 with measured data.6

Given that there are no apparent temporal trends in treatment impact month-to-month, we create a 6-month aggregate measure of adherence and plot k-density curves of its distribution across the three groups, for each level of baseline adherence as shown in Figure 8. Since the aggregate outcome measure does not account for baseline differences between groups, we also run DID regressions of treatment impact for Pooled vs. Control, Fixed Target vs. Control, Flexible Target vs. Control, conditional on each level of baseline adherence.

The first hypothesis 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 theoretically obtain greater psychological value and motivation from movements towards it. We find some suggestions of this. In Figure 8, we see that low adherers seem to benefit the most in the flexible group - among this group, we see 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 6 months in the study peaks around 50 percent; for the Fixed Target group it peaks around 65 percent; for the Flexible Target Group it is around 80 percent.

---

6 We used different cut-offs as robustness checks and observed results that were essentially identical. 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.
From plotting the means over time, we also see that among low-adherers, there is a large increase immediately after the intervention started among Flexible Target participants (Figure 9). From the estimation results in Table 7 this is equivalent to a 16.5 percentage point increase in adherence, with statistical significance at 5 percent (column 8). By contrast, the coefficient among low-adherers in the Fixed vs. Control regression is 1.5 percentage points and not statistically significant.

The second hypothesis discussed in the conceptual framework is that medium adherers do better in the Fixed Target group. From the k-density curves for 6-month mean adherence in Figure 8, the distribution shift towards higher adherence values is slightly greater for Fixed Target than for Flexible target. When comparing Fixed Target to Control in a regression framework, we also see that the average null effect found in Table 5 may be masking meaningful heterogeneity. The coefficient on the treatment x time interaction is statistically significant among those with baseline medium adherence, but not for high or low adherers.
Among this group, the Fixed Target arm improved adherence by 16.1 percentage points relative to the control group (p-value <0.01). The corresponding treatment effect among medium-adherers comparing Flexible vs. Control is slightly lower at 15.9 percentage points with significance at 5 percent.

Table 7. Heterogeneous treatment effects, by baseline adherence level

<table>
<thead>
<tr>
<th></th>
<th>Pooled vs. Control</th>
<th>Fixed vs. Control</th>
<th>Flexible vs. Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Treat x Post</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>0.104</td>
<td>0.155</td>
<td>0.165*** 0.167*** 0.028</td>
</tr>
<tr>
<td>Medium</td>
<td>0.179*** 0.002</td>
<td>0.189*** -0.017</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.077) (0.052) (0.026)</td>
<td>(0.075) (0.052) (0.031)</td>
<td>(0.074) (0.059) (0.026)</td>
</tr>
<tr>
<td>Treat</td>
<td>0.030</td>
<td>0.099</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>(0.053) (0.027) (0.013)</td>
<td>(0.064) (0.030) (0.013)</td>
<td>(0.063) (0.032) (0.019)</td>
</tr>
<tr>
<td>Post</td>
<td>0.216***-0.128** -0.020</td>
<td>0.216***-0.128** -0.020</td>
<td>0.216***-0.128** -0.020</td>
</tr>
<tr>
<td></td>
<td>(0.059) (0.049) (0.022)</td>
<td>(0.049) (0.049) (0.022)</td>
<td>(0.053) (0.049) (0.022)</td>
</tr>
<tr>
<td>Observations</td>
<td>503 495 571</td>
<td>308 342 395</td>
<td>375 309 334</td>
</tr>
<tr>
<td>Individuals</td>
<td>69 65 74</td>
<td>43 45 51</td>
<td>44 51 41</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.167 0.132 0.006</td>
<td>0.131 0.165 0.016</td>
<td>0.179 0.108 0.005</td>
</tr>
</tbody>
</table>

Note: Standard errors in parenthesis are clustered at the individual level. "Low" is baseline adherence <60%, "Medium" is baseline adherence between 60 and 90% and "High" is baseline adherence >90%.

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

Our third hypothesis is that high adherers will do better in the Flexible Target group, though the difference may be small since everyone in the Fixed Target 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. However, as 90 is known widely a clinically “good enough” standard, there is less room for further improvement up to 100 and any differences we do find may not be detected with our small sample size. In Figure 9, we find that among baseline high-adherers, all three groups peak at about 95 percent, those in the Flexible Target group have a distribution with less variance and greater density of observations between 90 and 100. From the estimation results, the coefficient on treat x time for baseline high-adherers shows an adherence improvement of 2.8 percentage points of Flexible Target vs. Control compared to a decline of 1.7 among Fixed Target vs. Control, though none of these are statistically significant.
6. Internal validity and robustness checks

In our primary analysis, 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, requiring a replacement cap due to malfunction or having lost it, or failing to show up for multiple clinic visits. 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. To address this concern, we did a series of pairwise comparisons for each baseline variable shown in Table 1, comparing between Pooled Treatment vs. Control, Fixed vs. Control
and Flexible vs. Control, for each month of the study. This yields a total of 18 balancing tables for the six study months and three sets of comparisons. 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 Figure 3-5).

In Appendix Figures 3-5, 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, meaning that treatment and control groups are balanced along these observables in each period of the study. As such, missing data in each period are not likely causing time-varying selection bias in the sample. Since our main analysis uses difference in differences, we are also controlling for time-invariant unobservable characteristics.

Another concern is that because game eligibility is contingent on EDM adherence as measured by bottle openings, participants may force bottle openings without taking the medication. We have 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 from qualitative evidence we know that most participants are only vaguely aware how their adherence is calculated, apart from knowing that the more pills they take the higher the adherence; to extract the data, they simply place the bottle upside-down against a flat reader. Participants are not aware that each opening counts as one dose. Second, if forced openings were common study coordinators would see multiple sequential openings on the EDM charts (Appendix Figure 1 shows an example of a EDM chart) every time they checked the participants’ data. This has never been reported and when we asked the study coordinators if they believe participants are trying to game the system, they have a strong opinion that it is not the case. Third, if indeed a participant did open their cap multiple times in a day to account for a number of future doses, this would be discounted in our analysis as we top code daily doses to be capped at the regimen. In addition to these reasons against believing EDM data do not accurately reflect actual pill-taking behavior, we can explicitly test for evidence of gaming the system in the following way.

Recall from the design of the study that participants were explicitly told that prize drawings were conditional on adherence within the last month of their game visit. In the EDM extraction software, the study coordinators can select the exact window of date to check adherence and
hence if a clinic visit corresponding to a game occurs March 20, they select February 20 – March 19 as the window. They reinforce this by informing participants, “in the last month, you did [not] reach your target.” This allows us to perform a robustness check: we discard the last four weeks before clinic visits and only examine those weeks in which adherence was not explicitly conditioned. Considering that there may be a level effect across all groups resulting from a final push to adhere well before a clinic visit we also discard such observations among the control group. This robustness analysis ensures that any potential forced bottle openings wherein a pill is not actually taken is not driving the intervention effect. If results are robust to this analysis, then it also suggests that participants are changing their behaviors even when there is 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. If a clinic visit occurs 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 more noisy than the full sample analysis.

Figure 10 shows mean weekly adherence over 24 weeks by study group, omitting the four weeks prior to clinic visits. The Flexible Target and the Control group track closely together in the pre-intervention weeks; once the intervention started, the Flexible Target group mean adherence increases and though there is a more noise week-to-week within groups, the Flexible Target group remains about 5-10 percentage points higher than control throughout the 24 weeks. We verify this using a difference in differences analysis over 24 weeks (Table 8). Even excluding the 4 weeks before each visit, 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 12 percentage points is of the same magnitude as that found in the main analysis using monthly adherence and the full sample. The pooled effect is also of similar magnitude – 6 percentage points, though is no longer statistically significant at p-value<0.1 as it was previously. Overall, our results appear to be robust to this sub-sample analysis.
Figure 10. Mean Weekly Adherence over 24 Weeks – Robustness Check

Notes: Excludes 4 weeks before clinic visits (control group) and game visits (treatment groups).

Table 8. Intervention Treatment Effect on Mean Adherence over 24 weeks, Group Comparisons - Robustness Check

<table>
<thead>
<tr>
<th></th>
<th>Pooled vs. Control (1)</th>
<th>Fixed vs. Control (2)</th>
<th>Flexible vs. Control (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treat x Post</td>
<td>0.063</td>
<td>0.005</td>
<td>0.123**</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.044)</td>
<td>(0.050)</td>
</tr>
<tr>
<td>Treat</td>
<td>0.057</td>
<td>0.119**</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>Post</td>
<td>0.076**</td>
<td>0.076**</td>
<td>0.076**</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>Individual + Time FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>4,541</td>
<td>4,541</td>
<td>3,008</td>
</tr>
<tr>
<td>Individuals</td>
<td>208</td>
<td>208</td>
<td>139</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.053</td>
<td>0.052</td>
<td>0.667</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
7. Discussion and limitations

This paper adds to the emerging literature demonstrating the effectiveness of small “nudge” incentives on health behaviors. We find that incentives in the form of small mobile airtime lottery rewards resulted in significant improvements in medication adherence among adolescents and young adults. A second contribution is that we tested whether allowing participants to set their own subgoals results in improved outcomes, and find that the overall result is driven by the Flexible Target group. Treatment impact is robust to several validity checks, such as discarding bottle openings for the months leading up to each game – the period where adherence was explicitly checked against the target (and participants knew this). In addition to ensuring that the results are not driven by false bottle openings during this time, this provides additional evidence that young patients in our sample are intrinsically motivated to reach a goal they personally set, even when they know that their good behavior does not affect the chances of receiving a reward.

We also find that starting point relative to a goal matters as predicted by the value function derived from Prospect Theory. Patients who have low adherence at baseline see large and statistically significant increases in adherence when they chose their own target, as opposed to no effect when given a fixed target that is relatively distant to their starting position. While the average treatment effect for the Fixed Target group was small and not statistically significant, patients who are “medium” adherers at baseline increased adherence by 15 percentage points (p-value<.01) when the goal reference point was high but within reach.

While we applied the value function framework as a useful way to guide this analysis of heterogeneity and results are generally in line with predictions, an important caveat is that the literature on Prospect Theory and goal-setting is largely silent on how goals should be assigned. Given that in our study design the adherence targets chosen by Flexible target participants were endogenous, we are unable to fully isolate the effect of “sub-goaling”. A strict test of “sub-goaling” against fixed goals requires the research team to set a series of exogenous, individualized goals that are adjusted in each game depending on previous performance.
Absent of this design that was not feasible as it would have required a much larger sample than the one we were able to recruit, other potential mechanisms could influence treatment effect. One separate but related mechanism is that goal choice may instill in participants is the 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 intrinsic motivation (Sue-Chan and Ong 2002). The same objective target of 90 percent could elicit greater intrinsic motivation in the participant who actively selected it compared to one to whom it was assigned. Our analysis assumed this is 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 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 our small experiment suggest this is a plausible mechanism driving our results. A related limitation of this study is due to our small sample size the heterogeneous analysis is not sufficiently powered to detect small differences. While those results should be interpreted with some caution, the fact that they are in line with pre-defined hypothesis based on a conceptual framework suggests that this is a promising avenue for future research in a study with a larger sample.

As an alternative way of gaining insights into the pathways through which the intervention 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?” The plurality of responses indicate that participants 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 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) Examples of such a response include:

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.

I consider my performance first, if I am at 90, then I make it 95 and upwards after seeing that the previous target was achieved but if I don’t make it, I leave it there (at previous target).

I would challenge myself. I started with 95% the first times basing on how I was taking my medicine then I went to 100% and even when I could not achieve it, I decided to put it at 100%.

At the very beginning I started with a lower percentage and then as time went on, I decide that I would increase the percentage to motivate me more to take my medicine.

Responses among some high baseline adherence participants suggested they anchored their goal choice to their pre-intervention behavior:

I wanted it to be 100 because I love my life and I have no problem taking my medicine.

For me, I hate missing drugs, so those days I missed, I went to work and forgot and I was very disappointed. I usually take my medicine, that is why I usually set it at 95 or 100%. I take my medicine.

Other responses suggested participants had a strong intrinsic motivation to adhere well; these participants saw the Flexible Target intervention as a way to boost intrinsic 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

I did not like to miss any medications that is why I used to put a high target to enable me fulfill my target and not to forget.

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, in line with current counselling practices. 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).

A final point is that adherence is a useful indicator for health but ultimately is a health output/behavior and not an outcome. We are currently collecting endline data on viral loads among the study sample and as such, in a future iteration of the paper will include an assessment of whether incentives are effective in improving health outcomes (viral suppression).

8. Conclusion

Incentives are increasingly used to encourage healthy behaviors with wide applications. Our study shows that an important consideration in the design of such incentives is starting point. If motivation and subsequent effort is starting-point dependent, then setting an absolute threshold across the distribution may not have a uniformly positive effect. We found that for low baseline performers, setting the conditionality threshold in a way that allows participants to choose their own targets had a pronounced treatment effect that is larger 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. The results from this study has
implications both for the use of such incentives to help the specific problem of ART adherence among youth, and for the design of conditional requirements for incentives in general.

References


Bor, J., et al. (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): 1459-1469.


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

Notes: This figure shows the time and date of each pill-taking event based on this participants’ electronically monitored adherence.
Appendix Figure 2. EDM Adherence from September 2016 to April 1 2017, once-daily regimen

Notes: This figure shows the time and date of each pill-taking event based on this participants’ electronically monitored adherence.
Appendix Figure 3. Comparing Pooled Treatment vs. Control along Baseline Variables in Each Study Month

Notes: 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 4. Comparing Flexible Target vs. Control along Baseline Variables in Each Study Month

Notes: 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 4. Comparing Fixed Target vs. Control along Baseline Variables in Each Study Month

Notes: 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.