

## Measuring Employment in Developing Countries

### Evidence from a Survey Experiment

*Rachel Heath<sup>1</sup>, Ghazala Mansuri<sup>2</sup>, Dhiraj Sharma<sup>2</sup>, Bob Rijkers<sup>2</sup> and William Seitz<sup>2</sup>*

**Abstract:** How do reference period and survey modality impact labor market reporting? The Ghana High Frequency Labor Market Data Experiment demonstrates that shorter reference periods in labor market surveys result in reporting of a significantly higher incidence of self- but not wage employment spells, and higher self-employment income. They also reduce the reported duration of both self- and wage employment spells in terms of days and hours worked. These impacts are in part driven by recall bias. Moreover, phone-based surveys yield significantly lower estimates of self-employment incidence, income, and duration than face-to-face interviews, while the labor market reports of the wage employed are not impacted by survey modality. Using recall periods shorter than a week may substantially improve the accuracy of labor statistics, especially in developing countries where self-employment is rife.

**Keywords:** Labor Market Surveys, Survey Design, Self-Employment, Africa, Phone-based surveys

<sup>1</sup>Department of Economics, The University of Washington, <sup>2</sup>The World Bank

We would like to thank Kathleen Beegle, Andrew Dabalen, David Newhouse, and Vasco Molini for helpful comments and suggestions, and Mary Hallward-Driemeier and Paolo Falco for their critical support in the early stages of the project. Arthur Lagrange, Jonathan Lain, Lokendra Phadera, Vaclav Tehle, and Asmus Zoch provided excellent research assistance. The research was supported by funding from the World Bank's Research Support Budget. The views expressed in this paper are entirely those of the authors; they do not necessarily represent the views of the International Bank for Reconstruction and Development /World Bank and its affiliated organizations or those of the Executive Directors of the World Bank or the governments they represent.

## 1 Introduction

Accurately measuring employment is essential for effective policy design, but labor market statistics are sensitive to the method by which they are collected. They vary *inter alia* with the wording of the question, who answers the question (Beegle et al., 2012), survey modality (Groves, 1990, De Leeuw, 1992) and the reference period over which questions are asked (see e.g. Bound et al, 2001, for a review of the literature). Current practice favors the use of a one-week long reference period, even in developing countries with large informal sectors and substantial shares of self-employment, where employment tends to be more volatile and short-lived, such that a shorter reference period might provide a sharper snapshot of economic activity (Hussmans et al., 1992). At the same time, new mobile phone based data collection methods render affordable repeated surveys at short intervals, which are especially valuable for monitoring outcomes subject to sudden and frequent changes (McKenzie, 2012), such as the earnings and hours worked of the self-employed.

Against this backdrop, this paper reports the results of the Ghana High Frequency Labor Market Data Experiment (GHFLMDE), conducted between August 2013 and May 2014, with respondents of the Ghana Urban Household Panel Survey (GHUPS), designed to examine how compliance and labor market reporting vary with reference period and survey modality, and how reporting differences depend on labor market status, i.e. between the wage employed, self-employed and those not working. The experiment tracked the time use and labor market behavior of a subset of 1579 respondents of the Ghana Urban Household Panel Survey for a period of six months using four separate monitoring instruments; (i) a baseline face-to-face interview, (ii) high-frequency interviews spanning 10 consecutive weeks, (iii) an endline face-to-face interview conducted approximately three months after the baseline, and (iv) a follow-up phone survey conducted three months after completion of the endline interview (i.e. six months after the start of the survey).

The experimental design was composed of three treatment and two controls arms. The control arms were only interviewed at baseline, endline, and in a three-month follow-up phone interview while the treatment arms also participated in repeated interviews. One group was interviewed by phone three times a week for a period of 10 weeks, another was interviewed by phone once every week over the same period, and the third was interviewed face-to-face each week. The impact of survey modality on labor market reporting is identified by comparing the labor market reports of those assigned to the

weekly phone arm with those of subjects in the weekly face to face arm. The impact of the reference period on reporting is identified in two ways. First, the labor market reports of individuals in the tri-weekly treatment arm are compared with those of individuals who took part in the weekly high-frequency interviews, to identify the impact of moving from a reference period of a week to a reference period of two days. Second, labor market reports obtained during the repeated high-frequency interviews are compared to reports provided by controls who took part in a one-shot retrospective endline survey to assess differences between quarterly and high-frequency (i.e. weekly and two-day) recall. The paper ends by comparing the aggregated high frequency reports of treated individuals with their own endline reports, which allows an examination of recall bias.

Previewing our main findings, shorter reference periods result in reporting of a higher incidence of work, but of shorter average duration, i.e. lower intensity, and higher incomes. The effects are economically meaningful; having a reference period of two days instead of a week increases the probability of reporting having done any work over the course of a week by 8% at the sample mean, but also results in reporting 0.7 fewer days and 5 fewer hours conditional on reporting any work that week. The higher incidence of work and higher reported incomes are driven by increased reporting of self-employment spells as well as recording of employment spells of those who were not working at the baseline interview. Shortening the reference period does not change the reported incidence of the wage employment significantly, but does reduce the number of hours and days reported by wage employees, as well as their weekly income.

Qualitatively similar effects are observed when comparing labor market reports based on repeated weekly interviews with a one-shot retrospective labor market report over a three-month period; repeated interviews with a reference period of one week are associated with an 11% higher chance of recording having done any work than surveys with a three month recall period at the sample mean. Repeated weekly surveys also result in reporting of significantly fewer weeks worked, as well as fewer days and hours per week conditional on working. The increased incidence of employment is again driven by the self-employed; for wage employees no significant differences in reported incidence of work are detected between repeated and retrospective interviews, though wage employees on average report working 0.7 fewer days and 8 hours less per week when interviewed each week as opposed to a one shot retrospective survey.

These differences are likely at least in part driven by recall bias; comparing the reports of treated individuals with their aggregated high-frequency reports reveals that even individuals who had been reporting regularly over the course of the high-frequency survey made systematic mistakes when asked to report on their aggregate labor market behavior over a three month period. Relative to their high frequency reports, they under-reported the odds of working by 2%, but over-reported the number of days worked per week by 0.6 days.

Survey modality impacts labor market reporting. Weekly phone-based surveys yield much lower estimates of employment, hours and days worked than weekly face-to-face interviews. Such reporting differences are mostly driven by differential reporting by the self-employed, for whom work tends to be characterized by greater volatility and less clearly delineated boundaries between work and professional tasks. For the wage-employed, differences are not significantly different, such that the null hypothesis that phone and face-to-face interviews are perfect substitutes for wage employees is not rejected. We speculate that this set of results may indicate the presence of greater social desirability bias for face-to-face interviews, which results in upward estimates of the labor input of self-employed individuals. An important contribution of the paper is thus to show how the impact of survey modality and reference period on labor market reporting varies with the type of employment.

The results also demonstrate the feasibility of cost-effective high-frequency labor market data collection in an urban context characterized by high mobility and substantial volatility. Survey compliance was high across all treatment arms, with over 99.9% of respondents completing the endline survey, 90.8% of people completing the 3-month follow-up survey, and 97.5% of respondents indicating a willingness to continue participating in future data collection efforts. These high compliance rates compare favorably to previous studies, which tend to suffer higher non-compliance (see e.g. Dillon, 2012, Croke et al., 2014; Garlick et al., 2015). The cultivation of trust and building rapport between enumerators and respondents appears to have been an important facilitator for high retention. Moreover, data quality was high, as verified by a random 5% check of responses, and exhaustive interview call logs.

The results suggest there may be scope for revisiting the decision to use a week as the standard reference period in labor market surveys. Substantial accuracy may be gained by using shorter recall periods, especially in developing countries where self-employment is rife, since it appears that the

labor market reporting of the self-employment is most sensitive to survey design. In addition, the paper attests to the potential of mobile phone based data methods to yield timely and cost-effective labour market data, but also draw attention to the need to be cognizant of potential modality induced changes in labor market reporting.

The rest of the paper is organized as follows. Section 2 elaborates on the experimental design. Section 3 analyses survey compliance. Section 4 analyzes the effect of survey frequency and modality on labor market reporting during the high frequency interviews. Section 5 examines differences in labor market reports obtained from repeated interviews with retrospective labor market reports. Section 6 examines recall bias by comparing individuals labor market reports at endline to their own reports provided during the high frequency survey. Conclusions are presented in section 7.

## **1. Experimental Design**

### **2.1 Hypotheses and related literature**

The GHFLMDE was designed to examine the impact of the length of the reference period and survey modality on labor market reporting.

A substantial literature has shown that retrospective reports are prone to recall bias, whose magnitude depends inter alia on the salience of events to be recalled, social desirability, and the reference period. Longer reference periods have been associated with increased recall bias as well as a changes in labor market reporting,<sup>1</sup> but few studies have focused on labor market reporting differences over short

---

<sup>1</sup> While summarizing the literature on recall error is beyond the scope of this paper (see e.g. Bound et al, 2001 and Beckett et al, 2001 for summaries of related literature), a number of earlier studies relating to the collection of labor market data merit note. Comparing company records of employment with reports recorded in the Panel Study of Income Dynamics, Mathiowetz and Duncan (1998), for example, find that response errors regarding employment are lowest for months of and prior to the actual interview (see also Horvath, 1982),<sup>1</sup> and when spells are either characterized by full employment or complete unemployment. Subject appear to have more difficulty recalling spells of short duration. Similarly, Pierret (1998) documents that, switching from annual to biennial interviews in the National Survey of Youth 1979 results in reporting of

reference periods,<sup>2</sup> even though the question what reference period is optimal when collecting labor market data has been the subject of extensive debate (Hussmans et al., 1992).<sup>3</sup> The ILO identifies both a day and a week as appropriate reference periods corresponding closely to an instantaneous measure of employment and being less vulnerable to memory-dependent errors which arise over longer period of recall. Most labor market surveys use a reference period of a week, because of the practicality of measurement and consistency with other sources. When fulltime paid employment is the norm, using weekly as opposed to daily recall has the additional advantage of resulting in a lower variance whilst still giving similar average results. However, the ILO also cautions that when intermittent work, casual work, and short-term employment are widespread, as is the case in developing countries, shorter reference periods might enhance accuracy. On the assumption that self-employment is more volatile than wage employment, one would anticipate a greater impact of the choice of reference period on the reporting of the self-employed than that of the wage employed.

A closely related issue is the measurement of the currently active versus the usually active population. Current activity is typically measured using short reference periods and provides a snapshot picture of the economy at a given point in time. In the presence of seasonality or substantial economic fluctuations a one shot measure may give a misleading representation of the labor market. In such cases, accuracy can be improved by either taking repeated measures or staggering the survey, or using longer reference periods, i.e. measuring the usually active population. These measures will tend to give different results not only because they are conceptually different, but also because of recall error. To the extent that self-employment is characterized by greater fluctuations, one might anticipate such recall error to disproportionately impact them rather than wage employees, whose employment may

---

both fewer employers and fewer spells of non-employment, since short spells of employment or non-employment appear to be forgotten.

<sup>2</sup> A notable exception are the time use studies of Duncan and Stafford (1980), Hammermesh (1990) and Robinson and Bostrom (1994) surveyed in Bound et al. (2001) in which reports based on daily time diaries are compared to labor force survey questions about the number of hours worked in a week. The diary-based estimates yield consistently lower estimates of the number of hours worked in a given week.

<sup>3</sup> There are a number studies that examine the impact of using short reference periods on the reporting health and care seeking behavior (e.g. Das et al., 2012) and consumption (e.g. Beegle et al., 2010) with longer recall periods being associated with under-reporting of health shocks and consumption.

be less prone to fluctuations and hence easier to recall. Measures of usual employment may also fail to pick up sporadic employment episodes by those who are typically not working.

The impact of the choice of reference period on labor market reporting is identified by comparing labor market reports of individual interviewed on a weekly basis with the aggregated weekly reports of those interviewed three times a week. These reports are also compared to the retrospective reports of a control group, in order to assess differences in reporting in retrospective surveys asking about individual's *usual* activity and repeated (high-frequency) surveys asking about their *current* activity. The survey design also allows the reports of treated individuals to be compared to their own retrospective labor market reports obtained after completing the high frequency surveys, enabling an assessment of recall bias.

Turning to modality, phone-based interviews may suffer higher rates of non-response (De Leeuw, 1992) and may result in different reports than face-to-face interviews. While the jury is out on exactly how interview mode impacts labor market reporting, most studies of differences between phone and face-to face interviews only find limited differences (see e.g. De Leeuw, 2005). To assess the impact of modality in the context of labor market surveys, some respondents were randomly assigned to an arm where they were visited by enumerators on a weekly basis and interviewed face-to-face, whereas others were assigned to weekly phone interviews.

One possible concern is that the act of participating in the high frequency survey itself might impact the behavior and/or reporting of respondents. For instance, receiving a phone could induce job search,<sup>4</sup> even though in our sample 95.7% of high frequency survey participants already owned a mobile phone. To test for this possibility, respondents were randomly assigned to two control arms. The first control arm participated in the baseline, endline, and three month rounds, and received a phone during the baseline interview. The second control arm participated in the same rounds as the first but did not receive a phone at baseline. Neither control group participated in the high frequency

---

<sup>4</sup> Mobile phone ownership and usage have inter alia been associated increased search (Tack and Aker, forthcoming), imparting of basic skills (Aker et al, 2012), and increased migration (Aker et al., 2011). More generally, access to information technology has also been associated with reductions in price dispersion (see e.g. Jensen 2007, Goyal, 2012) and improved risk sharing (Jack and Suri, 2014); See Aker and Mbiti, (2010) for an overview of the literature.

interviews. The difference in reporting of the two control groups allows an assessment of the impact of gifting a phone on labor market reporting.

## 2.2 Design

### *Sampling frame*

The sample for the experiment was drawn from the Ghana Urban Household Panel Survey (GHUPS). GHUPS is a panel labor market survey administered by the Centre for the Study of African Economies at the University of Oxford. Eight rounds of GHUPS were completed between 2004 and 2013.<sup>5</sup> The survey experiment was prepared prior to the fielding of the 2013 round. The sampling frame for the experiment consisted of individuals who had been interviewed in the GHUPS, excluding (i) individuals under age 20 or above age 60 in 2013, (ii) individuals not contacted in either 2010 or 2012, and (iii) individuals located in Takoradi-Secondi. The sampling frame consisted of 2251 individuals from 720 households. From this list, subjects were allocated into five treatment arms (described in more detail below) comprising 320 individuals each while ensuring balance on a range of observable characteristics including gender, age, occupation, marital status, dependency ratio (number of dependent-age household members/number of employment age household members), asset ownership, and mobile phone ownership. Randomization was done at the household level, with all individuals from the same household being allocated to the same treatment arm in order to avoid intra-household spillovers arising from assigning members from the same household to different treatment arms.

### *Treatment arms*

The high frequency experiment was composed of five arms with the following features, summarized in Table 1.

### *Baseline survey and enrollment*

A face-to-face baseline survey was fielded prior to interviewing respondents at high frequency, i.e., on a weekly or triweekly basis. The survey served multiple functions, including familiarizing respondents with the survey design and questionnaire, collecting baseline information on key variables of interest,

---

<sup>5</sup> The surveys were conducted in 2004, 2005, 2006, 2008 (with 2007 being a recall wave), 2009, 2010, 2012, 2013.

collecting contact information, and distributing cellphones to the designated experiment arms. All respondents provided their phone numbers, often more than one, and indicated their preferred phone number for completing phone-interviews. To incentivize participation, the respondents in the control arm that did not receive phones, and the treated individuals in the face-to-face weekly interview group which likewise did not receive phones, received 2 Cedis for completing the baseline interviews. Individuals assigned to the weekly phone, triweekly phone, and control with phone arms all received a phone (irrespective of whether they already owned one) with a SIM card and 1 Cedi of working credit.

### *Endline and follow-up surveys*

The endline survey was conducted after the completion of the high frequency interviews, approximately three months after the baseline survey. Another follow-up survey designed to track and monitor the longer term impact of the experiment on the reporting of labor market outcomes was conducted 3 months after the endline survey. The survey also collected data to help extrapolate tri-weekly observations with information on labor market behavior over the past two days to weekly observations.<sup>6</sup> At endline, individuals in the control arms received 2 Cedis (0.91 USD), while treated individuals received their cumulative payouts. At follow-up, all individuals received 4 Cedis (1.82 USD) worth of airtime credit.

## **2.3 Sample Description and Balance**

Individuals in the sampling frame were assigned to treatment arms such that there was balance in sex, education, age, occupation, marital status, and mobile phone ownership status of respondents across the arms (Table 1). At the household level, balance on the dependency ratio and asset ownership index was ensured. Aside from being balanced on variables used for stratification, the treatment arms were also balanced (due to random selection) in other dimensions as well, including on employment status,

---

<sup>6</sup> With the exception of a handful of interviews, the triweekly interviews were conducted on the same days of the week – Tuesday, Thursday, and Saturday – where the respondents were asked to recall their labor market outcomes in the last two days. As a result, Saturday was consistently missed in the high frequency reporting. The follow-up survey aimed to fill this gap by inquiring about respondents’ current and past labor market behavior on Saturday.

self-employment, number of jobs, income per week, hours of work per week, and the city of residence.

Table 2 documents the baseline characteristics of the high frequency survey participants. Note that the sample sizes per treatment arm differ slightly from the target of 320, with 318 individuals participating in weekly face to face interviews, 315 individuals in weekly phone interviews, 321 individuals in the tri-weekly phone arm, 314 controls receiving a phone and 311 pure controls. Just under three-fifths of respondents were female, and the average age was 35 years. Approximately 68 percent were employed, with 62 percent among those working reporting being self-employed. Very few – about 4 percent of the respondents – have more than one job. Average income per week is 87.24 Ghanaian Cedis, or approximately 40 USD<sup>7</sup> and the respondents work 34 hours per week on average. With the exception of income per week, the average values of key socio-economic variables reported by high frequency survey respondents are close to the average values reported by urban residents in the Ghana Living Standard Survey (GLSS) 2012/13, a nationally representative household survey (see Appendix A). The poverty profile of households in high frequency survey matches that of the GLSS households as well.<sup>8</sup> Thus, the sample used for the GHFLD experiment is broadly representative of Ghana’s urban population of working age.

## **2.4 Strategies to reduce attrition and enhance data quality**

Survey fatigue was a potential hazard in this setting. A variety of strategies were adopted to mitigate attrition and deterioration in data quality.

### *Minimize survey burden*

To minimize survey burden high-frequency surveys were designed to take no more than five minutes each. Using Computer Assisted Personal Interviewing (CAPI) software that allowed time invariant information to be stored in the questionnaire automatically, the high frequency survey focused specifically on those questions that were expected to vary at a high frequency. This approach was

---

<sup>7</sup> Using average exchange rate between the US dollar and Ghanaian new Cedis during the survey period, between August and December of 2013 (1 USD = 2.2 GHS) (<http://www.oanda.com/currency/historical-rates/>).

<sup>8</sup> While the GHUPs does not contain detailed consumption information, an asset index, derived from the first principal component of a bundle of assets, is similar for the samples of GLSS 6 and high frequency survey participants.

successful; the average length of phone interviews was less than four minutes. During the baseline interview, respondents were also asked to indicate a preference for the timing of interviews, to minimize the survey burden.

#### *Incentives for participation and retention*

The provision of phones and interview payments was designed to maximize reciprocity and retention. At baseline, a mobile phone was given to each respondent in three arms – triweekly phone interviews, weekly phone interviews, and control with phone – regardless of whether they owned a phone. To maximize compliance, those who already owned a SIM card were given the option to receive phone calls on their existing number to avoid forcing them to adopt a new number specifically for the survey. Payouts were disbursed after the completion of the survey to incentivize continued compliance.

#### *Enumerator incentives*

Many of the enumerators in the high frequency survey had a long-standing relationship with the survey team. The high frequency survey was implemented on the heels of the 2013 round of panel survey, which allowed the team to select the best performing enumerators. The prospect of future employment opportunities helped discipline enumerators, who were furthermore aware that a log of all calls made from the phones they were asked to use for the high frequency survey was kept. This verification check was designed to audit date, time, and duration of calls, and to record information about call attempts, network problems, and other usage statistics.

#### *Nudging mechanisms for temporary dropouts*

If respondents missed two interviews in succession, a supervisor followed up with them to inquire about the reason for the omission and to encourage them to continue to talk to the survey team. In practice, this protocol was not implemented very frequently.

#### *Minimize social distance and social desirability bias*

Reducing social distance between enumerators and respondents can be important in mitigating survey fatigue and therefore reducing attrition and increasing compliance (Croke et al. 2014). Enumerators were paired with respondents for the entirety of the survey across all arms. They were encouraged to interview respondents of their gender. In addition, the questions on wellbeing were asked first to convey interest in individuals' welfare before inquiring about labor market outcomes.

## 2.5 Data quality verification checks

A variety of data quality checks were performed throughout the survey. The use of a CAPI system facilitated the monitoring process as aberrant or incomplete surveys could be sent back immediately for completion or correction (see e.g. Caeyers et al., (2012) for evidence that CAPI systems help improve data quality). In addition, information entered by enumerators in the handheld device was verified against the data from the call logs stored in the phone. The call log cannot be tampered with easily and is thus arguably a more accurate source of information. For every completed interview in the handheld, we sought at least one record of a call made to a phone number that was provided by respondents at baseline or endline on the date of the interview registered in CAPI or a day before or after. By this criterion, 95.1% of the calls in the handheld can be matched with the call log. The small fraction of calls that remain unmatched is likely due to respondents asking the enumerators to call them at a number not in our record.

After the completion of the endline survey, five percent of the respondents were randomly selected for a verification survey. The respondents were asked to verify a number of responses and survey implementation actions. Specifically, they were asked whether they received a phone, if the phone was sealed in a box, how often were they interviewed per week, what their employment status at baseline was, and what economic activity they were engaged in at the time of the baseline interview. As seen in the Appendix their responses are remarkably consistent with their responses during baseline and attest to the credibility of the collected data.

## 2. Survey Compliance

Survey compliance over the course of the survey was remarkably high as is shown in Table 2, which documents attrition rates for the endline, follow up and high-frequency rates. Only 6% of scheduled weekly face to face interviews were not completed, while respectively 11,2% and 10,8% of the weekly and triweekly phone based high-frequency interviews were not completed. Remarkably, interviewing people more often does not lead to higher non-compliance, but there is a statistically significant difference in compliance between face to face and phone-based surveys consistent with earlier studies that have documented higher non-response in phone based surveys relative to face-to-face interviews (see e.g. De Leeuw, 1992). Appendix Table X which provided a breakdown of compliance rates by

round, demonstrating that compliance was not only high overall, but also did not decay over the course of the survey; if anything compliance was lowest in the first survey rounds, when respondents were still settling into their preferred schedules.

Moreover, attrition was typically temporary in nature; survey retention rates between baseline and endline are very high with only 2 out of 1578 individuals not being interviewed at both endline and baseline. The three month phone follow-up survey was less successful in retrieving subjects, yet still resulted in low attrition rates, with 9% of people who participated in the baseline not being surveyed. In the three month follow-up round, attrition rates were slightly, but not significantly, lower for the treated arms than the control arms, with those in the phone-based arms having the lowest attrition rates, perhaps reflecting habituation to phone-based surveys. The lower attrition between baseline and endline than between endline and follow-up survey likely reflect a combination of incentives for respondents to partake in the endline round, since the bulk of the compensation was conditional on completing an endline survey, and the three month follow-up survey being conducted over the phone only; while subjects that did not respond were contacted multiple times, they were not visited by enumerators.

Though it is hard to establish with certainty what drove high compliance it is likely that a combination of a light survey load and customized interview schedules, the cultivation of trust, and the focus on wellbeing may all have contributed to limited temporary and permanent attrition. In addition, financial incentives seemed to have been adequate and the majority of participants were satisfied with the survey, with the overwhelming majority claiming they would be willing to participate again.

One concern is that survey fatigue may manifest itself in a decay of data quality rather than non-compliance. For instance, respondents may learn that by claiming not to work they could speed up the survey, as this would allow them to skip a number of work-related questions (even though the interviews were short by design). However, the data do not show a drop in the share of people claiming to be employed over the course of the survey as is shown in Figure 1 which documents the share of people employed across different treatment arms by week, with the three tri-weekly observations being aggregated to a full week. The trends are fairly flat and, if anything, marginally upward sloping for the

people in the face-to-face weekly interview arm. Concerns about survey fatigue are furthermore alleviated by analyzing call logs which attest to the quality of the collected data.

### 3. Reporting Differences In High Frequency Surveys

#### 4.1 Empirical Strategy

To assess the impact of survey modality and frequency on labor market reporting within high frequency surveys, we estimate regressions of the form:

$$Y_{it} = \beta_{P1}W1Phone_i + \beta_{P3}W3Phone_i + \tau_t + \varepsilon_{it}$$

Where  $Y_{it}$  is a labor market variable of interest, notably whether individual  $i$  did any work that week, days worked per week, hours worked per week and log weekly income, reported by individual  $i$  in week  $t$  and  $W1Phone_i$  and  $W3Phone_i$  are dummy variables indicating whether individual  $i$  was assigned to the weekly phone or the triweekly phone treatment, respectively. The omitted category is thus that of weekly face-to-face interviews, which serve as a useful benchmark since face-to-face interviews remain the dominant survey mode for both labor and household surveys. The term  $\tau_t$  is a vector of calendar week dummies and  $\varepsilon_{it}$  is a random error term. Note that triweekly labor market reports are aggregated to the weekly level in order to be able to compare them with weekly labor market reports obtained in the other arms.

The coefficient  $\beta_{P1}$  provides an estimate of the impact of *survey modality* on reported labor market behavior and is informative about the extent to which phone-based reports differ from conventional face-to-face interviews. To assess the impact of *reference period* on labor market reporting, comparisons between the triweekly and weekly phone treatments are instructive, since these only differ in terms of frequency. Under the assumption that the effects of frequency and modality are additive, the differential between the triweekly treatment phone treatment and the weekly treatment,  $\beta_{P3}-\beta_{P1}$ , provides an estimate of the impact of frequency on labor market reporting (to see this, note that the triweekly phone treatment differs from the conventional face-to-face group both in terms of modality and frequency of reporting, whereas the weekly phone group differs from the face-to-face treatment

only in terms of modality). Under the null hypothesis of no impact of survey modality and reference period on labor market reporting:  $\beta_{P1} = \beta_{P3} = 0$ .

## 4.2 Results<sup>9</sup>

Table 4 presents the results using as dependent variables respectively, employment status in the past week (i.e. any work reported, column 1), days worked per week, days worked per week conditional on working (columns 2 and 3), hours worked, hours worked per week conditional on working, (columns 4 and 5) log income and log income per week conditional on working (columns 6 and 7). Note that for wage employed individuals earnings are computed by multiplying the wage rate with the number of hours they work; individuals who have paid leave, either vacation or sick leave, are excluded from this analysis. Bootstrapped standard errors are clustered by household since treatment was assigned at the household level.

### 4.2.1 Survey Modality Effect

Focusing first on the impact of modality, respondents assigned to the weekly phone treatment report significantly less employment and less income than respondents in the face-to-face high frequency survey and the effects are both statistically and economically significant. Relative to face-to-face interviews phone-based surveys are associated with a significant reduction in the likelihood of reporting any work of 11% at the survey mean. Reported number of days worked per week drop by 0.7 overall, and by 0.2 conditional on reporting any work; the reported number of hours worked drop by 8 overall, and by 5 conditional on reporting having done any work. Reported incomes are also significantly lower, though it should be noted that income conditional on reporting any work is not on average significantly different across the two treatment arms. These effects are driven by modality, since the face-to-face phone and weekly questionnaires are identical.

### 4.2.2 Reference Period Effect

To assess the impact of the reference period over which the question is asked it is instructive to compare the reports of subjects in the tri-weekly phone arm to those in the weekly phone arm, as the

---

<sup>9</sup> The results presented in this section are also robust to controlling for individual and household characteristics, as well as including initial conditions. These results are not presented here to conserve space, but available upon request.

reporting differences between these groups can be attributed to differences in the reference period over which questions are asked. Compared to the weekly phone arm, respondents in the triweekly arm are significantly more likely to report having done any work (column 1), but also report working 0.7 fewer days and 5 hours less conditional on doing any work (columns 3 and 5) such that the aggregate number of days (column 2) and hours (column 4) reported is not significantly different from those reported by their counterparts in the weekly phone treatment arm. Triweekly respondents also report significantly higher incomes (columns 6 and 7). Thus, shortening the reference period from a week to two days is associated with reporting significantly higher income and more employment spells, but not with significantly different total amounts of days and hours reported. One possible explanation for these findings is that more frequent reporting improves accuracy in the recording of the timing of employment spells, but not of the overall amount of time spent working.

Note that the triweekly phone respondents report significantly fewer hours and fewer days worked than those in the weekly face-to-face group, though not a lower incidence of work. Such reporting differences between face to face weekly respondents and those in the triweekly phone arms are driven by a combination of modality and reference period effects. These can countermand each other, as seems to be the case in the reporting of the incidence of work (with phone based surveys resulting in reporting of less work but shorter reference periods resulting in reporting of more work), or reinforce each other, as seems to be the case in the reporting of hours worked conditional on working.

### **4.3 Heterogeneity**

To what extent do these reporting effects vary with labor market status, if at all? As discussed in section 2.1 one might anticipate differences between individuals who are working and those who are not, or only engaging in sporadic employment which may be more accurately captured by questions with a shorter reference period. Moreover, results may vary with the type of job; Self-employment jobs are typically associated with greater flexibility and volatility in hours worked, and less clearly delineated boundaries between work and personal tasks. Wage jobs, by contrast, are characterized by explicit (though not necessarily written or formalized) contracts and a clearer delineation between professional and private tasks. Ex ante, one might expect greater impacts on reporting among the self-employed, whose labor market behavior is notoriously more difficult to capture adequately.

Reporting impacts vary across groups, as is shown in Table X which replicates the specifications presented in Table X above, separately for those reporting wage employment at baseline (panel A), those reporting self-employment at baseline (panel B), and those not reporting any work at baseline (panel C). Note that labor market status is very stable over the course of the survey so the type of employment at baseline is a good measure of the person's employment type during the course of the survey.

Modality has an especially strong impact on the self-employed. For the wage-employed, survey modality does not significantly impact labor market reporting; none of the labor market outcomes reported by those in the weekly phone treatment are on average different from those who were interviewed face-to-face once a week. By contrast, outcomes reported by the self-employed in the weekly phone are all significantly different from those reported by the self-employed who were interviewed face-to-face each week and the differences are large. Being assigned to the phone as opposed to the face-to-face treatment is associated with a lower propensity to report work, fewer days and hours worked, and less income, though not conditional on reporting any work. Those not working at baseline are an intermediate case; they report significantly fewer days and hours conditional on working.

The length of the reference period impacts the reporting of all three groups, but not in uniform fashion. Among the wage-employed the triweekly phone treatment is associated with reporting fewer days and hours worked and less imputed income but not a significantly different propensity to report working, both relative to the weekly phone treatment and to the weekly face-to-face groups. Among the self-employed, completing triweekly phone interviews is associated with reporting a higher incidence of work, more income and fewer days worked conditional on working relative to completing weekly phone interviews. Compared to self-employed individuals in the face-to-face group those in the triweekly phone treatment arm report fewer days and hours worked, but higher income conditional on hours worked. Amongst those not working at baseline, being assigned to the tri-weekly phone as opposed to the weekly phone treatment is associated with a reduction in the reporting of the number of days worked conditional on doing any, and an increase in earnings. Thus, the increased incidence of work and higher earnings associated with shortening the reference period observed for the sample as a whole is driven by the differential reporting of the self-employed.

To summarize, modality does not significantly impact reporting of the wage-employed and aggregate differences in labor market reporting between face-to-face and phone-based surveys are primarily driven by differential reporting of the self-employed. The choice of the reference period has a different impact on the labor market reports of the wage employed, self-employed and those not working. Shorter reference periods are associated with an increased incidence of self-, but not wage-employment. Incomes reported by the self-employed and those not working at baseline go up, but imputed income from wage work goes down, reflecting a reduction in the number of days and hours wage employees report having worked. Overall, the labor market reports of the self-employed appear most sensitive to survey design.

#### 4. Reporting Differences Between Retrospective and Repeated (High Frequency) Surveys

##### 5.1 Econometric Strategy

How do labor market reports obtained during repeated high-frequency survey compare with a one-shot retrospective labor market report using a three month reference period? Answering this question not only helps assess how the length of the reference period impacts reporting but is also informative about reporting biases induced by asking about respondents' *usual* instead of their *current* activity. Retrospective questions about labor market activity over a three month reference period asked about respondent's usual activity, e.g. "*over the past three months, on how many days per week did you do this work on average?*". By contrast, high-frequency interviews ask about their actual labor market behavior in the past week or preceding two-days. Comparing aggregated weekly and tri-weekly reports from repeated interviews to the one-time report of a "typical" week in the last three months thus reveals the extent of misreporting in usual activity questions. Since the high frequency survey spanned a 10 week period, whereas the endline retrospective questions cover a 3 month period, we convert labor market reports to a typical (i.e. average week). This also helps to avoid having to adjust for missing high frequency interviews.<sup>10</sup>

Reporting differences are identified by estimating the following regression:

---

<sup>10</sup> This difference arose as a consequence of keeping questions reasonably simple. It was decided that rather than refer to individual-specific interview dates, the questionnaire would be uniform and that questions would cover the period over which the high frequency survey was in the field (for all respondents).

$$Y_{it} = \beta_{F1}W1F2F_i + \beta_{P1}W1Phone_i + \beta_{P3}W3Phone_i + \beta_{CP}ControlPhone_i + \varepsilon_{it}$$

Where  $Y_{it}$  is a labor market report of interest,  $W1F2F_i$ ,  $W1Phone_i$ ,  $W3Phone$ , and  $ControlPhone_i$  are dummy variables indicating individuals were assigned to the weekly face-to-face, weekly phone, triweekly phone and control with phone arms respectively. Of primary interest is the coefficient  $\beta_{F1}$ , which reflects the differences in labor market reporting between repeated surveys with weekly recall conducted face-to-face and a one-shot survey over the three month period. Since these interviews were all conducted face-to-face, the differences in labor market reporting can be ascribed to differences in reference period and attendant frequency of reporting only. The reports of individuals in the weekly phone and triweekly phone treatment arms are not only impacted by shorter reference periods but also by a difference in the modality of the survey. Note that the coefficient  $\beta_{CP}$  measures the impact of receiving a phone on labor market reporting.

## 5.2 Basic Results

Table 5, panel A, presents the results using as dependent variables respectively, whether the individual in question reported any work over the course of the three month recall period or the 10 weeks of high frequency interviews (column 1), the number of weeks worked (column 2) including conditional on reporting any work over the period (column 3), the number of days worked per week on average (column 4) and conditional on working (column 5), the number of hours worked per week on average (column 6) and conditional on working (column 7) and log average income per week (column 8) and log average weekly income conditional on working (column 9). Bootstrapped standard errors are clustered by household since treatment was assigned at the household level.

Shortening the reference period from three months to a week impacts labor market reporting in a qualitatively similar way as shortening it from a week to two days. Relative to pure controls, i.e. controls who did not receive a phone, respondents in the weekly face-to-face treatment arm were significantly more likely to report having worked over the over the survey period, with the marginal probability of them reporting any work being 11% higher at the survey mean, and also reporting significantly higher log average earnings (column 8). However, they also reported having worked approximately half a day less (column 5) and approximately 6 hours less (column 7) per week than the

pure controls conditional on working that week. Shortening the reference period over which labor market questions are asked thus yields statistically significantly and economically meaningful different labor market reports, reflected in reporting of more frequent employment spells of shorter average duration and higher incomes. These results not only resonate with the results presented in section 4, but also with earlier work examining recall bias in labor market reports (see e.g. Bound et al., 2001).

Other findings of interest are that receipt of a phone is associated with a marginal increase in reported days worked conditional on working as well as an increase in the number of hours worked, albeit that differences between face-to-face high-frequency respondents and pure controls are only significant at the 10% level. In addition, individuals in the high frequency phone arms report systematically fewer days and hours and higher earnings than the two control groups. Individuals in the triweekly phone arm also have higher odds of reporting any work, while those in the weekly phone arm report fewer weeks. These results are consistent with shorter reference periods yielding reporting of more a higher incidence but lower average duration of employment spells.

## **5.2 Heterogeneity**

How do such reporting differences vary with labor market status? Table 5, panels B, C and D, replicate B replicates Table 5, panel A, separating workers who (i) reported being wage-employed at baseline from (ii) workers who reported being self-employed at baseline and (iii) workers who reported not to be working at baseline. The impact of shortening the reference period on reporting is clearly contingent on individuals' labor market status. While the aggregated labor market reports of those who were not working at baseline and participated in face-to-face weekly interviews are not statistically different from the labor market reports of pure controls over the same period, save for reporting higher earnings, labor market reports of those employed at baseline are very different. Those who were wage-employed at baseline and participated in the face-to-face weekly interviews reported working significantly fewer weeks conditional on reporting any, as well as fewer hours and days, both on average and conditional on working, relative to pure controls who reported being wage-employed at baseline. By contrast, the self-employed in the weekly face-to-face group are more likely to report having done any work, and report working more weeks, but fewer days and hours per week conditional on working than the pure controls who were self-employed at baseline. Note that these results are qualitatively similar to those obtained in section 4, with shorter reference periods disproportionately

impacting the reported incidence of self-employment, and also resulting in a reduction in the number of days and hours reported by the wage employed.

Interestingly, the impact of receiving a phone on labor market reporting also varies with labor market status, having little impact on the labor market reporting of the wage-employed, other than to increase the number of weeks reported conditional on reporting any, but a strong impact on the labor market reporting of the self-employed at baseline. Phone receipt increases the likelihood that the latter will report any work, as well as the number of weeks they report. It also increases the number of days and hours they report working, both on average and conditional on reporting any work. Among those not working at baseline the transfer of a phone is associated with a reduced probability of reporting any work, and a reduction in the reported number of weeks, days and hours worked.

## 5. Recall Bias

To what extent are the findings documented above driven by recall bias? Comparing endline reporting of labor market behavior over the 3 month period preceding the endline survey with a respondent's own high-frequency reports obtained during the same time period allows an assessment of the extent to which retrospective reports over a 3 month period are likely to be biased, if at all (on the premise that the reports obtained by means of the high frequency period are more accurate). Moreover, differences in recall bias across treated groups are informative about the impact of the length of the reference period and modality.

To assess the extent of recall bias, the high frequency reports are extrapolated to a three month period. Subsequently, recall bias is assessed by estimating the following regression:

$$\Delta\bar{Y}_i = Y_{iEndline} - \bar{Y}_{iHF} = c + \rho_{P1}W1Phone_i + \rho_{P3}W3Phone_i + v_i$$

Where  $Y_{iEndline}$  is a labor market variable reported during the endline survey, and  $\bar{Y}_{iHF}$  the corresponding variable derived from aggregating high frequency labor market reports and  $c$  is a constant reflecting average reporting differences for the omitted category, respondents who took part in weekly face-to-face interviews, and  $W1Phone_i$  are  $W3Phone_i$  dummy variables that take value 1

if individual are assigned to the weekly phone or triweekly phone arm respectively. If  $\Delta\bar{Y}_i > 0$  individuals *over-report* in the endline one-shot retrospective survey, whose design is similar to that of conventional labor market surveys enquiring about usual activity, relative to their own repeated reporting in the high-frequency surveys. The constant  $c$  is of particular interest since it provides an estimate of the misreporting of individuals in the omitted category, the face-to-face weekly interview arm; since both the endline and their high-frequency interviews were conducted in person a change in modality cannot be the cause of changes in labor market reporting; differences are due to the reference period only. Note that this analysis likely underestimates recall bias since, as is demonstrated in Appendix X, the retrospective endline reports of treated individuals differed somewhat from the reports of the controls, being closer to on average to their own aggregated weekly or triweekly reports than those of the pure controls. Put differently, the act of repeated surveying seems to help individuals remember their labor market behavior more accurately (again, assuming that more frequent reports are more accurate).

The results presented in Table  $\hat{\phantom{a}}$  are suggestive of significant and substantial misreporting. At endline, employment (column 1), the number of weeks worked (column 2 and 3), the number of hours worked per week conditional on doing any work (column 7), as well as earnings per week conditional on working (column 9) are under-reported relative to high frequency reports. This is evidenced by the significantly negative constant in these three regressions, which reflects the average misreporting of the omitted category, those who participated in the weekly face-to-face interviews. By contrast, days (columns 3 and 4) and hours per week conditional on working (column 7) are over-reported. The bias is economically meaningful. For example, using weekly aggregated data as opposed to retrospective reports would increase the probability of reporting any employment by 2.2%, but reduce the reported number of days worked by 0.6 days per week.

The extent of misreporting varies across treatment arms, consistent with the survey impacts report documented in section 4. The tendency to over-report days and hours worked was significantly more pronounced among respondents in the triweekly treatment. These respondents on average reported the lowest hours and days worked during the high frequency survey. Similarly, individuals in the weekly phone treatment were significantly less likely to underreport the number of weeks they worked (recall that the weekly phone treatment was associated with a significantly *lower* propensity to report work

during the high-frequency survey), higher over-reporting of hours worked, and less under-reporting of earnings (since they report lower earnings during the high-frequency survey).

Splitting the sample by whether or not individuals were wage or self-employed at baseline (as is done in Table 6B) provides additional supportive evidence that the reporting differences documented in section 4 are due to survey design. No significant differences in misreporting are detected between individuals in the face-to-face and phone weekly treatment arms that were wage-employed at baseline. By contrast, among respondents who were self-employed at baseline, misreporting varies between these arms. These results are consistent with modality impacting the reporting of the self-, but not the wage-employed.

Underreporting of the incidence and number of weeks worked was prevalent among the self-employed, but not the wage employed, consistent with the reference period disproportionately impacting the reporting of self-employment. Both the self-employed, wage employed and those not working over-report the number of days worked at endline and underreport earnings conditional on working.

Overall the results are suggestive of significant and economically meaningful misreporting of employment in retrospective labor market surveys due to the difficulties inherent in accurately recalling past employment spells. Even subjects who were primed to report accurately by repeated high frequency interviews misreport at endline, relative to their own high frequency reports. The pattern of misreporting is consistent with imperfect recall being one of the reasons for the documented impact of changing the reference period on reporting.

## **6. Conclusion**

Labor market data are central to our understanding of how economies function and evolve, but sensitive to the method by which they are collected. The Ghana High Frequency Labor Market Data Experiment assesses the impact of the length of the reference period and survey modality on a number of salient labor market indicators by means of a randomized control trial. The experiment attests to the feasibility of cost-effective high frequency labor market data collection in an urban context

characterized by substantial volatility, with only two out of 1579 participants not being interviewed at endline and compliance rates with the high frequency interviews exceeding 90%.

The experiment demonstrates that shortening the reference period over which labor market questions are asked results in statistically significantly and economically meaningfully different labor market reports. Compared to the standard reference period of a week, shortening the reference period to two days results in a higher incidence of self-, but not wage-employment, and higher reported self-employment income. Hours and days worked decrease among both the wage and the self-employed. Qualitatively similar results are obtained when comparing retrospective reporting over a three month period with labor market reports obtained from repeated weekly surveys.

Reporting differences are at least in part due to recall bias. Even individuals who partook in the high-frequency surveys and were primed to remember their labor market history tend to underreport working, but over-report the number of days and hours worked, conditional on reporting any work.

Reporting is also impacted by survey modality, with weekly phone-based interviews resulting in significantly lower reports of employment, hours worked, days worked and income amongst the self-employed. For the wage employed, modality does not appear to affect estimates.<sup>11</sup> These findings underscore the need to be cognizant of modality induced labor market reporting differences in labor market surveys.

Arguably the most important implication of the results presented in this paper is that using recall periods shorter than a week, which is currently the norm, may substantially improve the accuracy of labor statistics, especially in developing countries where self-employment is rife.

---

<sup>11</sup> Though establishing why these differences across employment categories arise is beyond the scope of the paper, one possible explanation is that phone-based interviews reduce social desirability bias to which the self-employed may be more susceptible as their jobs and working hours may be less clearly defined than those of wage worker.

## 7. References

- Aker, J.; Clemens, M. and C.Ksoll, (2011): “Mobiles and mobility: The Effect of Mobile Phones on Migration in Niger”, Proceedings of the German Development Economics Conference, Berlin 2011, No. 2
- Aker, J., C. Ksoll, and T. Lybbert. (2012) : "Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger." *American Economic Journal: Applied Economics* 4 (4) : 94–120.
- Aker, J. C., and I. Mbiti. (2010). “Mobile Phones and Economic Development in Africa.” *Journal of Economic Perspectives* 24 (3): 207–
- Bardasi, E., K.Beegle, A.Dillon and P.Serneels (2011) “Do Labor Statistics Depend on How and to Whom the Questions are Asked? Results from a Survey Experiment in Tanzania” *World Bank Economic Review* 25(3),418-447.
- Beegle, K., J. de Weerdt, J. Friedman, and J. Gibson (2012): “Methods of Household Consumption Measurement Through Surveys: Experimental Results from Tanzania,” *Journal of Development Economics*, 98(1), 3–18.
- Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., and Peterson, C. (2001). The quality of retrospective data: An examination of long-term recall in a developing country. *Journal of Human Resources*, 593-625.
- Caeyers, B., N. Chalmers, and J. De Weerdt (2012): “Improving Consumption Measurement and Other Survey Data through CAPI: Evidence from a Randomized Experiment,” *Journal of Development Economics*, 98(1), 19–33.
- Croke, K., A. Dabalén, G. Demombynes, M. Giugale and J. Hoogeveen, (2014):“Collecting High Frequency Panel Data in Africa using Mobile Phone Interviews,” *Canadian Journal of Development Studies*, 35(1), 186–207.

- Das, J., J. Hammer, and C. Sanchez-Paramo (2012): “The impact of recall periods on reported morbidity and health-seeking behavior” *Journal of Development Economics* 96, (2012) 76-88
- De Leeuw, E. (1992) “Data Quality in Mail, Telephone and Face to Face Surveys” Netherlands Organization for Scientific Research.
- De Leeuw, E. (2005) “To Mix or Not to Mix Data Collection Modes in Surveys” *Journal of Official Statistics*, Vol.21, No.2, 2005. pp. 233–255
- De Nicola, F. and X. Gine (2014): “How Accurate are Recall Data: Evidence from Coastal India?” *Journal of Development Economics* 106: 52–65. 2014.
- Dillon, B. (2012): “Using Mobile Phones to Collect Panel Data in Developing Countries,” *Journal of International Development*, 24, 518–27.
- Duncan, G. J., and Stafford, F. P. (1980). Do union members receive compensating wage differentials?. *The American Economic Review*, 355-371.
- Garlick, R., Orkin, K. and Quinn, S. (2015); “Call Me Maybe: Experimental Evidence on Using Mobile Phones to Survey African Microenterprises” *Mimeo*.
- Goyal, A. (2010). “Information, Direct Access to Farmers and Rural Market Performance in Central India.” *American Economic Journal: Applied Economics* 2 (3): 22–45
- Groves, R. M. (1990). “Theories and Methods of Telephone Surveys.” *Annual Review of Sociology*, 16, 221-40.
- Hamermesh, D. S. (1990). Shirking or productive schmoozing: Wages and the allocation of time at work. *Industrial & Labor Relations Review*, 43(3), 121S-133S.
- Husmanns, R, Mehran, F. and Verma, V. (1990). “Surveys of economically active population,

employment, unemployment and underemployment: An ILO manual on concepts and methods.” International Labour Office, Geneva.

Horvath, F. W. (1982). Forgotten unemployment: recall bias in retrospective data. *Monthly Labor Review*, 105(3), 40-43.

Mathiowetz, Nancy A & Duncan, Greg J, (1988). "Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment," *Journal of Business & Economic Statistics*, American Statistical Association, vol. 6(2), pages 221-29, April.

Mc Kenzie, D. (2012): “Beyond Baseline and Follow-up: The Case for More T in Experiments,” *Journal of Development Economics*, 99(2), 210–221.

Paul, G. (2002): “Biases in the Reporting of Labour Market Dynamics” Institute for Fiscal Studies Working Paper No 02/10.

Pierret, C (2001) Event History Data and Survey Recall. *Journal of Human Resources*, 36, 439-466.

Jack, W. and T. Suri. (2014): "Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution." *American Economic Review*, 104(1): 183-223.

Jensen, R., (2007) “The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector”, *Quarterly Journal of Economics*, 122 (3): 879-924.

Robinson, J. P., and Bostrom, A. (1994). The overestimated workweek? What time diary measures suggest. *Monthly Labor Review*, 11-23.

Tack, J. and J. Aker, (forthcoming) “Information, Mobile Telephony, and Traders' Search Behavior in Niger” *American Journal of Agricultural Economics*

## Tables

**Table 1: Survey Design**

<b>Interviews</b>							
	Baseline	High frequency interviews		Endline	3 Month Follow up		
		<i>weekly</i>	<i>3x weekly</i>				
<b><i>Treatment Arms</i></b>							
Weekly Face-to-face	✓	✓		✓		✓	
Called 1x week	✓	✓		✓		✓	
Called triweekly	✓		✓	✓		✓	
<b><i>Control Arms</i></b>							
Control without phone	✓			✓		✓	
Control with phone	✓			✓		✓	
<b>Incentives</b>							
	Received a phone	Weekly phone credit	<i>Participation Fees (per completed interview)</i>				
			Baseline	High frequency interviews		Endline	3 Month Follow up
				<i>weekly</i>	<i>3x weekly</i>		
<b><i>Treatment Arms</i></b>							
Weekly Face-to-face			3	3		3	4
Called weekly	✓	1	3	3		3	4
Called triweekly	✓	1	3		2	3	4
<b><i>Control Arms</i></b>							
Control without phone	✓		3			3	4
Control with phone	✓		3			3	4

**Table 2: Survey Compliance**

Non-Compliance (Attrition)				
	N	High Frequency Interviews	Endline (in person)	3-Month Followup (by phone)
<b><i>Treatment Arms</i></b>				
Weekly Face-to-face	318	6.4%	0.0%	9.4%
Called 1x week	315	11.2%**	0.3%	7.4%
Called triweekly	321	10.8%**	0.0%	8.4%
<b><i>Control Arms</i></b>				
Control without phone	314		0.0%	10.5%
Control with phone	311		0.0%	10.6%
Total	1579		0.1%	9.2%

Note: \*\* in columns 1 and 2 indicate attrition rates are significantly different from those of the pure controls who are not receiving phones at the 5% level. In column 3 they indicate non-compliance rates are significantly different from those observed for individuals in the face-to-face weekly interview arm.

**Table 3: Balance in Sample Frame**

	Mean	Mean difference with controls without phones (cn)				
		Treatment	Face-to-face	Weekly phone	3X week phone	Control with phone
	Control without phone	Control without phone	Control without phone	Control without phone	Control without phone	Control without phone
<b><i>Household Variables</i></b>						
Household size	6.230	-0.122	-0.197	-0.097	-0.067	-0.054
		-0.122	(0.499)	(0.482)	(0.500)	(0.500)
Dependency ratio	0.200	(0.411)	0.029	-0.004	-0.015	0.000
		0.003	(0.027)	(0.027)	(0.027)	(0.027)
Number of adults employed	0.970	(0.022)	-0.199	0.017	-0.121	-0.302**
		-0.104	(0.148)	(0.158)	(0.155)	(0.149)
Highest years of education by any	11.770	(0.125)	0.014	0.355	0.179	0.266
		0.179	(0.401)	(0.396)	(0.409)	(0.415)
Asset index	0.155	0.071	0.009	0.125	0.084	0.019
		(0.173)	(0.203)	(0.225)	(0.198)	(0.216)
N		457	233	225	229	223
<b><i>Individual Variables</i></b>						
Male	0.399	0.021	0.041	0.009	0.013	0.015
		(0.032)	(0.037)	(0.041)	(0.041)	(0.040)
Married	0.403	0.030	0.054	0.013	0.022	0.015
		(0.037)	(0.047)	(0.045)	(0.045)	(0.046)
Age	35.826	0.510	0.986	-0.048	0.586	0.152
		(0.745)	(0.874)	(0.856)	(0.921)	(0.834)
Years of education	9.798	-0.206	-0.470	-0.133	-0.011	-0.191
		(0.303)	(0.369)	(0.363)	(0.370)	(0.343)
Employed	0.594	0.014	0.008	0.024	0.009	0.038
		(0.034)	(0.042)	(0.043)	(0.041)	(0.042)
Self employed	0.282	0.029	0.061	0.009	0.015	0.034
		(0.030)	(0.038)	(0.037)	(0.036)	(0.037)
Log income per week (total)	2.478	0.080	0.194	0.025	0.021	0.158
		(0.139)	(0.170)	(0.173)	(0.163)	(0.166)
Hours worked per week (total)	31.470	-1.793	-1.984	-1.762	-1.630	2.934
		(2.097)	(2.487)	(2.598)	(2.627)	(2.664)
N		1211	603	600	600	598

Income per week is winsorized at 5th and 95th percentiles. Hours worked per week is winsorized at 1st and 99th percentiles.

**Table 4A: Reporting in High Frequency Surveys – All individuals**

<i>Dependent Variable</i>	Working	Days	Days	Hours	Hours	Income (log)	Income (log)
All Subjects			<i>(conditional on working)</i>		<i>(conditional on working)</i>		<i>(conditional on working)</i>
	Probit	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>marginal effects</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>
1X Phone ( $\beta_1$ )	-0.108*** (21.580)	-0.686*** (0.193)	-0.197** (0.092)	-7.727*** (2.336)	-4.705** (2.249)	-0.589*** (0.157)	-0.103 (0.110)
3X Phone ( $\beta_3$ )	-0.027 (6.420)	-0.683*** (0.196)	-0.882*** (0.110)	-7.321*** (2.083)	-9.770*** (2.096)	0.041 (0.174)	0.330*** (0.113)
Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reference Period	0.080	0.004	-0.685	0.406	-5.065	0.630	0.433
<i>p value</i> ( $\beta_3 = \beta_1$ )	0.019	0.984	0.000	0.828	0.000	0.006	0.000
N. Observations	8,537	8,537	5,125	8,537	5,117	7,360	3,880
R2		0.022	0.081	0.023	0.037	0.018	0.034
Adjusted R2		0.021	0.079	0.021	0.034	0.016	0.030
Pseudo R2	0.010						

Note: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ , bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly interviews. Standard errors in column 1 computed using the Delta Method

**Table 4B: Reporting in High Frequency Surveys By Initial Labor Market Status**

<i>Dependent Variable</i>	Working	Days	Days	Hours	Hours	Income (log)	Income (log)
All Subjects			<i>(conditional on working)</i>		<i>(conditional on working)</i>		<i>(conditional on working)</i>
	Probit	OLS	OLS	OLS	OLS	OLS	OLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>marginal effects</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>
<i>A. WAGE EMPLOYED AT BASELINE</i>							
1X Phone ( $\beta_1$ )	-0.013 (0.481)	-0.087 (0.252)	-0.016 (0.135)	-2.480 (3.218)	-2.271 (3.320)	0.013 (0.286)	0.080 (0.237)
3X Phone ( $\beta_3$ )	-0.026 (0.050)	-0.662** (0.301)	-0.661*** (0.188)	-8.424** (3.602)	-9.084*** (3.370)	-0.290 (0.337)	-0.090 (0.287)
Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reference Period	-0.013	-0.576	-0.645	-5.944	-6.814	-0.304	-0.170
<i>p value</i> ( $\beta_3 = \beta_1$ )	0.756	0.0386	0.000	0.040	0.011	0.3045	0.5008
N. Observations	2,254	2,256	1,829	2,256	1,828	1,270	1,015
R2		0.040	0.065	0.032	0.035	0.010	0.012
Adjusted R2		0.034	0.058	0.026	0.027	-0.001	-0.002
Pseudo R2	0.021						
<i>B. SELF EMPLOYED AT BASELINE</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1X Phone ( $\beta_1$ )	-0.160*** (28.858)	-1.031*** (0.226)	-0.265* (0.143)	-11.481*** (3.264)	-5.747* (2.937)	-0.721*** (0.178)	-0.118 (0.124)
3X Phone ( $\beta_3$ )	-0.028 (5.805)	-0.848*** (0.249)	-0.876*** (0.146)	-8.170*** (3.070)	-8.428*** (2.601)	0.236 (0.196)	0.495*** (0.111)
Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reference Period	0.132	0.183	-0.611	3.311	-2.680	0.957	0.613
<i>p value</i> ( $\beta_3 = \beta_1$ )	0.002	0.458	0.000	0.258	0.309	0.000	0.000
N. Observations	3,571	3,572	2,904	3,572	2,899	3,516	2,681
R2		0.049	0.077	0.043	0.032	0.042	0.066
Adjusted R2		0.045	0.072	0.039	0.027	0.038	0.061
Pseudo R2	0.035						
<i>C. NOT WORKING AT BASELINE</i>							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1X Phone ( $\beta_1$ )	-0.021 (0.041)	-0.183 (0.235)	-0.582** (0.287)	-2.926 (1.950)	-15.668*** (5.482)	-0.077 (0.120)	-0.328 (0.261)
3X Phone ( $\beta_3$ )	0.041 (0.044)	-0.062 (0.233)	-1.619*** (0.206)	-1.714 (2.043)	-20.449*** (3.962)	0.174 (0.159)	0.263 (0.373)
Week Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
( $\beta_3 - \beta_1$ )	0.062	0.121	-1.036	1.212	-4.781	0.251	0.590
<i>p-value</i> (1X Phone =	0.115	0.510	0.001	0.407	0.351	0.094	0.092
N. Observations	2,706	2,709	392	2,709	390	2,574	184
R2		0.013	0.249	0.015	0.207	0.011	0.095
Adjusted R2		0.007	0.223	0.010	0.180	0.006	0.026
Pseudo R2	0.018						

Note: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.10$ , bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly interviews. Standard errors in column 1 computed using the Delta Method

**Table 5A: Reporting Differences Between Retrospective and Repeated High Frequency Surveys**

<i>Dependent Variable</i>	<i>Working</i>	<i>Weeks</i>	<i>Weeks</i>	<i>Days</i>	<i>Days</i>	<i>Hours</i>	<i>Hours</i>	<i>Log Earnings</i>	<i>Log Earnings</i>
			<i>conditional on work</i>		<i>conditional on working</i>		<i>conditional on working</i>		<i>conditional on earning</i>
<b>Estimation Method</b>	<u>Probit</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>
	<i>marginal effects</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>marginal effects</i>	<i>coef/se</i>
<b>A. ALL INDIVIDUALS</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Face to Face	0.106*** (0.037)	0.678 (0.466)	-0.612*** (0.224)	0.045 (0.242)	-0.468*** (0.088)	-0.873 (2.869)	-6.344*** (2.449)	2.453*** (0.252)	-0.099 (0.104)
1X Phone	0.065* (0.035)	-0.909* (0.488)	-2.029*** (0.289)	-0.759*** (0.256)	-0.789*** (0.113)	-9.449*** (2.690)	-12.574*** (2.246)	2.063*** (0.256)	-0.271** (0.108)
3X Phone	0.117*** (0.034)	0.004 (0.477)	-1.558*** (0.288)	-0.824*** (0.242)	-1.601*** (0.109)	-10.255*** (2.633)	-19.332*** (2.113)	2.454*** (0.256)	0.103 (0.113)
Controls - phone	0.058 (0.115)	0.643 (0.492)	0.201 (0.201)	0.396 (0.246)	0.137* (0.079)	5.291* (2.978)	3.306 (2.486)	0.383 (0.335)	0.086 (0.094)
Constant (pure controls)		7.256*** (0.363)	11.044*** (0.148)	3.518*** (0.198)	5.776*** (0.060)	33.177*** (2.285)	54.186*** (1.609)	1.700*** (0.235)	4.488*** (0.071)
<i>p-value</i> (1X Phone = Control Phone)	0.855	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001
<i>p-value</i> (3X Phone = Control Phone)	0.096	0.156	0.000	0.000	0.000	0.000	0.000	0.000	0.876
N. Observations	1.428	1.428	1.026	1.428	1.027	1.428	1.025	969	798
R2		0.013	0.091	0.036	0.273	0.045	0.138	0.157	0.019
Adjusted R2	0.007	0.010	0.087	0.034	0.270	0.042	0.134	0.154	0.014
<b>B. WAGE EMPLOYED AT BASELINE</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Face to Face	0.028 (0.037)	-0.552 (0.543)	-0.770** (0.362)	-0.696** (0.290)	-0.547*** (0.132)	-7.645* (4.127)	-6.294** (3.208)	0.378 (0.382)	-0.316 (0.225)
1X Phone	0.027 (0.038)	-0.917* (0.517)	-1.138*** (0.350)	-0.891*** (0.254)	-0.608*** (0.116)	-10.990*** (3.533)	-8.612*** (2.656)	0.442 (0.357)	-0.257 (0.166)
3X Phone	0.007 (0.042)	-1.124* (0.592)	-1.148*** (0.413)	-1.571*** (0.272)	-1.464*** (0.174)	-18.226*** (3.610)	-17.671*** (2.559)	0.155 (0.370)	-0.337* (0.192)
Controls - phone	0.004 (0.042)	0.446 (0.533)	0.568** (0.258)	0.232 (0.269)	0.020 (0.112)	3.647 (4.349)	1.997 (3.576)	0.223 (0.468)	0.056 (0.159)
Constant (pure controls)		10.511*** (0.388)	11.145*** (0.232)	5.103*** (0.197)	5.807*** (0.078)	48.866*** (2.911)	55.157*** (1.997)	3.516*** (0.330)	4.388*** (0.104)
<i>p-value</i> (1X Phone = Control Phone)	0.502	0.006	0.000	0.000	0.000	0.000	0.002	0.455	0.577
<i>p-value</i> (3X Phone = Control Phone)	0.931	0.004	0.000	0.000	0.000	0.000	0.000	0.222	0.002
N. Observations	435	435	412	435	412	435	412	242	231
R2	0.003	0.031	0.080	0.125	0.269	0.100	0.108	0.557	0.061

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10, bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly interviews. Column 1 reports marginal effects.

**Table 5B: Reporting Differences Between Retrospective and Repeated High Frequency Surveys**

<i>Dependent Variable</i>	<i>Working</i>	<i>Weeks</i>	<i>Weeks</i>	<i>Days</i>	<i>Days</i>	<i>Hours</i>	<i>Hours</i>	<i>Log Earnings</i>	<i>Log Earnings</i>
			<i>conditional on work</i>		<i>conditional on working</i>		<i>conditional on working</i>		<i>conditional on earning</i>
<b>Estimation Method</b>	<u>Probit</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>	<u>OLS</u>
	<i>marginal effects</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>coef/se</i>	<i>marginal effects</i>	<i>coef/se</i>
<b>C. SELF EMPLOYED AT BASELINE</b>									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Face to Face	0.095*** (0.033)	0.771* (0.438)	-0.275 (0.284)	0.035 (0.245)	-0.428*** (0.133)	-1.634 (3.552)	-6.175* (3.276)	0.654*** (0.239)	-0.108 (0.136)
1X Phone	0.058 (0.028)	-1.063** (0.513)	-1.815*** (0.354)	-0.981*** (0.285)	-0.829*** (0.160)	-12.898*** (3.409)	-14.155*** (3.088)	0.223 (0.251)	-0.284* (0.147)
3X Phone	0.070** (0.036)	0.304 (0.471)	-0.487 (0.348)	-1.000*** (0.278)	-1.499*** (0.153)	-11.678*** (3.286)	-17.265*** (2.781)	0.960*** (0.244)	0.295** (0.129)
Controls - phone	0.091*** (0.036)	1.038** (0.460)	0.040 (0.269)	0.708*** (0.245)	0.243** (0.107)	10.550*** (3.636)	5.721* (3.047)	0.633** (0.264)	0.018 (0.113)
Constant (pure controls)		9.626*** (0.369)	10.960*** (0.199)	4.663*** (0.197)	5.765*** (0.088)	43.367*** (2.637)	53.712*** (2.229)	3.662*** (0.217)	4.617*** (0.093)
<i>b-value</i> (1X Phone = Control Phone)	0.247	0.000	0.000	0.000	0.000	0.000	0.000	0.038	0.023
<i>b-value</i> (3X Phone = Control Phone)	0.416	0.072	0.101	0.000	0.000	0.000	0.000	0.054	0.022
N. Observations	639	639	603	639	604	639	602	604	587
R2	0.021	0.042	0.064	0.113	0.266	0.109	0.142	0.036	0.024
Adjusted R2	0.015	0.036	0.058	0.108	0.261	0.103	0.136	0.069	0.013
<b>D. NOT WORKING AT BASELINE</b>									
Face to Face	0.093 (0.069)	0.476 (0.519)	-1.593 (1.018)	0.168 (0.258)	-0.259 (0.312)	1.627 (2.561)	-0.815 (8.157)	4.869*** (0.365)	-0.014 (0.280)
1X Phone	0.125* (0.066)	0.043 (0.449)	-4.536*** (1.060)	-0.097 (0.217)	-1.478*** (0.425)	-1.860 (1.968)	-21.399*** (8.280)	4.362*** (0.405)	-0.312 (0.325)
3X Phone	0.256*** (0.071)	0.668 (0.506)	-4.903*** (0.989)	-0.019 (0.221)	-2.552*** (0.381)	-1.027 (1.978)	-26.316*** (7.816)	4.432*** (0.482)	-0.200 (0.369)
Controls - phone	-0.108* (0.058)	-0.902** (0.386)	-2.614 (1.815)	-0.460** (0.200)	-0.314 (0.404)	-4.320** (1.915)	-7.207 (8.771)	-0.495* (0.284)	0.642** (0.327)
Constant (pure controls)		1.324*** (0.314)	10.214*** (0.693)	0.650*** (0.172)	5.714*** (0.294)	5.784*** (1.737)	49.607*** (7.473)	-1.672*** (0.214)	4.245*** (0.244)
<i>b-value</i> (1X Phone = Control Phone)	0.004	0.010	0.278	0.031	0.007	0.077	0.013	0.000	0.002
<i>b-value</i> (3X Phone = Control Phone)	0.000	0.001	0.198	0.008	0.000	0.018	0.000	0.000	0.015
N. Observations	501	501	103	501	103	501	103	241	67
R2		0.025	0.209	0.018	0.416	0.020	0.313	0.537	0.053
Adjusted R2	0.073	0.017	0.177	0.010	0.392	0.012	0.285	0.529	-0.008

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10, bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly interviews. Column 1 reports marginal effects

**Table 6A: Recall Bias – All Individuals**

<i>Dependent Variable</i>	$\Delta$ Working	$\Delta$ Weeks	$\Delta$ Weeks <i>conditional on working</i>	$\Delta$ days	$\Delta$ days <i>conditional on working</i>	$\Delta$ hours	$\Delta$ hours <i>conditional on working</i>	$\Delta$ Earnings	$\Delta$ Earnings <i>conditional on earning</i>
<b>Endline Report - HF Report</b>	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>ALL INDIVIDUALS</b>									
Constant (F2F)	-0.022** (0.010)	-0.403*** (0.146)	-0.432** (0.197)	0.588*** (0.077)	0.860*** (0.082)	0.249 (0.903)	-4.600*** (1.447)	-0.051 (0.054)	-0.295*** (0.059)
1X Phone	0.006 (0.017)	0.803*** (0.236)	0.942*** (0.288)	0.379*** (0.136)	0.493*** (0.142)	3.645** (1.441)	2.454 (2.062)	0.180* (0.096)	0.130 (0.096)
3X - Phone	-0.003 (0.019)	0.051 (0.202)	-0.187 (0.269)	0.775*** (0.123)	0.899*** (0.146)	3.358*** (1.256)	4.812** (1.896)	-0.106 (0.088)	-0.267*** (0.098)
<i>b-value</i> (1X Phone + Constant = 0)	0.261	0.033	0.029	0.000	0.000	0.001	0.137	0.113	0.023
<i>b-value</i> (3X Phone + Constant = 0)	0.131	0.012	0.001	0.000	0.000	0.000	0.858	0.025	0.000
N. Observations	948	948	668	948	667	948	665	818	515
R2	0.000	0.020	0.032	0.041	0.061	0.011	0.012	0.011	0.039
Adjusted R2	-0.002	0.018	0.029	0.039	0.058	0.009	0.009	0.009	0.036

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10, bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly interviews.

**Table 6B: Recall Bias – By Initial Labor Market Status**

<i>Dependent Variable</i>	$\Delta$ Working	$\Delta$ Weeks	$\Delta$ Weeks <i>conditional on working</i>	$\Delta$ days	$\Delta$ days <i>conditional on working</i>	$\Delta$ hours	$\Delta$ hours <i>conditional on working</i>	$\Delta$ Earnings	$\Delta$ Earnings <i>conditional on earning</i>
Endline Report - HF Report	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<b>A. WAGE EMPLOYED AT BASELINE</b>									
Constant (F2F)	-0.027 (0.019)	-0.226 (0.352)	-0.167 (0.355)	0.926*** (0.156)	1.006*** (0.154)	1.907 (1.902)	-3.674* (2.136)	-0.262* (0.135)	-0.312** (0.138)
1X Phone	0.027 (0.023)	0.735 (0.539)	0.638 (0.486)	0.161 (0.218)	0.130 (0.220)	2.180 (3.051)	0.942 (3.102)	0.399** (0.202)	0.264 (0.172)
3X - Phone	0.027 (0.031)	0.181 (0.452)	-0.131 (0.452)	0.603*** (0.233)	0.526** (0.229)	1.994 (2.885)	2.421 (3.002)	0.504** (0.233)	0.271 (0.224)
<i>b-value</i> (1X Phone + Constant = 0)	1.000	0.178	0.194	0.000	0.000	0.093	0.262	0.373	0.637
<i>b-value</i> (3X Phone + Constant = 0)	1.000	0.879	0.227	0.000	0.000	0.061	0.561	0.215	0.819
N. Observations	254	254	238	254	238	254	238	145	131
R2	0.005	0.011	0.015	0.026	0.022	0.003	0.003	0.036	0.020
Adjusted R2	-0.003	0.003	0.007	0.019	0.014	-0.005	-0.006	0.022	0.005
<b>B. SELF EMPLOYED AT BASELINE</b>									
Constant (F2F)	-0.020* (0.011)	-0.532*** (0.186)	-0.465** (0.194)	0.680*** (0.113)	0.757*** (0.108)	0.810 (1.254)	-3.358** (1.647)	-0.087 (0.074)	-0.293*** (0.071)
1X Phone	0.002 (0.024)	1.123*** (0.365)	1.253*** (0.379)	0.520** (0.206)	0.604*** (0.196)	4.545** (2.155)	3.195 (2.322)	0.162 (0.145)	0.075 (0.115)
3X - Phone	0.028 (0.020)	0.001 (0.272)	-0.172 (0.290)	1.042*** (0.191)	0.974*** (0.186)	5.107*** (1.867)	6.905*** (2.190)	-0.312** (0.124)	-0.418*** (0.107)
<i>b-value</i> (1X Phone + Constant = 0)	0.392	0.060	0.017	0.000	0.000	0.004	0.922	0.534	0.014
<i>b-value</i> (3X Phone + Constant = 0)	0.654	0.009	0.001	0.000	0.000	0.000	0.011	0.000	0.000
N. Observations	393	393	366	393	365	393	363	387	354
R2	0.005	0.036	0.053	0.076	0.081	0.021	0.030	0.030	0.073
Adjusted R2	-0.001	0.031	0.048	0.071	0.076	0.016	0.024	0.025	0.068
<b>C. NOT WORKING AT BASELINE</b>									
Constant (F2F)	-0.022 (0.015)	-0.338* (0.179)	-1.271 (0.854)	0.159* (0.082)	1.104*** (0.363)	-2.029** (0.940)	-18.762*** (4.835)	0.111 (0.115)	-0.250** (0.127)
1X Phone	-0.006 (0.038)	0.445 (0.332)	0.586 (1.043)	0.460** (0.210)	1.211** (0.572)	4.246** (1.829)	9.700 (7.507)	0.069 (0.171)	0.009 (0.350)
3X - Phone	-0.065* (0.039)	-0.022 (0.286)	-0.155 (1.029)	0.601*** (0.174)	1.415*** (0.509)	2.374* (1.270)	7.612 (5.326)	-0.137 (0.145)	-0.693** (0.306)
<i>b-value</i> (1X Phone + Constant = 0)	0.385	0.708	0.229	0.000	0.000	0.172	0.107	0.136	0.454
<i>b-value</i> (3X Phone + Constant = 0)	0.016	0.090	0.018	0.000	0.000	0.642	0.000	0.786	0.001
N. Observations	301	301	64	301	64	301	64	286	30
R2	0.010	0.011	0.012	0.031	0.120	0.026	0.042	0.006	0.162
Adjusted R2	0.003	0.004	-0.020	0.025	0.091	0.019	0.011	-0.001	0.100

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10, bootstrapped standard errors in parentheses, the omitted category is face-to-face weekly

## Appendices

### A. External Validity

To assess the representativeness of the sample used for the Ghana High Frequency Data Experiment, we compare its with the Ghana Living Standard Survey 2012/23 (GLSS6). GLSS6 is a LSMS-type survey that is representative at the urban/rural level within each region. Recall that the high frequency survey interviewed workers between the ages of 20 and 60 in Accra, Kumasi, and Cape Coast, the three largest Ghanaian cities. To make the sample equivalent, the comparison focuses on urban respondents aged 20 to 60 years in Greater Accra, Ashanti, and Central regions, the regions where the cities are located.<sup>12</sup>

Appendix Table 1 shows that the high frequency sample is similar to the restricted GLSS6 sample on a variety of dimensions, notably sex, age, employment and self-employment rates, and household asset index. Although the null hypothesis that the two means are equal is often rejected, the point estimates are close to each other given, all the more remarkable because the high frequency sample was not designed to be representative.

**Appendix Table A1: External Validity**

	High Frequency Sample (GHFLMDE)	GLSS 6 Restricted Sample	Difference	<i>p</i> -value (HF – GLSS)
Individual Characteristics	(N = 1579)	(N = 6332)		
Male	0.41	0.46	-0.05	0.00
Age	35.1	35.46	-0.36	0.25
Education	10.21	12.12	-1.91	0.00
Employed	0.68	0.73	-0.05	0.01
Self-employed	0.42	0.41	0.01	0.63
Household Characteristics	(N = 579)	(N = 3668)		
Asset index	0.02	0.00	0.02	0.80

Note: The GLSS 6 sample is restricted to respondents aged 20 - 60 years residing in urban areas in Central, Greater Accra, and Ashanti regions. Asset index is derived from the first principal component of the following assets: sewing machine, electric stove, gas stove, refrigerator, freezer, air conditioner, radio, television, bicycle, motorbike, car, and mobile phone.

<sup>12</sup> The sample cannot be restricted further because the survey is not representative at lower levels of administration.

## B. Data Verification

**Appendix Table B1: Audit of Random Five Percent of Respondents**

Were you given a phone at the time of baseline interview?						
	Triweekly	Called once a week	Face-to-face interview	Control with phone	Control without phone	Total
NA	0	0	28	0	15	43
No	0	1	0	1	0	2
Yes	24	29	0	15	0	68
Total	24	30	28	16	15	113

  

Was it a new phone (packed in a box)?		Total
NA		43
No		2
Yes		68
Total		113

  

How often were you interviewed per week on average?						
	Triweekly	Called once a week	Face-to-face interview	Control with phone	Control without phone	Total
NA	0	0	0	15	13	28
Once	3	29	26	1	2	61
Twice	2	0	1	0	0	3
Thrice	19	0	0	0	0	19
Can't	0	1	1	0	0	2
Total	24	30	28	16	15	113

  

Employment status	Baseline		
	Not employed	Employed	Total
I was not working at the time of baseline interview	28	3	31
I was working at the time of baseline interview	5	78	83
Total	31	81	114

In addition to matching interviews to call log, a number of diagnostic checks shed light into the implementation of the survey. The first is an assessment of the duration of interview. Average length of interview decreased significantly by 4.5 seconds per round in the weekly call arm ( $p$ -value = 0.00) and by 0.8 seconds in the tri-weekly arm ( $p$ -value = 0.00) (Appendix Table 5, Panel A), while there was no significant change in the face-to-face arm (-1.385;  $p$ -value = 0.24). The null hypotheses that the three coefficients are equal can be rejected at conventional levels of significance ( $p$ -value = 0.002).

There are two plausible explanations for the decline in interview length: learning or fatigue. Two pieces of evidence suggest that the drop was due to learning rather than fatigue. First, the drop in the duration of interview is steep in the first few rounds while it tapers off in the later rounds. This is

consistent with learning because if interview durations were shorter due to fatigue, we would expect the drop to occur in the later rounds. A direct test for fatigue also finds no evidence for it. .

Another margin where learning could have taken place is the number of call attempts per interview. Over time, enumerators could learn things like which phone number to call and what time of day to call to maximize the chance that respondents are reached and interviewed. Regression results show that call attempts per successful interview declined over time, although the decline was not significant. For the weekly call arm, number of call attempts declined by about 0.11 calls per person per interview round while for the tri-weekly arm, the decline was 0.09 calls per person per interview (Appendix Table 5, Panel C). This is further suggestive of learning on the part of enumerators.

### Appendix Table B2: Triangulation with Call Log Data

Triangulation with Call Log Data			
Panel A: Average duration of interview (in seconds)			
	Face to Face	Call weekly	3x weekly
Interview round	-1.385 (1.169)	-4.49** (1.124)	-0.803*** (0.274)
Constant	229.6*** (6.989)	215.76*** (6.771)	179.27*** (4.843)
Observations	3,232	3,048	9,815
R-squared	0.003	0.031	0.012
Number of individuals	314	315	320
Note: The regressions for face-to-face and weekly call arms include individual and month fixed effects. The regression for triweekly arm includes individual and week fixed effects. Duration of interview is winsorized at the 1 <sup>st</sup> and 99 <sup>th</sup> percentiles. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1			
Panel B: Call attempts per person interview			
	Call weekly	Triweekly	
Interview round	-0.11 (0.10)	-0.09 (0.06)	
Constant	4.68*** (0.29)	1.55*** (0.04)	
Observations	7,278	23,283	
R-squared	0.01	0.03	
Number of id	315	320	
Note: The regression for weekly arm includes individual and month fixed effects. The regression for triweekly arm includes individual and week fixed effects. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1			

### C. Attrition and Compliance

**Appendix Table C1: Compliance During the High Frequency Survey By Wave**

<i>Compliance During the High Frequency Survey</i>				
	Weekly Face to Face N=318	Weekly Phone N=315	Triweekly Phone N=321	
<i>Week</i>			<i>Interview</i>	
1	98.7%	100.0%	1	99.7%
			2	80.1%
			3	83.2%
2	76.7%	73.3%	4	84.7%
			5	90.3%
			6	84.4%
3	94.7%	78.7%	7	86.6%
			8	88.2%
			9	91.9%
4	96.2%	89.5%	10	92.2%
			11	90.0%
			12	89.7%
5	96.2%	88.6%	13	92.2%
			14	91.3%
			15	86.3%
6	92.5%	88.6%	16	91.6%
			17	92.8%
			18	91.9%
7	96.2%	92.7%	19	91.0%
			20	88.5%
			21	87.5%
8	93.4%	91.4%	22	87.5%
			23	87.9%
			24	92.2%
9	96.5%	93.7%	25	88.5%
			26	92.5%
			27	91.3%
10	95.0%	91.4%	28	86.0%
			30	87.2%
Average Compliance	93.6%	88.8%		89.2%

**Table C2: Determinants of Compliance in the High Frequency Survey;**

Table C3 presents simple probit models of compliance in the high frequency survey, with face to face weekly interviews being the omitted category, using as dependent variables respectively whether someone completes the next interview, conditional on missing the current one in columns 1-4, and whether someone completes the scheduled interview. Those in the weekly call arms and in the tri-weekly call arms are significantly less likely to comply, irrespective of whether we control for individual characteristics, as is done in columns 2 and 6. There is also weak evidence that self-employed individuals may be marginally less likely to comply (see columns 4 and 8), though the effect is small and moreover, the pattern of attrition is unlikely to drive the results documented in this paper; when the reference period is shortened the reported incidence of self-employment increases the most amongst the self-employed even though they might be marginally more likely to drop out.

<i>Dependent Variable</i>	<b>Compliance Probit</b>							
	(marginal effects, standard errors in parentheses)							
	<i>Completes next interview (conditional on completing the current one)</i>				<i>Completes scheduled interview</i>			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Call Weekly	-0.067*** (0.025)	-0.090*** (0.016)	-0.059*** (0.013)	-0.060*** (0.013)	-0.060** (0.024)	-0.084*** (0.015)	-0.078*** (0.014)	-0.080*** (0.015)
Call 3X Week	-0.045** (0.018)	-0.061*** (0.011)	-0.023** (0.009)	-0.023*** (0.009)	-0.048*** (0.018)	-0.064*** (0.011)	-0.058*** (0.010)	-0.059*** (0.010)
Self employed			0.008 (0.006)	0.011* (0.006)				
Wage employed			-0.003 (0.007)	-0.001 (0.007)				
Self employed when last interviewed							0.012 (0.008)	0.016** (0.008)
Wage employed when last interviewed							-0.002 (0.010)	0.002 (0.010)
Individual controls	No	Yes	No	Yes	No	Yes	No	Yes
Number of observations	15,006	14,912	13,497	13,443	15,960	15,860	15,890	15,830
Pseudo R2	0.006	0.013	0.007	0.009	0.006	0.014	0.011	0.013

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1, standard errors are heteroscedasticity robust and clustered by household. Individual controls: male, age, age squared, education, education squared, Ravens score, log family size, dependency ratio, location dummies for Cape Coast and Takoradi. The omitted category is that of controls who did not receive phones.

**Determinants of Attrition in the Follow Up Survey:** Table C3 presents simple probit models of attrition in the 3 month follow-up survey relative to the baseline. The omitted category is the group of controls who did not receive a phone. Overall, attrition is very difficult to predict, as is indicated by the low pseudo R2s, which never exceed 0.027. Column 1 controls for treatment status only, with pure controls (i.e. those that did not receive a phone) being the omitted category. Although subjects in the weekly phone arm are marginally less likely to attrit at the 10% level, the null hypothesis that the treatment dummies are jointly insignificant is not rejected, even after individual characteristics are controlled for, as is done in column 2. Column 3 adds indicators of whether the individual was self-employed or wage-employed at baseline, which do not have significant predictive power; thus attrition in the follow-up survey seems fairly random.

**Table C3: Determinants of Attrition in the Follow Up Survey**

<b>Attrition Probit</b>			
<i>Dependent Variable: Did not complete the follow up survey</i>			
<i>(marginal effects, standard errors in parentheses)</i>			
	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>
Face to Face Weekly	-0.011 (0.023)	-0.011 (0.022)	-0.003 (0.026)
Call Weekly	-0.034* (0.020)	-0.035* (0.020)	-0.029 (0.022)
Call 3X Week	-0.020 (0.023)	-0.020 (0.023)	-0.013 (0.025)
Control Phone	-0.001 (0.023)	-0.001 (0.022)	-0.001 (0.022)
Self employed at baseline			-0.021 (0.022)
Wage employed at baseline			-0.003 (0.023)
Individual controls	No	Yes	Yes
Pr F-test: Joint Sig Treatment	0.467	0.432	0.576
Number of observations	1,579	1,575	1,575
Pseudo R2	0.004	0.020	0.021
Observed	0.092	0.276	0.092
Predicted	0.091	0.432	0.088

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1, standard errors are heteroscedasticity robust and clustered by household. Individual controls: male, age, age squared, education, education squared, Ravens score, log family size, dependency ratio, location dummies for Cape Coast and Takoradi. The omitted category is controls who did not receive phones.

## D: Survey Satisfaction

Table X presents descriptive statistics demonstrating that self-reported survey satisfaction was high, in spite of a minority of respondents considering the survey burdensome; Approximately 1 out of every 10 respondents considered some of the questions uncomfortable and roughly 1 in 5 participants considered the survey invasive/a violation of their privacy. Interestingly, the share of respondents that were relatively dissatisfied with the survey does not appear to vary by treatment. Moreover, more than half of all respondents indicated they would have participated in the survey even if they had not received any compensation, with two thirds of the controls receiving phones most likely to indicate they would have done so. Roughly half of the respondents agreed with the statement that compensation sufficed, with subjects in the triweekly treatment arm, who had the highest survey load but also received the highest financial compensation, being most likely to agree with this statement. Respondents in the control arm without phones, who receives the lowest levels of compensation, were least likely to agree. Very tellingly, more than 97% of respondents indicated being willing to be contacted in three months' time, suggesting that compensation adequately incentivized compliance.

Thus, it seems that the strategy of keeping surveys short and enabling respondent to pick interview times that were convenient for them to minimize taxing respondents induced high levels of compliance. Moreover, compensation appeared adequate, and the strategy of establishing a clear report with subjects appears to have paid off. Another element of our design that might have been important, but is difficult to evaluate is that our respondents had some basic familiarity with surveys by virtue of having participated in the Ghana Urban Household Panel Survey.

### Appendix Table D1: Survey Satisfaction

Table X: Subjective Survey Satisfaction					
<i>Treatment</i>	<i>Uncomfortable Questions?</i>	<i>Privacy Violated?</i>	<i>Willing to participate without compensation?</i>	<i>Compensation Sufficed?</i>	<i>Willing to participate in 3 months?</i>
Control - No Phone	11.3%	22.2%	54.0%	35.4%	98.1%
Control - Phone	9.6%	12.7%***	64.0%**	43.0%	98.7%
Face to Face - Weekly	10.1%	20.8%	51.7%	53.0%***	95.9%
Callexd - 1X Weekly	9.9%	19.4%	52.5%	45.9%*	96.2%
Called - 3X Weekly	9.1%	17.8%	59.1%	65.3%***	98.4%

note: \*\*\*, \*\*, \*, indicate that the means are significantly different from the control - no phone group at the 10, 5, and 1 percent significance level respectively

## E. Aggregating High-Frequency Observations

The approach to calculating Income from work takes into account i) reporting frequency, ii) missing days of the week, and iii) differences in questionnaire design for reporting wage employment vs. self-employment income.

To make interviews for the triweekly arm comparable to the weekly treatment arms, totals for three consecutive triweekly interviews were aggregated into “waves” matching the frequency of the weekly treatment arms. But because the triweekly interviews were conducted on the same days of the week – Tuesday, Thursday, and Saturday – and respondents were asked to recall their labor market outcomes over the preceding two days, Saturday was consistently missed in the high frequency reporting.

### Appendix Table E1: Aggregating High-Frequency Observations

	Total	
	Interviews	Share
Sunday	153	1.6%
Monday	9	0.1%
Tuesday	2,997	31.4%
Wednesday	177	1.9%
Thursday	3,056	32.0%
Friday	131	1.4%
Saturday	3,035	31.8%

A follow-up survey filled this gap by gathering data on current and past labor market behavior on Saturdays. This information was then used to estimate indicators such as income and hours worked on Saturdays to be included in the aggregation for the triweekly treatment arm. To fill in this lack of coverage, an interpolation procedure was implemented when aggregating observations from the thrice weekly treatment arm to the weekly level. First, respondents were asked how often they work on Saturdays, and when they did, for how many hours on average. About 52% of respondents did not ever work on Saturdays. Among those respondents that did work on Saturdays, about 85% on average work on every Saturday throughout the month. Having established these patters, aggregated work

weeks in the high-frequency data were adjusted to reflect the estimated number of total hours for each respondent that reported working on some Saturdays by adding the estimated number of hours worked on Saturdays to the total observed hours for each week. For weeks during which no work was otherwise reported however, no adjustment was made to the total hours worked (i.e. corrections were made conditional on having worked during the week).

Information on wages, pay-period and other information about how income was calculated for respondents engaged in wage employment was gathered during the baseline interview. Wage employed respondents were then asked during each high frequency interview whether anything about their wage or income had changed since the preceding interview. For analysis, income from wage employment was calculated using the latest wage rate available for the respondent, and high frequency data regarding the number of hours, days, or pay-periods during which the respondent worked during a given wave.

The variability of self-employment income would not allow for a similar approach. Instead, self-employed respondents were asked during each interview about how much income they earned over the period of reference, the costs incurred for doing business, and any work completed on account, with the expectation for payment in the future. This allowed for analysis of aggregated income using three different definitions: i) pure income, without accounting for costs, ii) income after costs, and iii) income after costs and including earnings on account.

## F. Impact of the Experiment on Reporting at Endline and 3 Month Follow-up

Appendix Table F.1: Impact of the Experiment on LM reporting at Endline and 3 Month Followup

TX: Labor Market Reporting at Endline and at the Three Month Follow Up						
<i>Dependent Variable</i>	<b>Working</b>	<b>Weeks</b>	<b>Days</b>	<b>Hours</b>	<b>Income (log)</b>	
All Subjects	(1)	(2)	(3)	(4)	(5)	
	coef/se		coef/se		coef/se	
<b>Endline</b>						
Face to Face Weekly	0.252** (0.113)	0.359 (0.464)	0.387 (0.249)	-0.237 (2.770)	0.196 (0.188)	
1X Phone	0.158 (0.117)	-0.159 (0.485)	0.069 (0.258)	-4.254* (2.568)	0.006 (0.200)	
3X Phone	0.295*** (0.107)	-0.103 (0.472)	0.340 (0.235)	-5.498** (2.405)	0.157 (0.180)	
Control Phone	0.176 (0.117)	0.822 (0.502)	0.451* (0.263)	6.595** (2.830)	0.355* (0.198)	
Constant (Controls Without Phone)	0.341*** (0.084)	6.994*** (0.369)	3.679*** (0.189)	31.872*** (1.989)	2.789*** (0.149)	
<i>p-value</i> (Treatment)						
<i>p-value</i> (Control Phone =1X Phone)						
<i>p-value</i> (Control Phone = 3X Phone)						
N. Observations	1,577	1,577	1,577	1,577	1,577	
R2		0.005	0.004	0.021	0.004	
Adjusted R2		0.002	0.002	0.019	0.001	
Pseudo R2	0.005					
<b>3 Month Follow Up</b>						
Face to Face Weekly	0.166 (0.122)	0.733 (0.469)	0.246 (0.262)	3.834 (2.966)	0.344* (0.189)	
1X Phone	0.015 (0.121)	0.231 (0.497)	0.065 (0.239)	0.627 (2.560)	-0.048 (0.190)	
3X Phone	0.073 (0.113)	0.585 (0.474)	0.287 (0.240)	4.372* (2.602)	0.152 (0.186)	
Control Phone	0.016 (0.114)	0.355 (0.498)	0.152 (0.244)	3.162 (2.693)	0.110 (0.192)	
Constant (Controls Without Phone)	0.444*** (0.084)	7.004*** (0.339)	3.321*** (0.176)	30.521*** (1.801)	2.578*** (0.134)	
<i>p-value</i> (Treatment)						
<i>p-value</i> (Control Phone =1X Phone)						
<i>p-value</i> (Control Phone = 3X Phone)						
N. Observations	1,432	1,432	1,432	1,432	1,432	
R2		0.002	0.002	0.003	0.004	
Adjusted R2		-0.001	-0.001	0.001	0.001	
Pseudo R2						

Note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.10, bootstrapped standard errors in parentheses

Table X first documents reporting differences across treatment arms at endline and in the three-month follow-up round. The structure of the table is similar to that of Table X, which presented descriptive statistics on balance.

At endline, treated respondents were more likely to report having been employed at some point during the preceding three months (column 1). This effect is driven by a significantly higher propensity to report employment of those in the face-to-face treatment and the triweekly phone treatment arms (column 4). However, the treated groups do not report working more weeks (column 2), more days (column 3) or higher levels of income (column 5). The weekly and triweekly phone subjects report working fewer hours than respondents assigned to the control arms (column 4). Interestingly, receiving a phone seems to induce a change in subsequent labor market reporting, since the respondents in the control arm that received phones report working significantly more days and hours, as well as earning more income than the pure controls.

Disaggregating between respondents who at baseline reported wage employment, self-employed, and those not working, as is done in the Appendix, unveils that these average differences are almost entirely driven by the self-employed and those not working at baseline. That the treated on average reported working fewer hours, which is predominantly driven by the triweekly phone group, is the only significant difference among the wage-employed. The differences documented at endline are therefore consistent with reporting differences documented when comparing the high frequency reports of treated individuals to the endline reports of pure controls.

Three months later, at the follow-up survey, differences had diminished and largely disappeared. When examining the sample as a whole, that those who participated in the weekly face-to-face interviews report marginally higher levels of income three months later was the only significant difference (though, only at the 10% level). Those who partook in the triweekly phone interviews report working more hours than the controls.

Treatment effects are consistent with the reporting differences documented when comparing aggregated high frequency reports with labor market reports of the pure controls, but attenuate over time, as they are much smaller and typically insignificant in the 3 month follow-up survey than at endline.