Measuring Employment in Developing Countries: Evidence from a Survey Experiment

Rachel Heath[1], Ghazala Mansuri[2], Dhiraj Sharma[2], Bob Rijkers[2] and William Seitz[2]

Abstract: Do survey reference periods 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 results indicate that recall periods shorter than a week may substantially improve the accuracy of labor statistics in contexts were a substantial fraction of the work force in engaged in temporary work or in self-employment.

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

[1] University of Washington, [2] World Bank

We would like to thank Kathleen Beegle, Andrew Dabalen, 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 (Bardasi et al., 2011), 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 (Hussmans et al., 1992). At the same time, survey modality is changing rapidly in developing countries with the penetration of mobile phones. While quite a bit of work has been done on survey mode more generally, and phone based versus face to face surveys, in developed countries, this is among the first papers to look at the impact of survey mode, and the very first to look at the impact of survey reference period, on labor market statistics.

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 labor market reporting and survey compliance vary with reference period and survey modality, and whether reporting differences depend on initial labor market status (wage employed, self-employed or 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 the 10 week high-frequency interviews. Among the three treatment groups, 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 respondents assigned to the weekly phone arm with those of respondents assigned to the weekly face to face arm. The impact of the reference period on reporting is identified in two ways. First, the labor markets reports of individuals in the tri-weekly phone treatment arm are compared with those of individuals who took part in the weekly phone interviews, to identify the impact of moving from a reference period of a week to a reference period of two days, with data collected for two 24 hour periods. 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. Finally, the aggregated high frequency reports of individuals are compared with their own endline reports, which allows for an examination of recall bias within the high-frequency group.

We find that shortening the reference period to 48 hours (in two 24 hour recall questions) substantially increases both the odds of reporting any work and total income over the course of a week, and the effects are economically meaningful. The odds of reporting any work increase by 9 percentage points and income increases by close to 50 percent over the sample mean. Interestingly, conditional on reporting any work, the average duration of work spells declines. Over a week, respondents in the 2 day recall report working .6 days and 3.9 hours less than respondents who were asked about the last 7 days.

However, all of these effects are concentrated among the self-employed. Shortening the reference period does not significantly change the reported incidence of wage employment or wage income. However, wage employees do report fewer days and hours worked, conditional on working.

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; respondents in the weekly interview arm have a 11.5 percent higher odds of reporting any work than respondents the 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.

We show that these differences are largely due to 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, but over-reported the number of days worked per week.

Survey modality also 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 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

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 respondents 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 other studies, which tend to suffer from 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 check of responses for 5% of the respondents, and exhaustive interview call logs from the two high-frequency phone interview arms.

The results underscore the need to reconsider the use of a week as the standard reference period in labor market surveys. Till recently, the lack of reliability of conventionally measured

labor market statistics has mainly been a concern for developing countries with large informal sectors dominated by own account workers. The recent rise of own account, temporary and sporadic work in the US and other developed countries, even among groups that have traditionally been associated with full-time wage employment (often referred to as the 'gig economy'), has made this issue pertinent well beyond the developing world, however. This paper shows that for own account and sporadic workers, substantial accuracy may be gained by using shorter recall periods.

Finally, the paper attests to the potential of mobile phone based data collection methods to yield timely and cost-effective labor market data, given the low attrition rates even after 6 months, but it also draws 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 describes the design of the experiment and the data. Section 3 describes survey compliance results. Section 4 presents the empirical strategy and the main results. Section 4.1 discusses the impact of survey mode. Section 4.2 discusses the impact of shortening the Reference Period. Section 4.3 looks at differences in the impact of a shorter reference period and a shift in survey mode by job type and labor market participation. Section 5 turns to differences between the retrospective report of control respondents and the aggregated reports of high frequency respondents assigned to the three high frequency arms. Section 6 examines recall bias by comparing individuals' labor market reports at endline to their own reports provided during the high frequency survey. Section 7 concludes.

2. Experimental Design

2.1 Hypotheses and related literature

The G-HFLE 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,[1] but few studies have focused on labor market reporting differences over short reference periods,[2] 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).[3] 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

[1] 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),[1] 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 both fewer employers and fewer spells of non-employment, since short spells of employment or non-employment appear to be forgotten.

[2] 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.

[3] 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.

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 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, [4] 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

4 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.

same rounds as the first but did not receive a phone at baseline. Neither control group participated in the high frequency interviews. The difference in reporting of the two control groups allows an assessment of the impact of receiving 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.[5] The survey experiment was prepared prior to the fielding of the 2013 round. The sampling frame for the experimented 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 Appendix 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

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

familiarizing respondents with the survey design and questionnaire, collecting baseline information on key variables of interest, 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. [6] 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 3). 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

[6] 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.

as well, including on employment status, 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[7] 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.[8] 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

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 successful; the average length of phone interviews was less than four minutes.

[7] 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/).

[8] 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.

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. The bulk of the payout for respondent was 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.

3. Survey Compliance

Survey compliance over the course of the survey was consistently high as is shown in Table 4, which documents non-compliance (attrition) rates for the endline, follow up and high-frequency interviews. Only 6.4% 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. Interviewing respondents more frequently thus did not induce higher non-compliance, but the phone modality was associated with lower compliance than the face to face modality, 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 provides a breakdown of compliance rates by survey 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 completing an interview. 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. Lower attrition at endline than during the three month follow-up survey round likely reflects a combination of (i) incentives for respondents to partake in the endline round, since the bulk of the compensation was conditional on completing an endline survey, and (ii) 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 why compliance was high it is likely that a combination of the survey being very short, flexible interview schedules tailored to individual respondents particular (time-varying) preferences, the cultivation of trust, and the framing of the survey as being primarily about respondents' wellbeing all have contributed to low attrition. In addition, financial incentives seemed to have been adequate; the majority of participants were satisfied with the survey, and the overwhelming majority (97.5%) indicated 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 could have learned 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 are at odds with strategic manipulation of responses to minimize survey time; they 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. Note that the responses of individuals in the tri-weekly phone treatment group are being aggregated to weekly observations.

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 analysis of call logs which demonstrates that (i) interviews indeed took place and (ii) the duration of each interview did not diminish sharply over the course of the survey.

4. Empirical Framework and Results

To assess the impact of survey modality and frequency on the reporting of labor market outcomes we estimate a regression of the form:

$$Y_{iw} = \beta_m Phone_i + \beta_r 3XWeekly_i + \tau_w + \varepsilon_{iw} \quad (1)$$

Where Y_{iw} is a labor market variable of interest, notably whether individual *i* did any work that week *t*, days worked per week, hours worked per week and weekly income. $Phone_i$, is a dummy variable that takes the value 1 if respondent *i* was interviewed over the phone and 0 if she was interviewed in person; it thus takes the value 1 for both respondents in the weekly phone and the tri-weekly phone treatment groups. $3XWeekly_i$, is a dummy variable that indicates whether individual *i* was interviewed three times per week. 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 τ_w is a vector of calendar week dummies and ε_{iw} 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. Standard errors are clustered at the household level and bootstrapped.

The coefficient β_m 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. The coefficient β_r provides an estimate of the impact of shortening the *reference period* on labor market reporting. Under the null hypothesis of no impact of survey modality and reference period on labor market reporting: $\beta_m = \beta_r = 0$. Table 5 presents the results.

4.1 The Impact of Survey Mode on Reported Labor Market Statistics

Focusing first on the impact of modality, respondents assigned to the phone treatment arm report significantly less employment and less income than respondents interviewed in person and the effects are both statistically and economically significant. Relative to face-to-face interviews phone-based surveys are associated with a significant 11 percentage points reduction in the likelihood of reporting any work 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 not significantly different from those reported by respondents in the weekly face to face group, perhaps in part reflecting noise in the income measures.

4.2 Reference Period Effect

Respondents who were called tri-weekly, however, are 9 percentage points more likely to report any work. They also recall earning significantly more income. The income effect is not due to more work days or more hours worked per day, however. There is no change in the average reported days of work or hours worked per day, but conditional on working both hours worked and days worked fall. Respondents in the triweekly arm report working 0.6 fewer days and 4 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 treatment arms. Thus, shortening the reference period from a week to two days (with two 24 hour recall questions) yields more employment spells and more income, but not significantly different days or hours worked, conditional on working.

One possible explanation for these findings is that more frequent reporting improves accuracy in the recording of income and the timing of employment spells, but not the overall amount of time spent working. We show below, however, that survey reference periods have very different effects for wage workers and the self-employed.

4.3 Mode and Reference Period Effects: Wage Workers and the Self-Employed

We have seen above that shortening the reference period allows for a more accurate recall of work spells and of income. Are these effects similar across job types? Does a shorter reference

period only increase the odds of catching work spells among the employed or does it also allow for a more accurate assessment of the labor force participation rate? It is possible that individuals who only work sporadically may not view themselves as having been employed over longer reference periods, but would report some work if asked about the last 24 hours. We turn to this issue next.

It is well understood that labor market statistics are less reliable for the self-employed (or own account workers) who typically have both greater flexibility and greater volatility in days and hours worked, less clearly delineated boundaries between work and personal tasks, and a poorer separation of earnings and daily household expenditures. Sporadic workers share many of these characteristics. 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, therefore, one might expect greater impacts of recall periods on reporting among the self-employed and this is indeed the case.

We assess the impact of survey modality and frequency on labor market reporting by adding employment status and job type to equation 1 and looking at interactions between survey mode and reference period and job type. The resulting variation of estimation equation 1 takes the form:

$$Y_{iw} = \beta_m Phone_i + \beta_r 3XWeekly_i + \beta_S SE_i + \beta_{Sm} SE_i * Phone_i + \beta_{Sr} SE_i * 3XWeekly_i + \beta_N NW_i + \beta_{Nm} NW_i * Phone_i + \beta_N NW_i + \beta_{Nm} NW_i * Phone_i + \beta_{Nr} NW_i * 3XWeekly_i + \tau_w + \varepsilon_{iw} \quad (2)$$

Where SE_i is a dummy variable that takes the value 1 if respondent *i* was self-employed at baseline and zero otherwise, and NW_i a dummy variable indicating whether or not person *i* was not working at baseline. The omitted category is thus that of individuals who were wage employed at baseline.

Since labor market status is very stable over the course of the survey, baseline employment status is a good measure of the person's employment type during the course of the survey. The coefficients on the interaction terms enable us to test whether or not survey mode and reference period have a differential impact on the self-employed and those not-working, vis-a-vis the wage employed.

Survey mode has an especially strong impact on the reporting of 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 wage employed at baseline and assigned to the weekly phone treatment are on average different from those who were interviewed face-to-face once a week. By contrast, outcomes reported by those who were self-employed at baseline assigned to the weekly phone treatment arm 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 14 percentage point lower propensity to report work, 0.9 fewer days and 3.6 fewer hours worked, though not conditional on reporting any work. Put differently, the lower supply reports appear driven by the extensive margin. Those not working at baseline are an intermediate case; those who were assigned to the phone as opposed to the face to face treatment reported significantly fewer days and hours conditional on working, but not a lower aggregate labor supply and incomes: neither the reporting of work, unconditional days, unconditional hours nor incomes differs between those who were not working at baseline who were assigned to the face to face group and those assigned to the phone arms.

The length of the reference period impacts the reporting of all three groups, but not in uniform fashion. Among the wage-employed being assigned to the triweekly as opposed to one of the weekly treatment is associated with reporting fewer days and fewer hours conditional on reporting any. Among the self-employed, shortening the reference period from a week to two days is associated with a 16% higher incidence of reported work, 0.8 additional days and 9.5 additional hours worked. Average weekly incomes increasing by 59 CEDIS on average. Interestingly, days, hours and incomes conditional on reporting any work are not significantly impacted, suggesting the documented differential impact is due to the self-employed recording more spells when the reference period is shorter. Consistent with this finding, amongst those not working at baseline, those assigned to the triweekly treatment report more days and hours than those not working, but conditional on working their reports are not statistically different from those assigned to the weekly arms.

5. Reporting Differences: Retrospective and Repeated (High Frequency) Surveys[9]

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.[10]

Reporting differences are identified by estimating the following regression:

$$Y_{iw} = \beta_{F1} W1F2F_i + \beta_{P1} W1Phone_i + \beta_{P3} W3Phone_i + \beta_{CP} ControlPhone_i + c + \varepsilon_{iw} \quad (3)$$

Where Y_{iw} is a labor market report of interest, c is a constant, $W1F2F_i$ $W1Phone_i$, $W3Phone$, and $ControlPhone_i$ are dummy variables indicating individuals assigned to the weekly face-to-face, weekly phone, triweekly phone and control with phone arms respectively. Of primary interest is the coefficient β_{F1}, which reflects the difference in labor market reports between repeated weekly surveys conducted face-to-face and a one-shot endline survey enquiring about labor market behavior over the past three months. Since these interviews were all conducted face-to-face, the differences in labor market reporting can be ascribed to differences in reference period only. The

[9] 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.

[10] 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).

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. The coefficient β_{CP} measures the impact of receiving a phone on labor market reporting.

Table 7 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), the number of days worked per week on average (column 3) and conditional on working (column 4), the number of hours worked per week on average (column 5) and conditional on working (column 6) and log average income per week (column 7) and log average weekly income conditional on working (column 8). 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 form a week to two days, save for the reporting of income conditional on working. Relative to pure controls, i.e. controls who did not receive a phone, respondents in the weekly face-to-face treatment arm were 11.5% more likely to report having worked over the over the survey period, reporting 0.8 additional weeks worked. They do not report working fewer days (column 3) or hours (column 5) overall. However, conditional on reporting any work, they tend to work approximately half a day (column 5) and approximately 6 hours less (column 7) per week than the pure controls. Their reported earnings are not significantly different from those of the pure controls (column 7), but conditional on reporting any work they record earning 39 fewer CEDIS on average. 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. 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 weeks, days and hours worked (both unconditionally and conditional on working), albeit the differences with pure controls are only significant at the 10% level save for unconditional

hours. Consistent with the results described in the previous sections, individuals in the high frequency phone arms report systematically fewer days and hours than the two control groups in spite of a higher propensity to report any work at all. Shorter reference periods are thus associated with reporting of a higher incidence but lower average duration of employment spells.

6. 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 repeated high frequency interviews 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:

$$\Delta\overline{\mathrm{Y}}_i = \mathrm{Y}_{iEndline} - \overline{\mathrm{Y}}_{iHF} = \rho_{F1} W1F2F_i + \rho_{P1} W1Phone_i + \rho_{P3} W3Phone_i + v_i \qquad (4)$$

Where $\mathrm{Y}_{iEndline}$ is a labor market variable reported during the endline survey, and $\overline{\mathrm{Y}}_{iHF}$ the corresponding variable is obtained by aggregating high frequency labor market reports. $W1F2F_i$, $W1Phone_i$ are $W3Phone_i$ dummy variables that take value 1 if individual are assigned to the weekly face to face, weekly phone and triweekly phone arms respectively; the model we use is thus full saturated. If $\Delta\overline{\mathrm{Y}}_i > 0$ 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 coefficient ρ_{F1} is of particular interest since it provides an estimate of the misreporting of associated with a change in reference period alone; since both the face-to-face weekly high-frequency interviews and the endline interviews were conducted in person a change in modality cannot be the cause of differences in labor market reporting for subjects in this group.

Note that this analysis likely underestimates recall bias since, as is demonstrated in the Appendix Table 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 high frequency 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.

The results presented in Table 8 are suggestive of significant and substantial recall bias. At endline, employment (column 1), the number of weeks worked (column 2 and 3), and the number of hours worked per week conditional on doing any work (column 7), are significantly under-reported relative to high frequency reports. This is evidenced by the significantly negative constant in these four regressions, which reflects the average recall bias of the omitted category, those who participated in the weekly face-to-face interviews. By contrast, days (columns 3 and 4) are over-reported, as are overall earnings (column 9). The bias is economically meaningful. For example, using labor market indicators derived from repeated weekly observations as opposed to a one-shot retrospective report 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.

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. .

7. Conclusions

Labor market data are central to our understanding of how economies function and evolve, but they are 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 survey experiment.

The experiment demonstrates that shortening the reference period over which labor market questions are asked results in statistically significant and economically meaningful differences in labor market reports, particularly for the self-employed. 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. At the same time, hours and days worked, conditional on doing any work in a given week, 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 substantially 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 the incidence of employment, but over-report the number of days worked, when asked to recall the same statistics for an average week over a longer recall period. 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. Though establishing why these differences across employment categories arise is beyond the scope of this paper, a possible explanation, and one which resonates with the literature on survey modality effects, 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. Regardless, these findings underscore the need to be cognizant of modality induced labor market reporting differences in labor market surveys.

The key implication of the results presented in this paper, however, is that using recall periods of 24 to 48 hours, rather than the standard one week recall period, currently recommended by the ILO, may substantially improve the accuracy of labor statistics. While the paper shows this for a developing country urban labor market, the results seem relevant for both developing and developed countries. In the US, for example, the structure of the labor market has changed considerably in recent years, with far more workers engaged in some form of temporary or own account work, and permanent wage jobs accounting for a falling share of earnings and employment.

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. Dabalen, 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 FaceSurveys” 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.

Hussmanns, 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 Provide: 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*

Figure 1

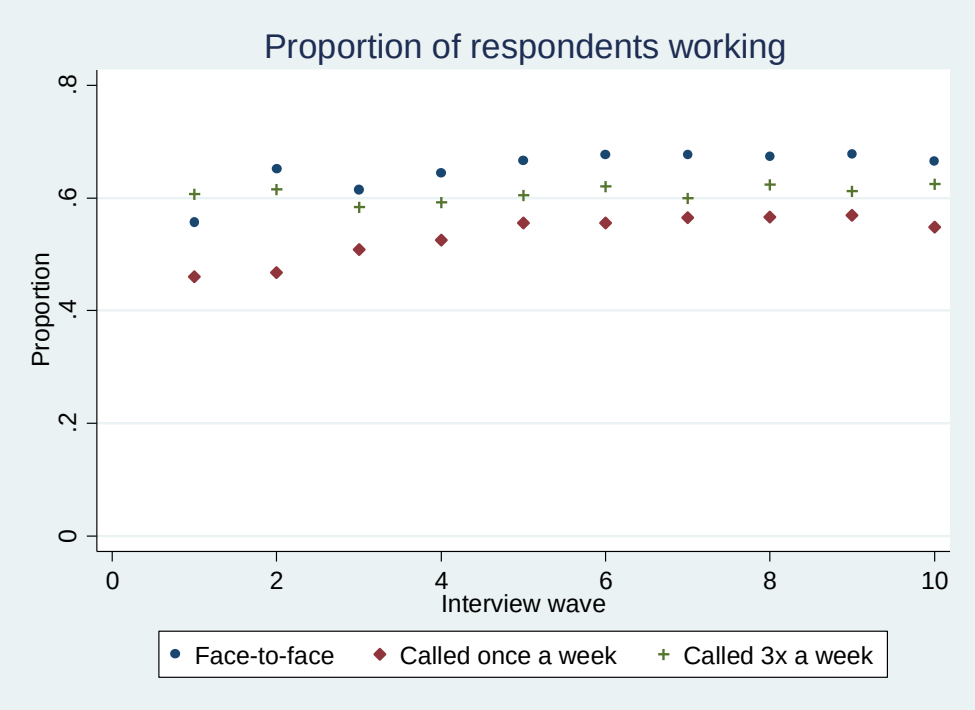

Proportion of respondents working
Proportion
0
.2
.4
.6
.8
0
2
4
6
8
10
Interview wave
Face-to-face
Called once a week
Called 3x a week

Table 1: Experiment Design and Survey Rounds by Treatment Arm

	Aug-Sept 2013	Aug-Oct 2013		Oct-Nov 2013	Jan-Feb 2014
	Baseline	High frequency interviews		Endline	3 Month Follow up
		weekly	3x weekly		
Weekly Face-to-face	✓	✓		✓	✓
Called 1x week	✓	✓		✓	✓
Called triweekly	✓		✓	✓	✓
Control without phone	✓			✓	✓
Control with phone	✓			✓	✓

Participation Fees (per completed interview) in Ghanaian CEDIS

	Received a phone	Weekly phone credit	Baseline	High frequency interviews		Endline	3 Month Follow up	Total payment (assuming perfect compliance)
				weekly	3x weekly			
Weekly Face-to-face			3	3		3	4	40
Called weekly	✓	1	3	3		3	4	50
Called triweekly	✓	1	3		2	3	4	80
Control without phone	✓		3			3	4	10
Control with phone	✓		3			3	4	10

Table 2 : Comparison of G-HFLS baseline with the Ghana Living Standards Survey 6 (GLSS6)

		G-HFLS	GLSS 6	Difference	*p* -value
		(1)	(2)	(3)	(4)
Individual Characteristics					
Male		0.410	0.454	-0.044	0.871
Age		35.100	35.465	-0.365	0.949
Education	Female	9.550	10.020	-0.470	0.929
	Male	11.160	12.890	-1.730	0.780
Employed	Female	0.640	0.768	-0.128	0.662
	Male	0.740	0.834	-0.094	0.752
of which Self-employed	Female	0.688	0.732	-0.044	0.908
	Male	0.480	0.414	0.067	0.890
Unemployed		0.070	0.030	0.040	0.681
Not in the labor force		0.250	0.170	0.080	0.670
Sample size		(N = 1579)	(N=13204)		
Household Characteristics					
Household size		4.370	3.640	0.730	0.734
Dependency ratio		0.195	0.310	-0.114	0.649
Maximum education level in the household		11.510	14.370	-3.110	0.650
Female headed households		0.350	0.340	0.009	0.984
Labor force participation rate (20 -60 year olds)		0.870	0.750	0.120	0.600
Asset index		0.000	0.100	-0.100	0.939
Sample size		N = 575	N = 6947		

Note: The GLSS 6 sample is restricted to urban households. For individual characteristics both samples are confined to respondents aged 20 - 60.

Table 3: Baseline Differences by Treatment Status

	No Phone Control	Treatment	Face-to-face Weekly	Phone Weekly	Phone Triweekly	Phone Control
	(1)	(2)	(3)	(4)	5.00	(6)
Household level variables						
Household size	5.345	-0.034	0.228	-0.143	-0.18	0.141
		-0.389	-0.501	-0.425	-0.46	-0.470
Dependency ratio	0.212	0.005	0.023	0.006	-0.01	0.008
		-0.023	-0.029	-0.029	-0.03	-0.029
Number of adults employed	2.504	-0.166	-0.104	-0.275	-0.12	-0.018
		-0.205	-0.262	-0.226	-0.26	-0.263
Highest Years of Schooling (HH)	12.363	0.121	0.110	0.041	0.21	0.394
		-0.298	-0.354	-0.364	-0.38	-0.352
Asset index	-0.191	0.201	0.300	0.031	0.27	0.349
		0.198	0.233	0.253	0.24	0.222
N		444	223	222	225.00	220
Individual Variables						
Male	0.399	0.01	0.026	0.003	0.001	0.025
		-0.032	-0.036	-0.04	-0.04	-0.04
Married	0.403	0.003	0.04	-0.009	-0.022	-0.022
		-0.038	-0.05	-0.046	-0.046	-0.047
Age	35.826	0.685	1.096	0.051	0.886	0.149
		-0.718	-0.85	-0.85	-0.906	-0.804
Years of education	9.798	-0.22	-0.45	-0.089	-0.12	-0.113
		-0.296	-0.379	-0.36	-0.364	-0.333
Employed	0.594	-0.01	0.04	-0.026	-0.044	-0.011
		-0.035	-0.044	-0.045	-0.042	-0.044
Self employed	0.282	0.04	0.083**	0.018	0.019	0.014
		-0.032	-0.04	-0.038	-0.039	-0.037
Log income per week (total)	2.478	-0.051	0.135	-0.101	-0.186	-0.019
		-0.152	-0.194	-0.19	-0.188	-0.191
Hours worked per week (total)	31.470	-1.931	1.274	-3.725	-3.351	-1.133
		-2.13	-2.754	-2.576	-2.559	-2.597
N		1211	603	601	610.00	605
p-value for F-test for joint		0.670	0.480	0.750	0.74	0.980

Note: *** p<0.01, ** p<0.05, * p<0.10. Income per week is winsorized at 5th and 95th percentiles. Hours worked per week is winsorized at 1st and 99th percentiles. Standard Errors are clustered at treatment level (Household).

Table 4: Non-Compliance: % missed interviews

	N	High Frequency Interviews	Endline (in person)	3-Month Followup (by phone)
	(1)	(2)	(3)	(4)
Weekly Face-to-face	318	6.4%	0.0%	9.4%
Called 1x week	315	11.2%**	0.3%	7.0%
Called triweekly	321	10.8%**	0.3%	8.4%
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 5: The Impact of Survey Mode and Reference Period on Labor Market Reporting in the High Frequency Survey

	Working	Days	Days-conditional on work	Hours	Hours-conditional on work	Income	Income (conditional on work)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Phone[(a)]	-0.106***	-0.684***	-0.201**	-7.742***	-4.796**	-6.285	13.533
	(0.036)	(0.206)	(0.086)	(2.162)	(2.350)	(8.604)	(19.343)
Triweekly[(b)]	0.090**	0.096	-0.602***	1.423	-3.987**	27.259***	43.411*
	(0.036)	(0.190)	(0.118)	(1.890)	(1.907)	(9.820)	(24.145)
Control Mean	0.650	3.499	5.378	31.424	48.345	47.469	80.718
Triweekly Phone vs F2F (a+b)	0.637	0.004	0.000	0.001	0.000	0.029	0.013
N. Observations	7,814	7,814	4,700	7,814	4,693	6,616	3,492
Adjusted R2	0.014	0.020	0.061	0.020	0.025	0.009	0.019

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors clustered by household in parentheses, the ommitted category in Panel A is face-to-face weekly interviews. The ommitted category in panel B is those who were wage employed at baseline assigned to weekly face-to-face interviews.

Table 6: Heterogeneity in Impact by Baseline Employment Status

	Working	Days	Days-conditional on work	Hours	Hours-conditional on work	Income	Income-conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Phone	-0.015	-0.110	-0.024	-2.657	-2.321	6.953	-11.004
	(0.042)	(0.274)	(0.145)	(3.452)	(3.300)	(23.964)	(29.511)
Tri-Weekly	-0.006	-0.484*	-0.562***	-4.799	-5.652**	-9.296	0.365
	(0.042)	(0.286)	(0.178)	(3.159)	(2.881)	(21.785)	(29.711)
Self-employed (SE)	0.035	0.280	0.114	0.529	-1.375	-11.094	-24.939
	(0.037)	(0.266)	(0.152)	(3.633)	(3.056)	(18.787)	(25.210)
Not working (NW)	-0.688***	-3.631***	0.143	-33.915***	-0.442	-71.572***	-30.870
	(0.049)	(0.297)	(0.161)	(3.512)	(4.006)	(18.296)	(35.730)
SE*Phone	-0.138***	-0.920**	-0.250	-8.932**	-3.657	-12.029	34.868
	(0.049)	(0.361)	(0.208)	(4.412)	(4.036)	(24.320)	(36.217)
SE*Tri-Weekly	0.155***	0.788**	0.012	9.447**	3.807	58.879**	47.646
	(0.056)	(0.344)	(0.226)	(4.082)	(4.074)	(24.873)	(39.556)
NW*Phone	-0.002	-0.049	-0.558*	-0.052	-12.844**	-11.393	-4.442
	(0.062)	(0.369)	(0.313)	(4.085)	(6.100)	(24.118)	(45.888)
NW*Tri-Weekly	0.068	0.656**	-0.236	6.488*	3.225	16.471	189.068
	(0.060)	(0.333)	(0.365)	(3.354)	(5.862)	(22.650)	(128.302)
Reference Period Effect S	0.149***	0.304	-0.550***	4.649	-1.846***	49.583***	48.011
p-value	0.000	0.231	0.001	0.137	0.002	0.003	0.125
Reference Period Effect N	-0.620	-2.975*	-0.093***	-27.428*	2.783**	-55.100	158.198
p-value	0.649	0.053	0.006	0.052	0.023	0.923	0.753
Mode Effect Self-Employ	-0.153***	-1.03***	-0.273**	-11.589***	-5.978*	-5.077	23.864
p-value	0.000	0.000	0.048	0.001	0.053	0.669	0.341
Mode Effect Not Workin	-0.017	-0.159*	-0.582***	-2.709*	-15.165**	-4.440	-15.446
p-value	0.649	0.053	0.006	0.052	0.023	0.923	0.753
Control Mean	0.833	4.419	5.297	41.003	49.232	76.726	100.152
N. Observations	7,814	7,814	4,700	7,814	4,693	6,616	3,492
Adjusted R2	0.420	0.378	0.067	0.297	0.038	0.079	0.024

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors clustered by household in parentheses, the ommitted category are those who were wage

Table 7: Reporting Differences Between High Frequency Reports and 3 Month Retrospective Recall Reports

	Any work	Weeks	Days	Days-conditional on work	Hours	Hours-conditional on work	Income	Income - conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Face to Face HF	0.115***	0.807*	0.110	-0.456***	-0.215	-5.994**	4.098	-39.127***
	(0.041)	(0.455)	(0.227)	(0.089)	(2.613)	(2.484)	(8.479)	(10.747)
Weekly Phone HF (WP)	0.071*	-0.580	-0.623***	-0.817***	-8.236***	-12.905***	-14.043*	-57.329***
	(0.043)	(0.459)	(0.235)	(0.108)	(2.477)	(1.923)	(8.161)	(10.143)
Tri-weekly Phone HF (TP)	0.106***	0.287	-0.627***	-1.456***	-8.081***	-17.381***	24.668**	-18.476*
	(0.040)	(0.460)	(0.220)	(0.113)	(2.309)	(1.950)	(9.630)	(11.016)
Control Phone (CP)	0.064	0.822*	0.482*	0.143*	6.595**	4.110*	12.693	8.084
	(0.041)	(0.476)	(0.248)	(0.080)	(2.839)	(2.228)	(7.992)	(10.596)
p-value (WP=CP)	0.847	0.001	0.000	0.000	0.000	0.000	0.000	0.000
p-value (TP = CP)	0.272	0.204	0.000	0.000	0.000	0.000	0.165	0.004
Control Mean	0.633	6.994	3.394	5.779	31.872	54.029	1,342	1,137
N. Observations	1,573	1,573	1,573	1,111	1,573	1,109	0.013	0.046
Adjusted R2	0.005	0.008	0.026	0.235	0.037	0.125	0.016	0.049

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors in parentheses, the ommitted category are controls without phones.

Table 8: Recall Bias (Difference between Restrospective Reports from the Endline and High Frequency Reports over three months)

	Any Work	Total weeks worked	Weeks worked conditional on any work	Days worked per week	Days worked per week conditional on any work	Hours	Hours worked per week conditional on any work	Weekly Income	Weekly Income-conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Face to Face HF (WF)	-0.022**	-0.403***	-0.432**	0.588***	0.860***	0.256	-4.590***	-12.967***	-23.681***
	(0.010)	(0.146)	(0.189)	(0.075)	(0.082)	(0.908)	(1.363)	(4.488)	(7.226)
Weekly Phone HF (WP)	-0.016	0.400**	0.510**	0.968***	1.356***	3.907***	-2.112	-4.506***	-9.859***
	(0.015)	(0.198)	(0.260)	(0.108)	(0.120)	(1.202)	(1.493)	(4.177)	(7.280)
Tri-weekly Phone HF (TP)	-0.003	-0.400***	-0.756***	1.246***	1.553***	2.630***	-1.130	-28.279***	-52.607***
	(0.017)	(0.146)	(0.181)	(0.095)	(0.118)	(0.853)	(1.246)	(4.904)	(9.139)
Modality effect (WP-WF)	0.006	0.803***	0.942***	0.381***	0.496***	3.651**	2.478	8.461	13.822
Frequency Effect (TP-WP)	0.013	-0.800	-1.266	0.277***	0.197***	-1.277*	0.982*	-23.773*	-42.749
N. Observations	947	947	663	947	662	947	660	797	496
Adjusted R2	0.005	0.023	0.044	0.288	0.430	0.029	0.026	0.503	0.100

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors in parentheses. The model is fully saturated.

Appendix Tables

Table : Characterisics of the Wage and Self-Employed at Baseline

	Wage Employed (1)	Self-Employed (2)	Difference (3)	*p*-value (4)
Male	0.57	0.36	0.21	0.00
Age	33.61	38.60	-4.99	0.00
Education	11.15	9.33	1.82	0.00
Days worked per week	5.61	5.74	-0.13	0.04
Hours worked per week	49.63	51.28	-1.65	0.17
log pay per week	4.46	4.20	0.26	0.00
Pay per week (Cedis)	145.14	121.41	23.73	0.04
Tenure	5.51	7.89	-2.38	0.00
Multiple Jobs	0.05	0.06	-0.01	0.58
Mobile phone ownership	0.98	0.97	-0.01	0.14
Public sector	0.09	0.00	0.09	0.00
Manufacturing	0.10	0.08	0.03	0.24
Trade	0.14	0.64	-0.50	0.00
Services	0.66	0.29	0.37	0.00
Contract	0.49			
Pension	0.40			
Paid sick leave	0.35			
Paid holidays	0.34			
Paid overtime	0.45			
Pays taxes/permit fees		0.57		
Business is Mobile		0.67		
Employer size<5	0.30			
5<=Employer size<20	0.33			
20<=Employer size<100	0.19			
Employer size>=100	0.15			
no employees		0.76		
1 employee		0.12		
2-5 employees		0.11		
5-10 employees		0.01		

Table : Compliance During the High Frequency Survey

	Weekly Face to Face N=318	Weekly Phone N=315	Triweekly Phone N=321	
Week			Interview	
1	98.70%	100.00%	1	99.70%
			2	80.10%
			3	83.20%
2	76.70%	73.30%	4	84.70%
			5	90.30%
			6	84.40%
3	94.70%	78.70%	7	86.60%
			8	88.20%
			9	91.90%
4	96.20%	89.50%	10	92.20%
			11	90.00%
			12	89.70%
5	96.20%	88.60%	13	92.20%
			14	91.30%
			15	86.30%
6	92.50%	88.60%	16	91.60%
			17	92.80%
			18	91.90%
7	96.20%	92.70%	19	91.00%
			20	88.50%
			21	87.50%
8	93.40%	91.40%	22	87.50%
			23	87.90%
			24	92.20%
9	96.50%	93.70%	25	88.50%
			26	92.50%
			27	91.30%
10	95.00%	91.40%	28	86.00%
			30	87.20%
Average	93.60%	88.80%		89.20%

Table X: Subjective Survey Satisfaction

Treatment	Uncomfortable Questions?	Privacy Violated?	Willing to participate without compensation?	Compensation Sufficed?
Control - No Phone	11.30%	22.20%	54.00%	35.40%
Control - Phone	9.60%	12.7%***	64.0%**	43.00%
Face to Face - Weekly	10.10%	20.80%	51.70%	53.0%***
Called - 1X Weekly	9.90%	19.40%	52.50%	45.9%*
Called - 3X Weekly	9.10%	17.80%	59.10%	65.3%***

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

5% Verification Check		
Congruent Responses	N	Average
Did you receive a phone at the time of the baseline interview?	113	98%
Was it a new phone (packed in a box)?	70	97%
How often were you interviewed per week?	84	90%
Employment status at baseline	114	93%
Occupation at baseline	86	80%

Endline Reports

	Any work	Weeks	Days	Days-conditional on work	Hours	Hours-conditional on work	Log Income	Log Income-conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Face to Face HF	0.090**	0.359	0.387	-0.181**	-0.237	-6.766***	0.196	-0.240**
	(0.043)	(0.474)	(0.255)	(0.088)	(2.797)	(2.488)	(0.188)	(0.096)
Weekly Phone HF	0.058	-0.159	0.069	-0.355***	-4.254	-10.167***	0.006	-0.224***
	(0.042)	(0.472)	(0.261)	(0.094)	(2.766)	(2.316)	(0.192)	(0.083)
Tri weekly phone HF	0.104***	-0.103	0.340	-0.330***	-5.498**	-14.403***	0.157	-0.287***
	(0.039)	(0.467)	(0.248)	(0.089)	(2.465)	(2.245)	(0.173)	(0.091)
Control Phone	0.064	0.822*	0.451*	0.143*	6.595**	4.839*	0.355*	0.105
	(0.039)	(0.485)	(0.261)	(0.077)	(3.011)	(2.490)	(0.199)	(0.077)
Control Mean	0.633	3.994	3.679	5.779	31.872	50.315	2.789	4.403
N. Observations	1,577	1,577	1,577	1,101	1,577	1,098	1,577	1,083
Adjusted R2	0.004	0.002	0.002	0.039	0.019	0.078	0.001	0.028

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors in parentheses, the ommitted category are controls withou phoness.

	Follow up Reports							
	Any work	Weeks	Days	Days-conditional on work	Hours	Hours-conditional on work	Log Income	Log Income-conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Face to Face HF	0.054	0.733	0.246	-0.223	3.834	0.110	0.350*	-0.039
	(0.043)	(0.507)	(0.257)	(0.182)	(2.841)	(2.878)	(0.180)	(0.121)
Weekly Phone HF	0.004	0.248	0.074	0.069	0.770	0.782	0.057	0.025
	(0.044)	(0.483)	(0.231)	(0.156)	(2.471)	(2.523)	(0.185)	(0.114)
Tri weekly phone HF	0.028	0.611	0.299	-0.064	4.491*	1.853	0.212	-0.131
	(0.043)	(0.464)	(0.240)	(0.180)	(2.483)	(2.671)	(0.177)	(0.120)
Control Phone	0.006	0.355	0.152	0.043	3.162	3.024	0.093	0.086
	(0.041)	(0.531)	(0.252)	(0.164)	(2.740)	(2.487)	(0.182)	(0.112)
Control Mean	0.671	7.004	3.321	5.139	30.521	47.231	2.296	4.104
N. Observations	1,423	1,427	1,427	973	1,427	973	1,577	941
Adjusted R2	-0.001	-0.001	-0.001	0.000	0.001	-0.002	0.001	0.000

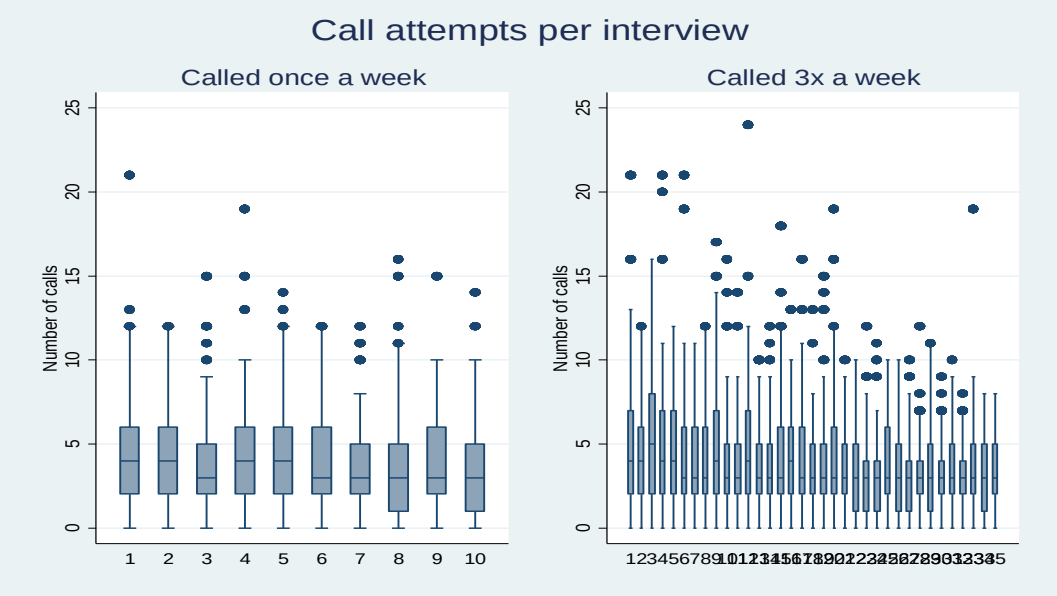
Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors in parentheses, the ommitted category are controls without phones.

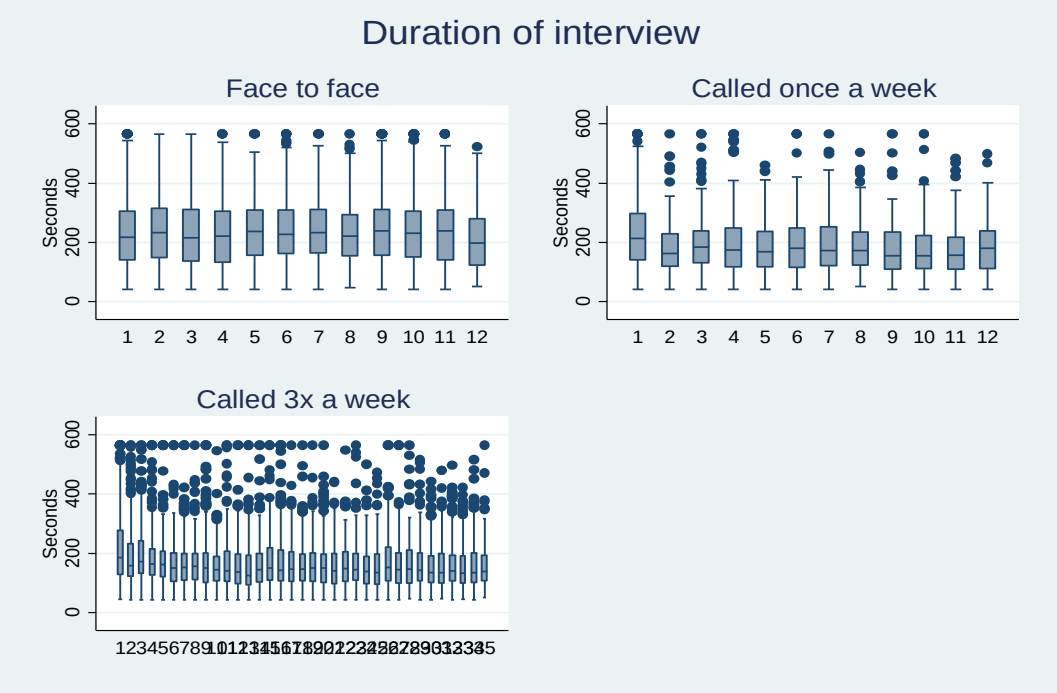
Table : Reporting Differences High Frequency Reports vs 3 Month Retrospective Recall Reports - Heterogeneity

	Any work	Weeks	Days	Days-conditional on work	Hours	Hours-conditional on work	Log Income	Log Income-conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Face to Face HF	0.028	-0.552	-0.696**	-0.547***	-7.645**	-6.294**	-0.169	-0.316
	(0.037)	(0.518)	(0.291)	(0.148)	(3.838)	(3.135)	(0.325)	(0.209)
Weekly Phone HF	0.027	-0.917*	-0.891***	-0.608***	-10.990***	-8.612***	-0.111	-0.236
	(0.036)	(0.528)	(0.275)	(0.119)	(3.566)	(2.619)	(0.310)	(0.169)
Tri-weekly Phone HF	0.007	-1.124*	-1.571***	-1.464***	-18.226***	-17.671***	-0.346	-0.337*
	(0.041)	(0.612)	(0.293)	(0.170)	(3.403)	(2.710)	(0.312)	(0.194)
Control Phone (CP)	0.004	0.446	0.232	0.020	3.647	1.997	0.181	0.056
	(0.044)	(0.548)	(0.287)	(0.113)	(4.332)	(3.470)	(0.324)	(0.150)
Self-employed (SE)	-0.054	-0.885*	-0.440*	-0.043	-5.498	-1.444	0.164	0.228
	(0.037)	(0.502)	(0.267)	(0.129)	(3.613)	(2.937)	(0.262)	(0.140)
Not working (NW)	-0.802***	-9.187***	-4.453***	-0.093	-43.081***	-5.549	-3.289***	-0.143
	(0.047)	(0.530)	(0.268)	(0.322)	(3.419)	(7.279)	(0.269)	(0.260)
SE*Face to Face HF	0.067	1.322**	0.731*	0.119	6.011	0.119	0.323	0.208
	(0.045)	(0.651)	(0.377)	(0.196)	(4.791)	(4.131)	(0.368)	(0.240)
SE*Weekly Phone HF	0.031	-0.146	-0.090	-0.220	-1.908	-5.543	-0.206	-0.048
	(0.049)	(0.694)	(0.376)	(0.193)	(4.839)	(3.986)	(0.373)	(0.214)
SE*Tri-weekly Phone HF	0.063	1.429*	0.572	-0.035	6.547	0.407	0.768**	0.632***
	(0.050)	(0.757)	(0.355)	(0.213)	(4.729)	(3.923)	(0.378)	(0.244)
SE*Control Phone	0.088*	0.592	0.477	0.223	6.903	3.725	0.235	-0.038
	(0.049)	(0.663)	(0.356)	(0.165)	(5.342)	(4.640)	(0.352)	(0.190)
NW*Face to Face HF	0.051	1.028	0.864**	0.288	9.272*	5.479	0.139	0.302
	(0.068)	(0.786)	(0.390)	(0.355)	(4.842)	(8.123)	(0.368)	(0.361)
NW*Weekly Phone HF	0.084	0.960	0.794**	-0.870*	9.130**	-12.787	-0.033	-0.076
	(0.069)	(0.726)	(0.359)	(0.468)	(4.148)	(8.220)	(0.354)	(0.361)
NW*Tri-weekly Phone HF	0.238***	1.792**	1.553***	-1.088**	17.199***	-8.645	0.514	0.138
	(0.069)	(0.774)	(0.366)	(0.426)	(4.039)	(7.878)	(0.370)	(0.403)
NW*Control Phone (CP)	-0.078	-1.347*	-0.692*	-0.335	-7.967	-9.204	-0.480	0.586
	(0.069)	(0.766)	(0.384)	(0.440)	(4.984)	(8.782)	(0.399)	(0.357)
Control Mean	0.932	10.511	5.103	5.807	48.866	55.157	3.792	4.388
N. Observations	1,575	1,575	1,575	1,119	1,575	1,117	1,357	886
Adjusted R2	0.590	0.594	0.555	0.320	0.425	0.164	0.608	0.057

Note: *** p<0.01, ** p<0.05, * p<0.10, bootstrapped standard errors in parentheses, the ommitted category in Panel A is controls without phones. The ommitted category in panel B is wage employed controls without phones.

Appendix A: Triangulation with Call Log Data





Appendix B: 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 B1: Survey Satisfaction

Table X: Subjective Survey Satisfaction

Treatment	*Uncomfortable Questions?*	*Privacy Violated?*	*Willing to participate without compensation?*	*Compensation Sufficed?*	*Willing to participate in 3 months?*
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

Appendix C. 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 C1: 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 ii) income after costs and including earnings on account.