Measuring Employment: Experimental Evidence from Ghana

May 2017

Rachel Heath, Ghazala Mansuri, Bob Rijkers, William Seitz, and Dhiraj Sharma

Abstract: A randomized survey experiment in Ghana demonstrates how the length of the reference period and interview modality affect how people respond in labor surveys. When the reference period is shorter than the traditional one week survey participants report more self-employment spells, but not more wage-employment spells. Use of shorter reference periods also reduces the reported duration of both self- and wage-employment spells. Finally, phone interviews yield lower estimates of employment, hours worked, and days worked among the self-employed as compared to in-person interviews. The results imply that labor surveys would benefit from the use of reference periods shorter than one week.

Keywords: Labor statistics, labor force surveys, reference period, interview mode, self-employment; survey design, phone-based surveys

JEL Codes : J20, J21, J22, J46

Rachel Heath, email: rmheath@uw.edu (Assistant Professor, University of Washington); Ghazala Mansuri:gmansuri@worldbank.org (Lead Economist, World Bank); Bob Rijkers, email: brijkers@worldbank.org (Economist, World Bank); William Hutchins Seitz, email: wseitz@worldbank.org (Economist, World Bank) and Dhiraj Sharma, email: dsharma5@worldbank.org (Economist, World Bank).

The authors thank Erhan Artuc, Kathleen Beegle, Andrew Dabalen, Adriana Kugler, Vasco Molini and conference participants at the 2016 Annual Bank Conference on Development Economics, the World Bank, and Oxford University for helpful comments. Paolo Falco and Mary Hallward-Driemeier provided support in the early stages of the project. Arthur Lagrange, Jonathan Lain, Poorna Mazumdar, Lokendra Phadera, Mohammed Y. H. Saleh, Vaclav Tehle, and Asmus Zoch provided excellent research assistance. The research was supported by funding from the World Bank's research support budget. Views expressed in this paper do not necessarily represent the views of the World Bank Group, its affiliated organizations, the executive directors of the World Bank Group or the governments they represent.

1. Introduction

The nature of employment is changing rapidly. Stable full-time wage employment is declining in advanced economies, and the growth of salaried employment is slowing in developing economies, where informal and self-employment are widespread (ILO 2015). The increasing importance of own-account, temporary, and sporadic jobs—often referred to as the "gig economy"—has rejuvenated interest in whether current labor market data collection methods are appropriate.

Current practice favors the use of a one-week reference period in labor market surveys in both developed and developing countries, even though a shorter reference period, such as a day, might provide a sharper snapshot of labor market activity when such activity is short-lived and transitory (Hussmans, Mehran, and Verma 1992). At the same time, with the penetration of mobile phones in developing countries, new modalities of collecting data have become feasible.

This paper speaks to both these issues. It reports the findings of a randomized survey experiment conducted in urban Ghana to assess the impact of the reference period and survey modality on labor statistics. The experiment tracked the labor market behavior of 1,579 respondents of the Ghana Urban Household Panel Survey for six months, using four monitoring instruments: (a) a baseline in-person interview, (b)a high-frequency sequence of interviews spanning 10 consecutive weeks, (c) an endline in-person interview conducted approximately three months after the baseline, and (d) a follow-up phone survey conducted three months after completion of the endline interview (i.e., six months after the start of the survey). Survey respondents were randomized into 5 'treatment' arms for the high frequency interviews which varied both the reference period and survey modality. One fifth of respondents were interviewed three times a week (using two separate 24-hour recall questions) by phone, two fifths were allocated to a weekly interview, but faced a distinct survey mode. In one arm, respondents were interviewed in person, while in the other, they were interviewed over the phone. Finally, two fifths were held as control. Phones were provided to all respondents interviewed on the phone, as well as to all respondents in one of the control arms.

The results indicate significant and large reference period and survey mode effects. Participants report significantly more self- but not wage-employment spells when the reference period is shorter than a week: self-employed workers in the tri-weekly arm are 15 percentage points more likely to report having done any work over the course of a given week than self-

employed individuals who were asked to recall their labor market behavior over the previous week. This finding suggests that the reference period of one week (the norm in most labor force surveys) systematically underestimates the employment rate of self-employed workers. Conditional on reporting working, both the self-employed and wage employees also report fewer days and hours, registering 0.6 fewer days per week worked and 5.6 and 1.8 fewer hours per week worked, respectively. Qualitatively similar patterns are observed when comparing retrospective labor market reporting over a three-month period. In this case, the quarterly report of the treated groups, based on aggregating their high-frequency interviews, is compared against the endline report of control groups that did not take part in the high-frequency survey. These results underscore the need to reconsider using a week as the standard reference period in labor market surveys, especially where self-employment is widespread.

The paper also finds interesting survey mode effects. Weekly phone-based surveys yield lower estimates of employment, as well as hours and days worked, as compared to weekly in-person interviews and these effects are again concentrated among the self-employed. This may reflect the presence of social desirability bias in in-person interviews for the self-employed, resulting in higher estimates of their labor input.

The rest of the paper is organized as follows. Section 2 reviews the literature on recall bias and interview mode in labor surveys and elaborates on the paper's hypotheses. Section 3 describes the experiment. Section 4 presents the data. Section 5 presents the empirical strategy and main results. The last section summarizes the paper's main conclusions. Appendix A defines the variables used in the paper. Appendix B provides statistics on the survey design, sample, compliance, and data quality checks. Appendix C elaborates on the aggregation of triweekly interviews to weekly labor reports. Appendix D assesses recall.

2. Recall Bias and Interview Mode in Labor Surveys

A substantial body of literature shows that retrospective reports are prone to recall bias, [1] the magnitude of which depends on factors such as the salience of the events to be recalled, social

[1] Summarizing the literature on recall error in labor market reporting is beyond the scope of this paper (see Bound, Brown, and Mathiowetz 2001 and Beckett et al. 2001 for summaries). A number of studies are particularly relevant to the current study, however. Comparing company records of employment with reports from the Panel Study of Income Dynamics, Mathiowetz and Duncan (1998) find that response errors regarding employment are lowest in the

desirability, respondent characteristics (Bardasi et al. 2011), and the reference period.[2] Longer reference periods (e.g., reporting labor market behavior over the course of a year) have been associated with increased recall bias that leads to differences in reported labor outcomes. Studies that use time diaries, in which the reference period is typically very short, yield lower provides a snapshot of the economy at a given point in time. In the presence of seasonality and/or other substantial economic fluctuations, a one-shot measure may not be a good representation of labor market activity. In such cases, accuracy can be improved by taking repeated measures, staggering the survey, or using longer reference periods (i.e., measuring the usually active population). These measures tend to yield different results not only because they are conceptually different but also because they are less subject to recall error. Measures of usual employment may also fail to pick up sporadic employment episodes by people who are typically not working. One would thus expect repeated surveys of "current" activity to yield higher estimates of employment but not necessarily higher estimates of days and hours worked.

---

months of and before the interview and when spells are characterized by either full employment or complete unemployment (see also Horvath 1982). Subjects appear to have more difficulty recalling spells of short duration. Pierret (2001) shows that switching from annual to biennial interviews in the National Survey of Youth 1979 results in the reporting of both fewer employers and fewer spells of nonemployment, suggesting that short spells of employment or nonemployment are easily forgotten. estimates of both total labor supply (Duncan and Stafford 1980; Hamermesh 1990; Robinson and Bostrom 1994; Bonke 2005; Robinson et al. 2011; Juster and Stafford, 1991), responsiveness of labor supply to wages (Barrett and Hamermesh 2016) and fertility (Carlin and Flood 1997) than standard labor surveys.

[2] The choice of reference period in labor market surveys has been the subject of extensive debate (Hussmans, Mehran, and Verma 1990; Stewart 2014). The International Labour Organization (ILO) identifies both a day and a week as appropriate reference periods, as they correspond closely to an instantaneous (stock) measure of employment and are less vulnerable to the memory-dependent errors that arise over longer periods of recall.

Most labor market surveys use a reference period of one week, because of both the practicality of measurement and consistency with other sources. When full-time formal-sector employment is the norm, using weekly as opposed to daily recall has the additional advantage of resulting in a lower variance while giving similar average results. The ILO notes that when intermittent, casual, and short-term employment is widespread, as is the case in developing countries, shorter reference periods may enhance accuracy. If self-employment is more volatile than wage employment, one would expect the choice of the reference period to have greater impact on reporting of self-employment than wage-employment.

Another strand of the literature examines the impact of survey mode in social surveys. Phone interviews may suffer higher rates of nonresponse (De Leeuw 1992; Holbrook, Green, and Krosnick 2003) and yield slightly different responses than in-person interviews (Groves 1990) because they tend to be less susceptible to social desirability bias (De Leeuw 2005). While the literature on phone surveys and social desirability bias in developing countries has tended to focus on issues such as stigmatized health behaviors (Gregson et al. 2002; Langhaug, Sherr and Cowan 2010), social desirability bias could also prompt respondents to inflate their employment reports in labor surveys. If it does, phone interviews could be associated with lower estimates of days and hours worked, especially in jobs where the distinction between professional and personal tasks is blurry.

3. Experiment Design

The experiment was designed to examine the impact of the length of the reference period and survey modality on labor market reporting. It included three treatments and two controls arms (table 1). Participants in all five groups were interviewed in-person at baseline and endline and by phone in a three-month follow-up. The treatment groups comprised (a) one group that was interviewed by phone three times a week for 10 weeks (30 interviews),[3] (b) another that was interviewed by phone once a week for 10 weeks (10 interviews), and (c) a third that was interviewed in person once a week for 10 weeks (10 interviews). The questionnaire was identical for both phone and in-person interviews, and the two types of interviews were conducted at the same times. All participants interviewed by phone received cellphones to avoid selection bias associated with phone ownership.[4]

[3] All but a handful of the triweekly interviews were conducted on the same days of the week (Tuesday, Thursday, and Saturday). Respondents were asked to recall their labor market outcomes in over the previous 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 Saturdays. Appendix D explains how triweekly interviews were aggregated to render them comparable to weekly labor reports.

[4] People who already owned a SIM card were given the option of receiving phone calls on their existing number (to avoid forcing them to adopt a new number specifically for the survey).

It is possible that the act of participating in the survey itself and/or the receipt of a phone affects the behavior and/or reporting of respondents.[5] This concern is limited by the fact that 95.7 percent of survey participants owned a cellphone before the experiment began. To nonetheless control for the possible effect of the cellphone, respondents were randomly assigned to two different control arms, one that received a cellphone at baseline and one that did not. The difference in reporting of the two control groups allows assessment of the impact of receiving a phone on labor market reporting.

To incentivize participation, all respondents were paid.[6] Participants in both the treatment and control arms received 3 Cedis ($1.36) for completing the baseline survey, 3 Cedis ($1.36) for completing the endline survey, and 4 Cedis ($1.82) worth of airtime credit for completing the three-month follow-up. In addition, participants who were interviewed once a week (by phone or in-person) received 3 Cedis ($1.36) for each completed interview, and participants who were interviewed three times a week received 2 Cedis ($0.91) for each completed interview. All payments, except the follow-up survey payment, were made at endline. Individuals assigned to the weekly phone, triweekly phone, and control with phone arms received phones (whether or not they already owned one) with a SIM card and 1 Cedi ($0.46) of phone credit.

The sample was drawn from the Ghana Urban Household Panel Survey (GUHPS).[7] A baseline in-person interview was conducted with all respondents before the high-frequency

[5] Mobile phone ownership and usage have been associated with increased job search (Tack and Aker 2014); the imparting of basic skills (Aker, Ksoll, and Lybbert 2012); and increased migration (Aker, Clemens, and Ksoll 2011). Access to information technology has also been associated with reductions in price dispersion (see, e.g., Jensen 2007 and Goyal 2010) and increased risk sharing (Jack and Suri 2014). See Aker and Mbiti (2010) for an overview of the literature.

[6] The total costs per respondent were $82.61 for weekly in-person interviews, $89.48 for weekly phone interviews, and $109.29 for triweekly phone interviews. The in-person interview arm was less expensive than the phone arms because participants did not receive a cellphone or weekly phone credit top-ups. However, in-person visits required enumerators to visit households and entailed higher enumerator and supervisor wages as well as higher transportation costs. Had the survey duration been extended to 16 weeks, the cost of weekly phone interviews would have fallen below the cost of in-person interviews in terms of both the cost per respondent and the cost per interview.

[7] The Ghana Urban Household Panel Survey (GUHPS) is a panel labor market survey administered by the Centre for the Study of African Economies at the University of Oxford. The sampling frame for the experiment consisted of

survey was initiated. It served multiple functions, including collecting baseline information on key variables of interest, collecting contact information, distributing cellphones to participants, familiarizing respondents with the survey questions, and cultivating trust between enumerators and respondents. All respondents provided their phone numbers (often more than one) and indicated their preferred phone number for completing phone interviews.

Subjects were allocated to one of the five arms. The arms were balanced on a range of observable characteristics, including gender, education, age, occupation, marital status, dependency ratio, asset ownership, and mobile phone ownership. Randomization was applied at the household level, so that everyone in a household was assigned to the same arm, in order to avoid intrahousehold spillovers arising from assigning members from the same household to different arms.

A number of strategies helped reduce attrition and enhance the quality of the data. The in-person baseline interview was key to this process. At baseline, enumerators were paired with specific respondents, of the same gender, for the entirety of the survey. To convey interest in individuals' welfare, interviewers asked questions about well-being before inquiring about labor market outcomes. The high frequency interviews were also short. They were designed to take no more than five minutes. This was assisted by the use of Computer Assisted Personal Interviewing (CAPI) software that prepopulated time-invariant information, allowing interviewers to focus on questions that were expected to vary over the course of the survey. A log was kept of all calls made from the phones used for the survey. This allowed for a verification of the date, time, and duration of calls and also maintained a record of call attempts, network problems, and other usage statistics. Information entered by enumerators in the hand-held device was also verifiable against data from the call logs automatically stored in the phone assigned to each enumerator, enabling better monitoring. During the baseline interview, respondents were also asked to indicate when they preferred to be interviewed. Finally, data quality checks were performed throughout the survey, (see Caeyers, Chalmers, and De Weerdt 2012 for evidence that CAPI systems help improve data quality).

---

respondents who had been interviewed in the GUHPS, excluding individuals under the age of 20 or above the age of 60 in 2013, individuals not contacted in either 2010 or 2012, and individuals located in Takoradi-Secondi (to cut costs, the number of locations in which the survey experiment was conducted was limited). The resulting sample frame consisted of 2,251 individuals from 720 households.

This approach was successful: The average phone interview took less than four minutes.[8] More than 95 percent of the interviews were matched with a record in the call logs; unmatched interviews largely reflected the fact that respondents asked to use a number that was not provided at the time of the baseline or endline interview. After completion of the endline survey, 5 percent of respondents were randomly selected for a verification survey. They were asked to verify whether they had received a phone, whether the phone was sealed in a box, how often they were interviewed, what their employment status at baseline was, and what economic activity they were engaged in at the time of the baseline interview. Their responses are highly consistent with their baseline responses and attest to the credibility of the collected data (see appendix table B.2).

While phone interviews were associated with somewhat lower compliance than in-person interviews (consistent with earlier studies such as De Leeuw, 1992), survey compliance was very high overall: Only 6.4 percent of weekly in-person interviews, 11.2 percent of weekly phone interviews, and 10.8 percent of triweekly interviews were not completed (table 4). Interviewing respondents more frequently thus did not induce higher noncompliance. Survey retention rates between baseline and endline interviews were even higher, with only 2 of 1,579 individuals not interviewed at endline and just 9 percent of respondents not available for an interview at the three-month follow-up.[9] Moreover, the overwhelming majority of participants (97.5 percent) indicated that they would be willing to participate again. These rates of compliance are much higher than in other studies (see, e.g., Dillon 2012; Croke et al. 2014; Garlick, Orkin, and Quinn 2015) and probably reflect a variety of factors, including a short questionnaire, flexible interview schedules, adequate compensation, the cultivation of trust, and the framing of the survey as being primarily about respondents' well-being.

[8] The duration of interviews did not decline over the course of the survey, suggesting that respondents did not (start) responding strategically to reduce the duration of the interview (by responding that they had not worked, they could have reduced survey duration by about a minute).

[9] Attrition was probably lower at endline than at the three-month because of follow-up incentives for respondents to participate in the endline round (when the bulk of the compensation was remitted) and the fact that the three-month follow-up interview was conducted only over the phone.

4. Data

Table 2 documents the baseline characteristics of the participants and compares them with the characteristics of urban residents who took part in the Ghana Living Standard Survey (GLSS) 2012/13, a nationally representative household survey. Just under three-fifths of respondents in the experiment were women, and the average respondent age was 34. Approximately two-thirds of respondents were employed, and approximately three-fifths of those working reported being self-employed. The null hypothesis that the average socioeconomic characteristics of experiment participants and their households are not statistically different from GLSS participants is not rejected; the poverty profile of households in the high-frequency survey also matches that of the GLSS households well.[10] In short, the sample used for the high-frequency data experiment is broadly representative of Ghana's urban working-age population.

Table 3 assesses balance across treatment arms at baseline for the sample that participated in the survey (sampling was done before the survey was fielded, so examining balance at baseline is a strong test of whether randomization was successful). Sample sizes per treatment arm differ slightly from the target of 320, with 318 respondents participating in weekly in-person interviews, 315 in weekly phone interviews, 321 in triweekly phone interviews, 314 in the control arm that received phones, and 311 in the control arm that did not receive phones. The treatment arms balanced on all variables that were used to stratify the sample, with the exception of mobile phone ownership, which is lower for the triweekly treatment arm but still high (92 percent of respondents owned mobile phones). Aside from being balanced on variables used for stratification save phone ownership. The treatment arms were also balanced (as a result of random selection) in other dimensions, including employment status, self- employment, and hours of work per week.

[10] Although the GUHPS 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 participants in this experiment.

5. Empirical Framework and Results

The following equation was estimated to assess the impact of survey modality and reference period on the reporting of labor market outcomes by respondents:

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

where Y_{iw} is a labor market outcome of interest (whether individual *i* reported doing any work in week *w*, days worked per week, and hours worked per week, both unconditionally and conditional on having reported any work that week); $Phone_i$ 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; $3XWeekly_i$ is a dummy variable that indicates whether individual *i* was interviewed three times a week (the omitted category is thus the weekly in-person interviews, which serve as a useful benchmark, as in-person interviews remain the dominant survey mode for both labor and household surveys); τ_w is a vector of calendar week dummies; and ε_{iw} is a random error term. [11] Triweekly labor market reports are aggregated to the weekly level in order to compare them with weekly labor market reports obtained in the weekly arms (see appendix C). Standard errors are bootstrapped and clustered at the treatment level (by household).

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

5.1 Impact of Survey Mode

Respondents assigned to the phone treatment arms reported significantly less employment, fewer hours, and fewer days worked than respondents interviewed in person. The effects are both statistically and economically significant. Relative to in-person interviews, phone-based interviews are associated with an 11 percentage point reduction in the likelihood of reporting any

[11] The results presented in this section are robust to controlling for individual and household characteristics as well as including initial conditions. Those results are not presented here, to conserve space. They are available upon request.

work at the survey mean. The reported number of days worked per week is 0.7 fewer overall (0.2 fewer conditional on reporting any work). The number of hours worked reported is 8 fewer overall (5 fewer conditional on reporting having done any work).

5.2 Impact of Reference Period

Respondents in the triweekly arm (who were asked to respond to two 24-hour recall questions) were 9 percentage points more likely to report having done any work in a given week than respondents with a one-week reference period. There was no change in the average reported number of days of work or hours worked overall. However, conditional on reporting having done any work that week, the number of hours fell by 4 and the number of days fell by 0.6. Shortening the reference period from one week to two days thus yielded more employment spells but not a significantly different number of days or hours worked overall. One explanation for these findings is that more frequent reporting improves accuracy by enabling respondents to better recall both short-lived employment spells and disruptions.

In sum, shortening the reference period is associated with reporting more work spells of shorter average duration. Conducting interviews over the phone instead of in person reduces both the incidence and duration of reported employment spells.

5.3 Differential Impacts on Wage- and Self-Employed Workers

Since self-employed workers typically have both greater flexibility and greater volatility in days and hours worked, as well as less clearly delineated boundaries between work and personal tasks, ex ante one might expect greater impacts of recall periods on reporting among the self-employed.

To check this, the respondent's baseline employment status was added to equation 1 and interacted with both survey mode and the reference period:

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

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 is a dummy variable that is equal to 1 if person i was not working at baseline. The omitted category is respondents who were wage-employed at baseline. As labor market status was very stable over the course of the survey, baseline employment status is a good measure of a person's employment type during the course of the survey and avoids potential concerns about the endogeneity of employment status to treatment.

The results are reported in table 6 and indicate that survey modality does not significantly affect labor market reporting by the wage-employed: None of the labor market outcomes reported by respondents who were wage-employed at baseline and assigned to the weekly phone treatment is different on average from the outcomes of respondents who were interviewed in-person. In contrast, all outcomes reported by the self-employed in the weekly phone treatment arm are significantly different from those reported by the self-employed interviewed in-person—and the differences are economically meaningful.[12, 13] Being interviewed over the phone is associated with a 15 percentage point lower likelihood of reporting any work in a given week, plus one fewer day worked and 12 fewer hours worked. Conditional on reporting any work, the self-employed who were interviewed over the phone also reported 0.2 fewer days and 2 fewer hours worked, but the overall effect was driven primarily by the reporting of more spells of work.

The length of the reference period also affects self-employed and wage-employed workers differently. Among the wage-employed, participants who were interviewed triweekly reported 0.6 fewer days worked overall (0.5 fewer days and 6 fewer hours conditional on reporting any) than participants who completed weekly interviews. They were not, however, more likely to report working.

Among the self-employed, those in the triweekly group, were 15 percentage points more likely to report doing any work than those interviewed weekly. Shortening the reference period

[12] People not working at baseline are an intermediate case. Respondents 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. The reporting of work, unconditional days, and unconditional hours does not differ between people who were not working at baseline, by whether they were assigned to the face-to-face or phone arms.

[13] The modality effect for the self-employed is the sum of the coefficients on the phone dummy and the interaction between the phone arm and being self-employed.

from a week to two days was also associated with reporting fewer days and hours conditional on working but not with reporting fewer days and hours overall. The effect of reporting more work spells (an "extensive margin" effect) and fewer days and hours conditional on reporting work (an "intensive margin" effect) roughly offset each other, so that total days and hours worked did not vary significantly between weekly and triweekly self-employed respondents. The finding that the overall labor supply reported by self-employed individuals does not decline when moving from a reference period of one week to two 24-hour recall periods whereas labor supply reported by wage employees does thus seems to reflect the self-employed reporting more work spells.

5.4 Differences between Current and Retrospective Reporting

How do labor market reports obtained during repeated weekly or triweekly surveys compare with a one-time retrospective labor market report using a three-month reference period? Answering this question not only helps assess how the reference period affects reporting, it is also informative about reporting differences induced by asking about respondents' usual instead of current activity that are not driven by seasonality, which is limited in urban Ghana. Retrospective questions about labor market activity over a three-month reference period asked about respondent's usual activity ("over the past three months, on how many days per week did you do this work on average?"). In contrast, the high-frequency interviews conducted as part of the experiment asked about current labor market behavior during the preceding week or two days.

Reporting differences are identified by estimating the following regression:

$$\mathrm{Y}_i = \beta_{F1} W1F2F_i + \beta_{P1} W1Phone_i + \beta_{P3} W3Phone_i + \beta_{CP} ControlPhone_i + \mathrm{c} + \varepsilon_i \quad (3)$$

where Y_i is a labor market outcome of interest (the endline report for control groups, aggregated weekly or triweekly interviews for treatment groups); c is a constant; and $W1F2F_i$ $W1Phone_i$, $W3Phone$, and $ControlPhone_i$ are dummy variables indicating assignment to the weekly in-person, 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 interviews conducted in-person and a one-time endline interview inquiring about labor market behavior over the previous three months. Because these endline interviews

were conducted in-person, the differences in labor market reporting can be ascribed to differences in the reference period only. The reports of individuals in the weekly and triweekly phone treatment arms are affected not only 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.

Shortening the reference period from three months to one week affects labor market reporting in a qualitatively similar way as shortening the reference period from one week to two days. Relative to respondents in the control group who did not receive a phone, respondents in the weekly in-person treatment arm were 11.5 percentage points more likely to report having worked over the survey period, and they reported working 0.8 additional weeks. However, conditional on reporting any work, this group tended to report working about half a day less a week and about 6 hours less a week than the control group that did not receive a phone. Shortening the reference period thus yields statistically significant and economically meaningful differences in labor market indicators, driven by reporting of more frequent employment spells of shorter average duration. These results resonate both with the results presented above and with earlier work examining recall bias in labor market reports (see, e.g., Bound, Brown, and Mathiowetz 2001).

High-frequency respondents interviewed by phone also reported systematically fewer days and hours than both control groups, despite a higher likelihood of reporting any work at all. Shorter reference periods were thus once again associated with reporting more work spells of shorter average duration.[14]

Comparing the endline labor market reports of participants who took part in the interviews with their own aggregated reports suggests that the reporting differences discussed in this section are related to recall bias (see appendix D).

[14] Comparing the labor reports of the controls who received phones with controls that did not suggests that receiving a phone is associated with small increases in reported weeks, days, and hours worked (both unconditionally and conditional on working), although the differences are in most cases significant only at the 10 percent level (an exception is unconditional hours, which is significant at the 5 percent level).

6. Conclusion

A randomized survey experiment was conducted in urban Ghana to understand the impact of reference periods and survey modality on reported labor market statistics. The reference period was shortened for some respondents to a 24-hour recall over 2 days as compared to the more usual one-week recall. Respondents were also randomly assigned to either an in-person or phone interview.

Shortening the reference period from one week to a 24- to 48-hour recall, results in a higher incidence of reported work among self- but not wage-employed participants. Conditional on doing any work in a given week, the number of hours and days worked decreased 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/triweekly interviews.

These findings have important implications for the design of labor force surveys, which typically ask respondents to recall their experience in the labor market over the previous week. Using shorter recall periods—of, say, 24–48 hours—is likely to substantially improve the accuracy of labor statistics, especially for the self-employed.

Relative to in-person interviews, phone-based interviews result in significantly lower reports of employment, hours worked, and days worked among the self-employed. For the wage-employed, modality does not appear to affect labor market reporting. While establishing why survey modality matters is beyond the scope of this paper, a possible explanation, 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, because their jobs and working hours may be less clearly defined than those of wage workers.

Although the experiment was conducted in urban Ghana, the findings are quite general and thus of likely relevance to both developing as well as advanced economies where a substantial fraction of the work force in engaged in self-employment.

Table 1: Interviews conducted

	Baseline (August–September 2013)	Weekly (10) (August–October 2013)	Triweekly (30) (August–October 2013)	Endline (October –November 2013)	Three-month follow-up (March 2014)
Treatment group					
Weekly face-to-face interview	✓	✓		✓	✓
Weekly phone interview	✓	✓		✓	✓
Triweekly phone interview	✓		✓	✓	✓
Control group					
Control without phone	✓			✓	✓
Control with phone	✓			✓	✓

Table 2: External validity: comparison with representative household survey

	Experimental data (GHFLS)	Representative household survey (GLSS 6)	Difference	p-value
Individual characteristics				
Male	0.415	0.431	–0.016	0.977
Age	33.734	35.553	–1.819	0.884
Years of education				
Women	8.343	9.621	–1.278	0.913
Men	9.873	12.871	–2.998	0.819
Employed				
Women	0.639	0.758	–0.119	0.853
Men	0.740	0.804	–0.064	0.925
Of which self-employed				
Women	0.687	0.748	–0.061	0.940
Men	0.485	0.460	0.025	0.980
Total sample size	1,579	13,204		
Household characteristics				
Household size	4.375	3.765	0.610	0.774
Dependency ratio	0.439	0.285	0.154	0.498
Maximum education level household (years)	11.484	14.914	–3.430	0.624
Female-headed household	0.354	0.292	0.062	0.878
Asset index	0.000	0.102	–0.102	0.939
Total sample size	573	6,947		

Note: The Ghana Living Standard Survey (GLSS) 6 is a representative household survey. For purposes of comparison, we use only urban households and focus on respondents in the age range 20-60 for individual characteristics.

Table 3 Baseline characteristics of respondents by treatment status

	No phone control mean	Treatment group (arms 3, 4, and 5)	In-person weekly	Phone weekly	Phone triweekly	Phone control
	(1)	(2)	(3)	(4)	(5)	(6)
Household-level variables						
Household size	5.310	–0.016	0.226	–0.118	–0.156	0.129
		(0.387)	(0.500)	(0.426)	(0.459)	(0.470)
Dependency ratio	0.207	0.009	0.026	0.014	–0.011	0.011
		(0.023)	(0.028)	(0.029)	(0.029)	(0.029)
Number of adults employed	2.578	–0.164	–0.105	–0.284	–0.105	–0.045
		(0.208)	(0.266)	(0.232)	(0.263)	(0.266)
Maximum education level in the household	12.491	–0.060	–0.073	–0.097	–0.010	0.275
		(0.299)	(0.346)	(0.358)	(0.385)	(0.349)
Asset index	–0.132	0.147	0.298	–0.080	0.221	0.299
		(0.195)	(0.227)	(0.251)	(0.236)	(0.219)
Number of observations		442	224	223	225	223
p-value for *F*-test for joint significance		0.827	0.581	0.795	0.916	0.738
Individual variables						
Male	0.403	0.010	0.026	0.003	0.001	0.025
		(0.032)	(0.036)	(0.040)	(0.040)	(0.040)
Married	0.465	0.003	0.040	–0.009	–0.022	–0.022
		(0.038)	(0.050)	(0.046)	(0.046)	(0.047)
Age	34.665	0.685	1.096	0.051	0.886	0.149
		(0.718)	(0.850)	(0.850)	(0.906)	(0.804)
Years of education	10.360	–0.220	–0.450	–0.089	–0.120	–0.113
		(0.296)	(0.379)	(0.360)	(0.364)	(0.333)
Employed	0.623	–0.010	0.040	–0.026	–0.044	–0.011
		(0.035)	(0.044)	(0.045)	(0.042)	(0.044)
Self-employed	0.290	0.040	0.083**	0.018	0.019	0.014
		(0.032)	(0.040)	(0.038)	(0.039)	(0.037)
Hours worked per week (total)	30.891	–1.904	1.274	–3.700	–3.298	–1.053
		(2.111)	(2.733)	(2.555)	(2.544)	(2.583)
Owns mobile phone	0.977	–0.029**	–0.013	–0.017	–0.055***	0.010
		(0.012)	(0.014)	(0.015)	(0.020)	(0.011)
Number of observations	303	1,211	606	601	610	605
p-value for *F*-test for joint significance		0.356	0.605	0.717	0.072	0.854

Sources: Household and demographic variables as well as education are from the Ghana Urban Household Panel Survey (GUHPS). Labor market variables and mobile phone ownership are from the Ghana high-frequency labor survey (GHFLS) baseline survey.
Note: Figures in parentheses are standard errors, clustered by household.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4: Percent of interviews missed

	Sample size	High-frequency interviews	Endline (in person)	Three-month follow-up (by phone)
Weekly in-person interview	318	6.4	0.0	9.4
Weekly phone interview	315	11.2**	0.3	7.0
Triweekly phone interview	321	10.8**	0.3	8.4
Control without phone	314		0.0	10.5
Control with phone	311		0.0	10.6
Total	1,579		0.1	9.2

Note: ** Significantly different from the "control without phone" arm at the 5 percent level.

Table 5: Impact of survey mode and reference period on labor market reporting in high-frequency surveys

	Working	Days	Days conditional on work	Hours	Hours conditional on work
	(1)	(2)	(3)	(4)	(5)
Phone	–0.106***	–0.684***	–0.201**	–7.742***	–4.795**
	(0.036)	(0.206)	(0.086)	(2.162)	(2.350)
Triweekly	0.089**	0.096	–0.601***	1.425	–3.973**
	(0.036)	(0.190)	(0.117)	(1.890)	(1.907)
Mean in-person weekly (omitted category)	0.650	3.499	5.378	31.424	48.345
Interview week dummies	Yes	Yes	Yes	Yes	Yes
Triweekly plus phone versus in-person	–0.016	–0.588***	–0.801***	–6.318***	–8.768***
Number of observations	7,813	7,813	4,699	7,813	4,692
Adjusted R^2	0.014	0.020	0.061	0.020	0.025

Note: Figures in parentheses are bootstrapped standard errors, clustered by household. Regressions are estimated by ordinary least squares. The omitted category is in-person weekly interviews. *Phone* is a dummy variable that takes the value 1 if the respondent was interviewed over the phone and 0 if she was interviewed in-person. *Triweekly* is a dummy variable that indicates whether the respondent was interviewed three times a week.

*** $p < 0.01$, ** $p < 0.05$.

Table 6: Impact of survey mode and reference period on labor market reporting in high-frequency surveys – heterogeneity by baseline employment status

	Working	Days	Days conditional on working	Hours	Hours, conditional on working
	(1)	(2)	(3)	(4)	(5)
Phone	–0.015	–0.110	–0.024	–2.657	–2.321
	(0.042)	(0.274)	(0.145)	(3.452)	(3.300)
Triweekly	–0.006	–0.482*	–0.558***	–4.777	–5.606*
	(0.042)	(0.286)	(0.178)	(3.161)	(2.880)
Self-employed (SE)	0.035	0.280	0.114	0.529	–1.375
	(0.037)	(0.266)	(0.152)	(3.633)	(3.056)
Not working (NW)	–0.688***	–3.631***	0.143	–33.915***	–0.443
	(0.049)	(0.297)	(0.161)	(3.512)	(4.006)
SE * Phone	–0.138***	–0.920**	–0.250	–8.932**	–3.657
	(0.049)	(0.361)	(0.208)	(4.412)	(4.036)
SE * Triweekly	0.155***	0.786**	0.008	9.426**	3.761
	(0.056)	(0.344)	(0.226)	(4.084)	(4.076)
NW * Phone	–0.002	–0.049	–0.558*	–0.052	–12.844**
	(0.062)	(0.369)	(0.313)	(4.085)	(6.100)
NW * Triweekly	0.068	0.654**	–0.240	6.466*	3.180
	(0.060)	(0.333)	(0.365)	(3.355)	(5.863)
Reference period effect self-employed	0.149***	0.304	–0.550***	4.649	–1.845***
(p-value)	0.000	0.231	0.001	0.136	0.511
Reference period effect not working	0.062	0.172	–0.798**	1.689	–2.426
(p-value)	0.145	0.403	0.013	0.263	0.621
Modality effect self-employed	–0.153***	–1.03***	–0.273**	–11.589***	–5.977*
(p-value)	0.000	0.000	0.048	0.001	0.053
Modality effect not working	–0.017	–0.159	–0.582**	–2.708	–15.165***
(p-value)	0.701	0.493	0.035	0.179	0.006
Mean wage-employed at baseline interviewed weekly in person (omitted category)	0.833	4.419	5.297	41.003	49.232
Interview week dummies	Yes	Yes	Yes	Yes	Yes
Number of observations	7,813	7,813	4,699	7,813	4,692
Adjusted R^2	0.420	0.378	0.067	0.298	0.038

Note: Figures in parentheses are bootstrapped standard errors, clustered by household. Regressions are estimated by ordinary least squares. *Phone* is a dummy variable that takes the value 1 if the respondent was interviewed over the phone and 0 if she was interviewed in person. *Triweekly* is a dummy variable that indicates whether the respondent was interviewed three times a week. *Self-employed (SE)* and *Not working (NW)* are dummy variables that take the value 1 if the respondent was self-employed or not working) at the time of the baseline survey. The omitted category is respondents who were wage employed at baseline assigned to weekly in-person interviews.
*** $p < 0.01$, ** $p < 0.05$.

Table 7: Retrospective reporting of "usual" employment (endline reports of controls) versus repeated reporting of "current" employment (aggregated high-frequency reports of treated)

	Any work	Weeks with any work	Average days per week	Average days per week conditional on work	Average hours per week	Average hours per week conditional on work
	(1)	(2)	(3)	(4)	(5)	(6)
Weekly in-person interview	0.115***	0.807*	0.110	–0.456***	–0.215	–5.994**
	(0.041)	(0.455)	(0.227)	(0.089)	(2.613)	(2.484)
Weekly phone interview (WP)	0.071*	–0.580	–0.623***	–0.817***	–8.236***	–12.905***
	(0.043)	(0.459)	(0.235)	(0.108)	(2.477)	(1.923)
Triweekly phone interview (TP)	0.105***	0.279	–0.641***	–1.470***	–8.173***	–17.459***
	(0.040)	(0.461)	(0.221)	(0.113)	(2.312)	(1.945)
Control with phone (CP)	0.064	0.822*	0.482*	0.143*	6.595**	4.110*
	(0.041)	(0.476)	(0.248)	(0.080)	(2.839)	(2.228)
p-value (WP = CP)	0.847	0.001	0.000	0.000	0.000	0.000
p-value (TP = CP)	0.293	0.197	0.000	0.000	0.000	0.000
Mean control without phone (endline report) (omitted category)	0.633	6.994	3.394	5.779	31.872	54.029
Number of observations	1,575	1,575	1,575	1,112	1,575	1,110
Adjusted R^2	0.005	0.008	0.027	0.237	0.038	0.126

Note: Figures in parentheses are bootstrapped standard errors clustered by household. Regressions are estimated by ordinary least squares. Omitted category is controls without phones.
Treatment arms: aggregated HF reports; Control arms: endline reports.
*** p <0.01, ** p <0.05, * p < 0.10.

References

Aker, J., M. Clemens, and C. Ksoll. 2011. "Mobiles and Mobility: The Effect of Mobile Phones on Migration in Niger." Proceedings of the German Development Economics Conference, Berlin.

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): 207–32.

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): 418–47.

Barrett, G., and D. S. Hamermesh. 2016. "Labor Supply Elasticities: Overcoming Nonclassical Measurement Error Using More Accurate Hours Data." NBER Working Paper 22920. National Bureau of Economic Research, Cambridge, MA.

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., J. Da Vanzo, N. Sastry, C. Panis, and C. Peterson. 2001. "The Quality of Retrospective Data: An Examination of Long-Term Recall in a Developing Country." *Journal of Human Resources* 36: 593–625.

Bonke, J. 2005. "Paid Work and Unpaid Work: Diary Information versus Questionnaire Information." *Social Indicators Research* 70 (3): 349–68.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics*, ed. J. Heckman and E. Leamer, 3705–843. Amsterdam: Elsevier.

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.

Carlin, P. S., and L. Flood, 1997. "Do Children Affect the Labor Supply of Swedish Men? Time Diary vs. Survey Data." *Labour Economics* 4(2): 167–83.

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: 76–88

De Leeuw, E. 1992. "Data Quality in Mail, Telephone and Face to Face Surveys." Netherlands Organization for Scientific Research, The Hague.

———. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys." *Journal of Official Statistics* 21 (2): 233–55

de Nicola, F., and X. Gine. 2014. "How Accurate Are Recall Data: Evidence from Coastal India?" *Journal of Development Economics* 106: 52–65.

Dillon, B. 2012. "Using Mobile Phones to Collect Panel Data in Developing Countries." *Journal of International Development* 24: 518–27.

Duncan, G. J., and F. P. Stafford. 1980. "Do Union Members Receive Compensating Wage Differentials?" *American Economic Review* 70 (3): 355–71.

Garlick, R., K. Orkin, and S. Quinn. 2015. “Call Me Maybe: Experimental Evidence on Using Mobile Phones to Survey African Microenterprises.” CSAE Working Paper WPS/2016-14. Centre for the Study of African Economies, Oxford University.

Goyal, A. 2010. “Information, Direct Access to Farmers and Rural Market Performance in Central India.” *American Economic Journal: Applied Economics* 2 (3): 22–45.

Gregson, S., T. Zhuwau, J. Ndlovu, and C.A. Nyamukapa, 2002. “Methods to Reduce Social Desirability Bias in Sex Surveys in Low-Development Settings: Experience in Zimbabwe.” *Sexually Transmitted Diseases* 29 (10): 568–75.

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–33S.

Holbrook, A. L., M. C. Green, and J. A. Krosnick. 2003. “Telephone versus Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Response Bias.” *Public Opinion Quarterly* 67 (1): 79–125.

Horvath, F. W. 1982. “Forgotten Unemployment: Recall Bias in Retrospective Data.” *Monthly Labor Review* 105 (3): 40–43.

Hussmanns, R., F. Mehran, and V. Verma. 1990. *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods.* Geneva: International Labour Office.

ILO (International Labour Organization). 2015. *World Employment and Social Outlook: Trends 2015*. Geneva.

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.

Juster, F. T., and F. P. Stafford. 1991. “The Allocation of Time: Empirical Findings, Behavioral Models, and Problems of Measurement.” *Journal of Economic literature* 29(2): 471–522.

Langhaug, L. F., L. Sherr, and F. M. Cowan, 2010. “How to Improve the Validity of Sexual Behaviour Reporting: Systematic Review of Questionnaire Delivery Modes in Developing Countries.” *Tropical Medicine and International Health* 15 (3): 362–81.

Mathiowetz, Nancy A., and Greg J. Duncan. 1988. “Out of Work, Out of Mind: Response Errors in Retrospective Reports of Unemployment.” *Journal of Business and Economic Statistics* 6 (2): 221–29.

Pierret, C. 2001. “Event History Data and Survey Recall.” *Journal of Human Resources* 36: 439–66.

Robinson, J. P., and A. Bostrom. 1994. “The Overestimated Workweek? What Time Diary Measures Suggest.” *Monthly Labor Review* 11–23.

Robinson, J. P., S. Martin, I. Glorieux, and J. Minnen. 2011. “The Overestimated Workweek Revisited.” *Monthly Labor Review* 134 (6): 43–53.

Stewart, J. 2014. “The Importance and Challenges of Measuring Work Hours.” *IZA World of Labor.* Available at https://wol.iza.org/articles/importance-and-challenges-of-measuring-work-hours/long.

Tack, J., and J. Aker. 2014. “Information, Mobile Telephony, and Traders’ Search Behavior in Niger.” *American Journal of Agricultural Economics* 96 (5): 1439–54.

Appendix A Variable Definitions

This appendix consists of five tables (tables A.1–A.5) that define the variables used in the paper.

Table A.1 Definition of variables used from 2010, 2012 and 2013 rounds of the Ghana Urban Panel Survey

Variable	Definition
Age	Age in years
Married	Married = 1, 0 otherwise
Male	Dummy variable = 1 if male, 0 if female
Employed	Dummy variable = 1 if respondent did any work for pay, profit, or gain in last seven days, even if only for one hour, or has job or work he or she will definitely return to, 0 otherwise
Self-employed	Dummy variable = 1 if primary work activity is self-employment, 0 if primary work activity is wage employment, out of labor force, or unemployed
Years of education	Years of formal schooling
Household size	Number of household members
Dependency ratio	Number of household members younger than 15 or older than 64 divided by household size divided by household size
Maximum education level in household	Years of formal schooling of most-educated household member
Number of adults employed	Number of household members 15–64 employed
Female-headed household	Dummy variable = 1 if household head is female, 0 if male
Asset index	Asset index derived from principal component analysis of the following household assets: radio, TV, sewing machine, air conditioner, refrigerator, freezer, gas stove, electric stove, cellphone, bicycle, motorbike, car

Table A.2 Definition of baseline variables

Variable	Definition
Employed	Dummy variable = 1 if respondent has stable work for pay or gain or reports working regularly, 0 otherwise
Self-employed	Dummy variable = 1 if primary work activity is self-employment, 0 if primary work activity is wage employment, out of labor force, or unemployed
Wage-employed	Dummy variable = 1 if primary work activity is wage employment, 0 if primary work activity is self-employment, out of labor force, or unemployed
Not working	Dummy variable = 1 if respondent is not working (unemployed or out of the labor force), 0 if respondent is self-employed or wage-employed
Hours worked per week	Total number of hours worked in a typical week (calculated by multiplying the number of days worked per week with the number of hours worked per day)
Owns mobile phone	Dummy variable = 1 if respondent owns mobile phone, 0 otherwise

Table A.3 Definition of variables in weekly/triweekly interviews

Variable	Definition
Phone	Dummy variable = 1 if respondent was interviewed over the phone (i.e., assigned to weekly phone interview treatment or triweekly phone interview treatment), 0 otherwise
Triweekly	Dummy variable = 1 if respondent was assigned to triweekly treatment, 0 otherwise
Called once week	Dummy variable = 1 if respondent was assigned to weekly phone interview treatment
Weekly in-person	Dummy variable = 1 if respondent was assigned to weekly in-person interview treatment
Three times a week by phone	Dummy variable = 1 if respondent was assigned to triweekly phone interview treatment
Control with phone	Dummy variable = 1 if respondent assigned to control but received a phone
Control without phone	Dummy variable = 1 if respondent assigned to control without receiving a phone
Working	Dummy variable = 1 if respondent completed work over the reference period
Days	Days worked per week
Days conditional on work	Days worked per week conditional on having reported any work that week
Hours	Hours worked per week (number of days worked per week X number of hours worked per day)
Any work	Dummy variable = 1 if respondent reported working in any of the high-frequency interviews
Weeks with any work	Number of weeks respondent reported working in repeated high-frequency interviews, extrapolated to a three-month period (by multiplying number of weeks with any work in the high frequency surveys by 12/*c*, where *c* is the number of weeks with high-frequency data)
Hours conditional on work	Hours worked per week (calculated by multiplying number of days worked per week times number of hours worked per day) conditional on having reported any work that week
Average days per week	Average number of days respondent indicated working per week in repeated high-frequency interviews
Average days per week conditional on work	Average number of days respondent indicated working per week in repeated high-frequency interviews, calculated over weeks respondent reported having worked
Average hours per week	Average number of hours respondent indicated working per week in repeated high-frequency interviews, calculated over weeks respondent reported having worked
Average hours per week conditional on work	Average number of hours respondent indicated working per week in repeated high-frequency interviews

Table A.4 Definition of variables in endline interviews

Variable	Definition
Any work	Dummy variable = 1 if respondent reported doing any work for pay or gain, even if only for a day, over preceding three months
Weeks with any work	Number of weeks worked over preceding three months
Average days per week	Average days worked per week over preceding three months
Average days per week conditional on working	Average days worked per week over preceding three months conditional on working
Average hours	Average hours worked per week over preceding three months
Average hours per week conditional on working	Number of weeks worked over preceding three months

Table A.5 Definition of variables in the Ghana Living Standards Survey 6

Variable	Definition
Age	Age in years
Male	Dummy variable = 1 if male, 0 if female
Years of education	Years of formal schooling
Employed	Dummy variable = 1 if respondent has stable work done for pay or gain and expects to continue doing it for next three months OR has done any work regularly, 0 otherwise
Self-employed	Dummy variable = 1 if primary work activity is self-employment, 0 if primary work activity is wage employment, out of labor force, or unemployed
Household size	Number of household members
Dependency ratio	Number of household members younger than 15 or older than 64 divided by household size
Maximum education level in the household	Maximum years of formal schooling by a household member
Female-headed households	Dummy variable = 1 if household head is female, 0 otherwise
Asset index	Index derived from first principal component of following assets: sewing machine, electric stove, gas stove, refrigerator, freezer, air conditioner, radio, TV, bicycle, motorbike, car, cellphone

Appendix B Additional Descriptive Statistics

This appendix consists of three tables (B.1–B.3) that provide additional descriptive statistics.

Table B.1 Characteristics of the wage- and self-employed at baseline

Characteristic	Wage-employed	Self-employed	Difference	p-value
Male	0.57	0.36	0.21	0.00
Age (years)	33.61	38.60	–4.99	0.00
Education (years)	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
Tenure (years)	5.51	7.89	–2.38	0.00
Multiple jobs (dummy)	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		
Less than 5 employees	0.30			
5–19 employees	0.33			
20–100 employees	0.19			
More than 100 employees	0.15			
No employees		0.76		
1 employee		0.12		
2–5 employees		0.11		
5–10 employees		0.01		

Table B.2 Results of verification check

	Congruence (percent)
Did you receive a phone at the time of the baseline interview?	98
Was it a new phone, packed in a box?	97
How often were you interviewed per week?	90
Employment status at baseline	93
Occupation at baseline	80

Note: Check administered to 5 percent of participants. Congruence means that the respondent gave an answer that was consistent with his or her treatment assignment or the answer he or she provided at baseline.

Table B.3 Completed high-frequency interviews as percent of scheduled interviews

Week	Weekly in-person (N = 318)	Weekly phone (N = 315)	Triweekly phone (N = 321)	
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	93.6	88.8		89.2

Appendix C Aggregating Triweekly Interviews

Interviews for respondents participating in the triweekly treatment arm took place on a set schedule (table C.1). The majority of labor market questions in this treatment arm referred to a 24-hour period, in particular "yesterday" and "the day before yesterday." As each interview covered two days, they typically covered only six days of any given week. Usually, Saturday was missed, as interviews were conducted on Tuesday, Thursday, and Saturday.

Table C.1 Interview schedule

Day of week interviewed	Number of interviews conducted	Percent of total
Sunday	153	1.60
Monday	9	0.09
Tuesday	2,997	31.36
Wednesday	177	1.85
Thursday	3,056	31.97
Friday	131	1.37
Saturday	3,035	31.75

To fill in this lack of coverage, we used an interpolation procedure when aggregating observations from the triweekly treatment arm to one-week periods. First, in the three-month follow-up we asked respondents how often they worked on Saturdays and when they did for how many hours on average. About 52 percent of respondents reported never working on Saturdays. Among respondents who did work on Saturdays, about 85 percent reported working every Saturday throughout the month. Having established these patterns, we adjusted aggregated work hours and days in the weekly data to reflect the estimated number of total hours for each respondent who did at least some work on Saturdays. For weeks during which no other work was reported, no adjustment was made to the total hours worked (i.e., adjustments were made conditional on having worked during the week). For respondents who worked all Saturdays, we added one day of work and the average number of hours worked on Saturday. For respondents who reported working only some Saturdays we followed the same procedure but instead of adding a day and hours to each week, we randomly selected a number of weeks in which they reported doing any work. Selection probabilities corresponded to the number of times respondents typically worked on a Saturday in any month (e.g., if a respondent reported working two Saturdays each month, then additional days and hours would be added to 50% of the weeks they reported doing any work).

Appendix D Assessing Recall Bias

To what extent did recall bias drive the findings? Comparing endline reporting of labor market behavior over the three-month period preceding the endline survey with a respondent's weekly or triweekly reports obtained during the same time period allows an assessment of the accuracy of retrospective reports over a three-month period. Differences in recall bias across treated groups are also informative about the impact of the length of the reference period and the modality.

To assess the extent of recall bias, we estimated the following regression:

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

where $\mathrm{Y}_{iEndline}$ is a labor market variable reported during the endline survey; $\overline{\mathrm{Y}}_{iHF}$ is the corresponding variable obtained by aggregating the weekly/triweekly labor market reports; and $W1F2F_i$, $W1Phone_i$ and $W3Phone_i$ are dummy variables that take the value 1 if the individual is assigned to the weekly in-person, weekly phone, and triweekly phone arms, respectively. The model is thus fully saturated. If $\Delta\overline{\mathrm{Y}}_i > 0$, individuals overreport in the endline one-time retrospective interview (whose design is similar to that of conventional labor market surveys inquiring about usual activity) relative to their own repeated reporting in the weekly/triweekly surveys. The coefficient ρ_{F1} is of particular interest, because it provides an estimate of the misreporting associated with a change in the reference period alone (on the assumption that high-frequency interviews yield more accurate data). As both the in-person weekly interviews and the endline interviews were conducted in person, a change in modality cannot be the cause of differences in reporting for subjects in this group. The coefficients ρ_{P1} and ρ_{P3} reflect a combination of both reference period and modality effects.

This analysis likely underestimates recall bias, because the retrospective endline reports of treated individuals differed somewhat from the reports of the controls (table D.1). Treatment respondents were significantly more likely to report having done any work (this effect is not significant for respondents assigned to weekly phone interviews). They also reported significantly fewer days and hours conditional on working. These effects do not persist, however. During the three-month follow-up interview, no statistically significant differences between the labor market reports of treatment and control arms were evident.

Table D.1 reporting at endline and three-month follow-up

	Any work	Weeks	Days	Days conditional on work	Hours	Hours conditional on work
Panel A: Endline interview						
Weekly in-person interview	0.090**	0.359	0.387	–0.181**	–0.237	–6.766***
	(0.043)	(0.479)	(0.258)	(0.086)	(2.704)	(2.466)
Weekly phone interview	0.058	–0.159	0.069	–0.355***	–4.254*	–10.167***
	(0.044)	(0.488)	(0.252)	(0.091)	(2.581)	(2.404)
Triweekly phone interview	0.104***	–0.097	0.340	–0.330***	–5.440**	–14.323***
	(0.040)	(0.439)	(0.234)	(0.082)	(2.406)	(2.216)
Control with phone	0.064	0.822*	0.451*	0.143*	6.595**	4.839**
	(0.044)	(0.486)	(0.246)	(0.077)	(3.064)	(2.354)
Mean control without phones (omitted category)	0.633	6.994	3.679	5.779	31.872	50.315
Number of observations	1,577	1,577	1,577	1,101	1,577	1,098
Adjusted R^2	0.004	0.002	0.002	0.039	0.019	0.077
Panel B: Follow-up interview						
Weekly in-person interview	0.061	0.684	0.223	–0.223	3.627	0.110
	(0.042)	(0.475)	(0.262)	(0.182)	(3.004)	(2.788)
Weekly phone interview	0.006	0.263	0.081	0.074	0.817	0.782
	(0.044)	(0.513)	(0.234)	(0.172)	(2.608)	(2.814)
Triweekly phone interview	0.027	0.560	0.275	–0.064	4.284*	1.853
	(0.041)	(0.456)	(0.233)	(0.190)	(2.565)	(2.635)
Control with phone	0.007	0.332	0.142	0.043	3.081	3.024
	(0.041)	(0.507)	(0.247)	(0.174)	(2.850)	(2.640)
Mean control without phones (omitted category)	0.667	7.106	3.370	5.139	30.968	47.231
Number of observations	1,413	1,413	1,413	974	1,413	974
Adjusted R^2	–0.001	–0.001	–0.002	0.000	0.000	–0.002

Note: Figures in parentheses are bootstrapped standard errors clustered by household. Regressions are estimated by ordinary least squares.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The results presented in table D.2 suggest significant and substantial recall bias. At endline, participants in the in-person weekly treatment arm underreported having done any work, as well as the total number of weeks worked, and they overestimated the number of days worked. Respondents in the weekly phone treatment arm overreported the number of weeks, days, and hours worked. This overreporting at endline is the mirror image of the underreporting documented during the weekly/triweekly surveys. Participants in the triweekly phone treatment

arm underreported the number of weeks worked but overreported the number of days worked, hours per week worked, and days worked conditional on working.

Table D.2 recall bias

	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 in-person interview (WF)	–0.022**	–0.403***	–0.432**	0.588***	0.860***	0.256	–4.590***
	(0.010)	(0.146)	(0.189)	(0.075)	(0.082)	(0.908)	(1.363)
Weekly phone interview (WP)	–0.016	0.400**	0.510**	0.968***	1.356***	3.907***	–2.112
	(0.015)	(0.198)	(0.260)	(0.108)	(0.120)	(1.202)	(1.493)
Triweekly phone interview (TP)	–0.003	–0.399***	–0.753***	1.257***	1.568***	2.659***	–1.073
	(0.017)	(0.146)	(0.181)	(0.096)	(0.120)	(0.849)	(1.240)
Modality effect (WP – WF)	0.006	0.803***	0.942***	0.381***	0.496***	3.651**	2.478
p-value	0.727	0.001	0.002	0.005	0.001	0.018	0.217
Frequency effect (TP –WP)	0.013	–0.799	–1.263	0.289***	0.213***	–1.248*	1.039*
p-value	0.317	0.984	0.222	0.000	0.000	0.050	0.051
Number of observations	948	948	664	948	663	948	661
Adjusted R^2	0.002	0.020	0.040	0.287	0.428	0.026	0.022

Note: Table shows difference between endline report and aggregated weekly/triweekly reports over three months. Figures in parentheses are bootstrapped standard errors clustered by household. Model is fully saturated. Regressions are estimated by ordinary least squares.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The results suggest significant and economically meaningful misreporting of employment in retrospective labor market surveys, probably because of the difficulties inherent in accurately recalling past employment spells. Even subjects who were primed to report accurately by repeated weekly/triweekly interviews misreported at endline relative to their previous reports.