Using Sensors to Measure Technology Adoption in the Social Sciences

Adina Rom,[*][†] *Isabel Günther,*[‡] *and Yael Borofsky*[§]

October 10, 2019

Abstract

Empirical social sciences rely heavily on surveys to measure people's well-being and behavior. Previous studies have shown that such data are prone to systematic biases caused by social desirability, recall challenges, and the Hawthorne effect, as well as random errors. Moreover, researchers often cannot collect high frequency data with surveys, which might be important for outcomes that vary over time. Innovation in sensor technology might address some of these challenges. In this study, we use sensors to describe the adoption of solar lights in Kenya and analyze the extent to which survey data are limited by systematic and random error. Sensor data reveal that households used solar lights almost every day and for four hours per day on average. On average, self-reported use does not differ from use measured with sensors, however, random measurement errors in surveys are large. Households that used the solar light a lot were likely to underreport use, while households that used it very little were likely to overreport. Whether a household received the solar light for free did not correlate with the household's tendency to over- or underreport. Asking about general usage provided more accurate information than asking about disaggregated use for each hour of the day. Frequent visits from surveyors for a random sub-sample increased solar light use as the Hawthorne effect would predict, but it had no long-term effects.

Key Words: Sensor, Self-Report Surveys, Measurement Error, Social Desirability Bias, Technology Adoption, Hawthorne Effect

[*] *ETH4D, ETH Zurich, Clausiusstrasse 37, 8092 Zurich, Switzerland*
[†] *Corresponding Author: adina.rom@nadel.ethz.ch*
[‡] *Development Economics Group, ETH Zurich*
[§] *Development Economics Group, ETH Zurich*

1 Introduction

Since the 1980s, advances in research design and analytical tools have increased the scientific impact and policy relevance of applied microeconomics, which Angrist and Pischke (2010) called a "credibility revolution." The increased use of natural experiments and randomized control trials (RCTs) were of particular importance to this development (Angrist & Pischke, 2010; Duflo, Glennerster & Kremer, 2008). Alongside this trend, there has been an increase in the collection of household-level survey data. While methodological and data advances have been remarkable, much of the research in applied microeconomics in low-income countries still relies heavily on self-reported survey data, which are prone to measurement errors and can be expensive to collect. Recent technological breakthroughs have pushed the field to a new frontier: improving the accuracy and precision of measurements of household well-being and behavior through the use of entirely new types of data, such as satellite imagery, cortisol stress tests, cell phone network data, and information from sensors. The hope is that these novel measurement techniques can help circumvent some of the challenges associated with self-reported survey data, including social desirability bias, sampling bias, recall bias, and the Hawthorne effect.

These biases may affect self-reported survey data in a number of ways. First, respondents may be prone to social desirability bias, meaning that they tend to answer with what they think the surveyor wants to hear, or what they think is socially desired (Bertrand & Mullainathan, 2001; Zwane et al., 2011; Nederhof, 1985). This bias might be particularly large when attempting to measure the adoption of a technology that is "desired," in the sense that using the technology has positive externalities. Examples of such socially-desired technologies include improved cookstoves (Ramanathan et al., 2016; Ruiz-Mercado et al., 2013; Thomas et al., 2013; Wilson et al., 2016), water filters (Thomas et al., 2013), latrines (Garn et al. 2017; Gautam, 2017), vaccines (Banerjee et al., 2010), and bed nets (Dupas & Cohen, 2010).

Second, the technological device might be used by a large number of people within a household and it may be infeasible to interview them all (sampling bias). Serneels et al. (2016) test, for example, if information about returns to education varies depending on whether the concerned person was asked directly or another household member answered for them. They find that it did not make a difference whether the information about employment and wages was self-reported or provided by an another household member.

Third, being surveyed (frequently) may make certain decisions more salient or remind people of certain (desirable) behaviors and thus influence respondents' behavior (Zwane et al., 2011; Smits & Günther, 2018). Simply knowing that one is being observed can also change respondents' behavior, which is referred to as the Hawthorne effect. The extent to which this effect influences social science research has been hotly debated (Adair et al., 1989; Leonard & Masatu, 2006; Levitt & List, 2011; Clasen et al., 2012; McCambridge et al., 2014; Simons et al., 2017). This effect is not specific to surveys, but can occur whenever participants know they are being observed. All of these problems can create systematic errors and thus, reduce the accuracy of any measurement. In experimental set-ups that focus on comparisons between different treatment groups, these errors are particularly problematic if they have varying effects across different treatment groups.

In addition to these well-known biases, respondents may simply not recall the answers correctly or surveyors may make mistakes in recording them (Beaman et al., 2014; Bertrand & Mullainathan, 2001; Das et al., 2012). Such recall errors become larger as more time passes between the event or behavior and the survey. However, some studies still ask about events that happened (far) in the past, since in many cases, collecting high frequency data is nearly impossible because it is intrusive, expensive, and logistically challenging. Thus, survey data tend to be noisy for everything that fluctuates over time, even if the average across the population is accurately estimated (e.g., incidents of diarrhea). While these sources of error do not necessarily lead to systematic error, they tend to add noise to the data and reduce the precision of estimates. Hence, these random errors reduce the chances of detecting an effect of a new technology or differences between sub-groups. Moreover, these types of error can still lead to systematic biases if they are more pronounced for certain sub-groups. Loken and Gelman (2017) even argue that measurement errors can increase the chances of finding spurious correlations in small sample sizes.

Recognizing these challenges, researchers have begun comparing different types of survey questions and methods. Typically, the goal of these studies is to measure the extent of the problem and to optimize survey tools. A number of studies discuss recall biases and optimal recall periods, for example. Das et al. (2012) find large differences among 1,621 individuals in Delhi in the answers to questions about weekly and monthly medical expenditures, morbidity, doctor visits, time spent sick, and whether a school or work day was lost due to illness. Beegle et al. (2012), on the other hand, use variation in the time between conducting the survey and harvest in three African countries and find little evidence of large recall biases.

Relatedly, some studies compare recall answers with diaries, where respondents are asked to fill out information on their own at a higher frequency. De Mel et al. (2009) find that Sri Lankan micro-enterprise owners report higher revenues and higher expenses with diaries as compared to recall surveys, however, the reported profits were similar. Along similar lines, Deaton & Grosh (2000) summarize several studies about household expenditures and find that respondents report substantially higher food expenditure in diaries compared with recall questions. There is a related debate about whether asking aggregated questions versus disaggregated questions leads to more accurate and precise estimates (Arthi et al., 2016; Daniels, 2001; De Mel et al., 2009; Grosh & Glewee, 2000; Serneels et al., 2016; Seymour et al., 2017). In the same study with micro-enterprise owners in Sri Lanka, the researchers find that owners' reports of overall firm profits tended to be more accurate than when they were asked about all the details concerning revenues and expenses. As a benchmark, they had research assistants surprise the enterprises and observe transactions (De Mel et al., 2009). Other studies, however, find that asking more detailed questions does lead to more accurate results. Serneels et al. (2016) suggest that asking one question about labor market participation instead of several detailed ones leads to significant biases when estimating returns to education. Seymour et al. (2017) conclude that asking individuals about their activities for specific time intervals throughout the day (time diaries) leads to more accurate answers than asking how much time individuals spend overall on certain activities. An important challenge for these types of studies is that they often compare different self-reported data. Thus, they tend to rely on benchmarks whose accuracy remains unclear.

In the pursuit of ways to mitigate these biases and measurement errors, researchers in various fields have turned to sensors as a means of complementing self-reported survey data and, hopefully, improving measurement. This shift has been enabled by the fact that prices for sensor technology[1] have dropped significantly and more "off-the-shelf" solutions have become available (IPA, 2016; Pillarisetti et al., 2017), allowing sensors to be used to collect data in studies with large sample sizes. This increase in accessibility provides opportunities for researchers to use sensor data, which allows them to avoid some of the problems posed by survey data described above. Sensors can be used to measure the adoption of various technologies, such as water filters, cookstoves, or, in our case, solar lights. At this point in time, however, they do not allow researchers to measure who uses the technology.

A small, but burgeoning body of research uses sensor data to understand technology adoption in low- and middle-income countries. Some of these studies also compare sensor data to survey data and discuss different types of systemic biases, as well as random errors. In a field experiment in Guatemala, Ruiz-Mercado et al. (2013) used stove use monitors in 80 households to study the use of improved cookstoves. The research team additionally administered quarterly surveys, which included questions about the frequency of stove use. They find that answers from the surveys were relatively consistent with sensor data. Wilson et al. (2016) studied cookstoves in 141 households in Darfur for 4-12 weeks and find that most respondents (83%) said they used cookstoves for every meal each day, while sensor data reveal that participants only used them for half of their meals. They also find that when surveyors announced their visits, use increased amongst those who had hardly used the stove before. Similarly, Ramanathan et al. (2016) find a tendency to overreport cookstove use among 456 households in rural India (which they observed over 17 months). They find little correlation between self-reported use of cookstoves and sensor data. In a field study in Rwanda, Thomas et al. (2013) compared reported usage of water filters (N = 63) and cookstoves (N = 70) from monthly surveys with sensor data from the same respondents. Since the sensors they used were expensive (US $500 each), they deployed 50 sensors and rotated them every two weeks over the course of the five-month study. They find that respondents significantly overreport (by 17 percentage-points) the use of the improved cookstove. The overreporting for water filters was less pronounced, but still fairly large (6 percentage-points).

[1] IPA (2016) defines a sensor as a "device used to measure a characteristic of its environment—and then return an easily interpretable output, such as a sound or an optical signal. Sensors can be relatively simple (e.g., compasses, thermometers) or more complex (e.g., seismometers, biosensors)."

Most existing studies on different technologies suggest that survey data can significantly overestimate the adoption of socially desired technologies. Two mechanisms might explain that. First, respondents may overreport if they think they are expected to use the new technology (social desirability bias). Second, frequent interviews may change respondents' actual behavior (Wilson et al., 2016; Zwane et al., 2011). Yet, we still know little about the conditions and types of technologies that accentuate these biases. For example, if "reciprocity" is a source of the bias – in the sense that people feel they need to be "nice" (i.e., report that they use the product a lot) because they received a free good – it could well be the case that the bias is weaker or even absent when households purchase the relevant product.

In this study, we use data from 220 sensors and a corresponding household survey to describe patterns of solar light usage. As Rom & Günther (2019) show, switching to renewable energy sources and more energy efficient appliances can have important health and environmental benefits. However, these benefits only occur if households actually use the solar light and reduce the use of kerosene accordingly. As has been shown for the case of cookstoves, even very promising technology can fail to be effective because it is simply not used (Hanna, Duflo & Greenstone, 2016). Therefore, it is crucial to get an accurate understanding of households' solar light use patterns to estimate the effect of this technology. In the second part of the paper, we compare sensor data with survey data, testing several hypotheses put forward in the literature about the accuracy and precision of survey data. Sensor attrition was a problem, as 23.2% of sensors stopped functioning within the first 3.5 months of the study and 37.7% before the end of the study.[2] For this reason, we focus the analysis on measurements taken in the first month of the study, when 93.2% of the sensors still worked. Survey response attrition at the end of the study was 5.9% for adults and 9.1% for pupils.

In contrast to much of the previous literature on cookstoves, we do not find systematic overreporting of usage. In fact, the averages of survey data and sensor data look fairly similar and the sensor measures are even slightly higher. Households that hardly used the solar light, however, tend to overreport use, which is consistent with social desirability bias. Households that use the solar

[2] It was difficult to confirm why sensors stopped functioning without potentially damaging a respondent's light, however, we know the most likely reasons for attrition are that the sensor simply malfunctioned, the battery failed if the light was not in use for several days, or the light broke and disabled the sensor.

lights frequently, on the other hand, tend to underreport use. Second, and consistent with the Hawthorne effect, we find that more frequent household visits from surveyors increased use of solar lights initially, but had no effect in the long run. Third, we find that time diary questions reflect usage patterns throughout the day, on average. However, there is little correlation between the time-use diary estimates and the sensor data at the households level, and there is less than when using aggregated questions. Finally, increased precision of the sensor data allows us to see usage patterns of sub-groups more clearly, which reveal that poorer households tend to have higher solar light use.

Our findings have a number of implications for survey and sensor measurements. First, the added value of sensors seems to be particularly high when biases are expected to be large, especially when adoption of the technology is low or when precise estimates are needed to answer the survey questions, e.g., if the sample size is small or sub-group analysis is important. Second, for surveys, our data suggest that asking about global use estimates provides more accurate results than asking two household members about their individual use throughout the day (time diary) and combining them. Thus, while time diary questions are relevant for understanding use patterns over the course of the entire day, they do not seem to be ideal for understanding global use of a shared technological device. Third, we find that frequent interactions with field staff can temporarily increase use of new technologies, suggesting that researchers need to think carefully about how interactions with the field staff could bias results and, if this is a concern, aim to find ways to measure these surveyor effects. Finally, since sensor attrition can be high, study designs should allow researchers to answer their main questions early on, and a first round of sensor data should be collected very soon after baseline, if sensor data collection does not happen remotely. Researchers using sensors should also put a protocol in place in case the studied technology (in our case, solar lights) or the sensors break before the study ends.

2 Study Design, Technology, and Data

The sensor data used in this paper is part of a larger randomized controlled trial (RCT) conducted between June 2015 and March 2016 in two sub-counties in Busia, western Kenya. The sample contained 1,410 randomly selected households. In total, 400 households were assigned to the control group, 400 received a solar light for free, and 610 households randomly received an offer to buy a solar light at either 900 KES (US $9), 700 KES (US $7), or 400 KES (US $4). The randomization was conducted at the household level.

Households which received free solar lights were given either a Sun King Eco or a Sun King Mobile light (see Appendix A, Figure A.6 and A.7 for pictures), both manufactured by Greenlight Planet and quality assured by Lighting Global, a joint initiative of the World Bank and the International Finance Cooperation. At the time of the study, the Sun King Eco sold for US $9 in Kenya and the Sun King Mobile for US $24. According to tests conducted by Lighting Global, the Sun King Eco provides light for 5.8 hours when used at its maximum brightness of 32 lumens. The Sun King Mobile can be used for 5.4 hours on its brightest mode (98 lumens) and can also charge a mobile phone (Greenlight Planet, 2016; Lighting Global, 2015). For comparison, a simple tin lamp, which is what was most often used for indoor light in our sample, provides around 7.8 lumens and a kerosene lantern provides 45 lumens (Mills, 2003). Thus, both types of solar lights provide much stronger light than the tin lights. Half of the households that received a solar light for free got a Sun King Eco and half received a Sun King Mobile. Discount vouchers were offered for the Sun King Eco model.

Of the 400 solar lights that were distributed for free to households, 164 were equipped with a sensor that measured usage. Households only learned about the sensors when we asked for permission to download their data for the first time, which was a few months after baseline. The research team only accessed the data if the respondent gave permission for them to do so. Of the 130 solar lights that were sold to households at either 900 KES (US $9) or 700 KES (US $7), a sub-sample of 56 solar lights was equipped with a sensor that collected data. Thus, in total, we had 220 solar lights equipped with a functioning sensor (see also, Section 2.1). This design also allowed us to compare use between households who received a solar light for free and households who purchased one.[3]

[3] A total of 610 households received an offer to buy a lamp, but only 274 bought one. Out of these, 130 were sold at either 900 KES (US $9) or 700 KES (US $7) and the remaining 144 were sold at 400 KES (US $4).

2.1 Sensor Data

We have sensor data for a total of 220 households for at least part of the study period starting in August 2015 and ending in March 2016. By the household survey endline (February-March), around a third of sensors had stopped recording data, such that we were left with 147 sensors. The sensors stopped recording data either because the battery life ended, the sensor was faulty (manufacturing errors), or the solar light stopped working. It is possible that the point in time at which the sensor stopped working is correlated with usage. It could be that some sensors may have stopped working because the solar light was not used for a number of consecutive days. However, it is also likely that solar lights that are used more intensively tend to break more often. When we compare lights that broke in the previous month of usage with lights that did not, the coefficients go in different directions and we cannot conclude that one effect dominated the other (see Table B.3 in Appendix B). For these reasons, it is possible that we under- or overestimate usage when using data from the end of the study. To avoid possible biases in sensor measurements, we focus most of our analysis on the first month of sensor data collection only, when 93.2% of the sensors were still working by the end of the month. We replaced data points with missing values once the sensor stopped logging data. In this sense, all results should be interpreted as "usage conditional on the lights functioning."

For the sensor data, we report the following measures of average daily solar light use:

- Entire Study (all): recorded use by sensors, no matter how long they worked (N = 220). Data were used from all the days that we have data for. Once a sensor stopped working the remaining days were coded as missing. Months included: August 2015-March 2016. Variable: Sens (All)
- Entire Study (worked entire study): sensors that worked until the end of the study. Data were used from all the days we have data for (N = 147). Months included: August 2015-March 2016. Variable: Sens (All) worked until End.
- First Month (all): recorded use by sensors, no matter how long they worked (N = 220). Data were used from the first month of the study.

Once a sensor stopped working the remaining days were coded as missing. Month included: August 2015. Variable: Sens (Aug)

- Previous Day: sensors that worked until the end of the study (N = 147). Data were used from the day before endline data collection. Days include: Varying days in February and March 2016, depending on the day the endline was conducted in each household. This measurement was used since we asked about solar light use on the previous day in the survey, which is easy to compare with sensor data. Variable: Sens (Yest.)

Sensors tracked when the solar lights were turned on and off. Based on this information, we calculated the total number of minutes a solar light was used on any given day of the study. Independent of the measure used, we first calculated average use by sensor, meaning that we always weight each sensor equally, regardless of the number of days of data we have.

A random sub-sample of those with a solar light sensor (37.1%) were subject to around five additional household visits. Other studies have found that more frequent interactions between households and surveyors led to increased use, so we also use this variation to see whether additional household visits lead to more solar light use in our study (Wilson et al., 2016; Zwane et al., 2011).

2.2 Sensor Technology

We used Bluetooth-enabled Solar Lamp Usage Monitors (referred to as sensors or solar sensors throughout this paper) to determine when the lamp was in use by measuring the change in voltage of the solar lamp's light emitting diode (LED).[4] This sensor was installed by soldering the sensor to the board inside the light. Using smartphones enabled with Bluetooth and an iPhone App ("Lamplogger") developed specifically for these sensors, field officers visited households and wirelessly uploaded data directly from the sensor to the phone. These sensors, along with the iPhone application, were specifically developed for this study. Since the use of sensors in field experiments is still relatively new and other researchers may find themselves in a similar situation to ours, we share a few key lessons learned about implementing and managing sensor technology in the field in Appendix C.

[4]Sensors were developed by Bonsai Systems: https://www.bonsai-systems.com

2.3 Survey Data

The endline household survey was conducted in February and March 2016 and contained, among others, questions about household light use habits (the full survey is available from the authors upon request). Information about solar light use came from two separate questions:

- Aggregated Question (see Appendix D): one question asked the adult respondent for an estimate of total light used by the household on the previous day (N = 161). Variable: Surv (Aggr.)

  It is important to note that a respondent was only asked the Aggregated Question if they indicated that "any of their solar lights still works," due to a skip pattern in our survey instrument. A total of 53 households reported that their solar light did not work. Of these 53 households, 21 (40%) still had a working solar light and had, according to the sensor data, used it the previous day, suggesting that either they did not understand the question, did not know that their light still worked, or intentionally deceived the surveyors.[5] Thus, we only have an answer to the Aggregated Question from 161 respondents, from whom we also have sensor data.

- Detailed Question (see Appendix D): a separate battery of questions asked each individual about their activities and light use[6] for specific half-hour time slots between 7:00 pm and 7:00 am, corresponding to nighttime (dark) hours in Kenya. We faced a trade-off between level of detail and survey length. Ultimately, we only asked for this level of detail about light usage at night in order to limit both financial costs and the opportunity cost to respondents in terms of patience and attentiveness (N = 215 for adults, N = 205 for children). Variable: Surv (Detail)

We asked both adults and pupils the Detailed Questions. We aggregated the half-hour time slots for adults and for children separately and combined. To cal-

[5] According to sensor data, households which indicated that at least one of their solar lights worked during endline did not use their solar lights for different amounts of time per day than households that said that none of their solar lights worked.

[6] Options: Electricity-powered lighting, Solar home system powered lighting, Tin Lamp, Kerosene lantern/Hurricane, Fire, Wood, Battery-powered torch/lantern, Candle, Solar lantern/solar torch, Pressurized Kerosene Lantern, Other rechargeable lantern, Cell phone light, No lighting used, Matchsticks, Other.

culate these measures, we created a dummy indicating whether the adult/child used the solar light during each time slot. For the combined measure we created a separate dummy equal to one if either the child or the adult used the solar light and zero otherwise.

3 Use of Solar Lights

Sensors provide detailed information on how usage of a technology varies throughout the day, throughout the week, and throughout the month. As discussed in the previous section, we focus on results from the first month of the study (August 2015) for the analysis of solar light use, since about 93% of the sensors worked through August, whereas by March 2016, an additional 13.6 sensors had dropped out each month (on average). That said, results for the entire study period are very similar to results from the month of August. For each table and graph presented in this section (focusing on August 2015), we also refer to the corresponding table and graph reporting the results for the entire study period in Appendix A and B.

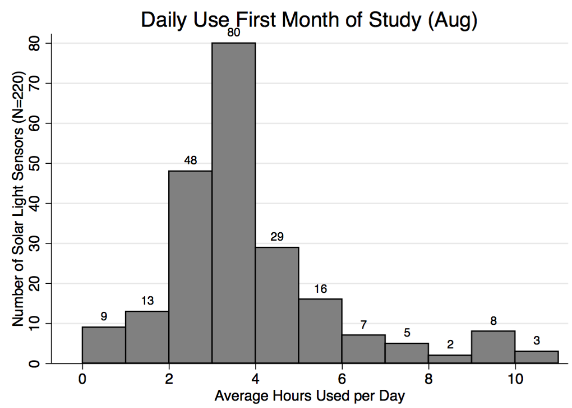
Households used the solar light on average 6.4 out of seven weekdays and 58.6% of households used the solar light on every single day of the study. Households used the solar light for 3.86 hours per day and 71% of households used the solar lights between two and five hours per day (see Figure 3.1 and Table 4.1, Row 3). Daily use across the entire study period is actually slightly higher (4.07 hours per day), possibly since schools were still closed during the first two months of the study (Table 4.1, Row 1). There are only nine households (4% of all households with sensors) who used the solar light for less than one hour per day on average (Figure 3.1). The corresponding distribution for the entire study period can be found in Figure A.3 in Appendix A.

These findings of high rates of solar light usage across all households contrast with recent findings about improved cookstoves. Wilson et al. (2016), for example, find that 29% of households hardly used the technology.[7] In addition, sensors allow researchers to collect data over a long period of time. Such information is usually very time consuming and intrusive to collect with surveys, especially if a technological device is used by several people, who all need to be

[7] They defined "non-users" as those using the cookstove less than once on 10% of days.

asked about the timing of their usage individually and repeatedly. For example, in our case, the adults we interviewed simply might not know whether their children used the solar light at night. One would have to separately ask all household members to get the full picture.

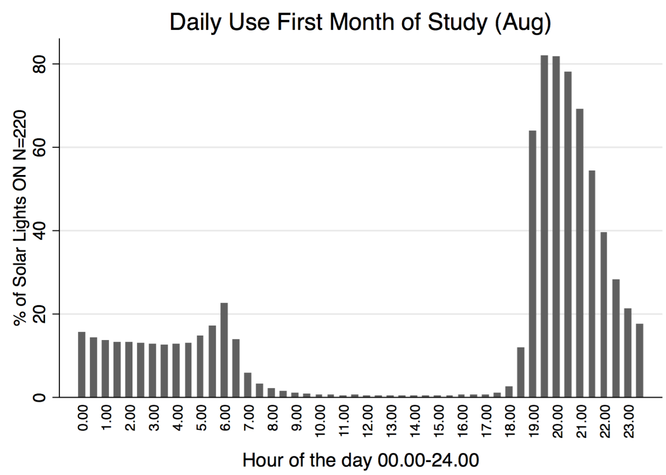
Figure 3.1: Average Hours Solar Lights are Used per Day



Notes: This graph shows sensor data about the average number of hours the solar lights were used per day during the first month of the study.

Figure 3.2 shows the share of solar lights that were used, reported in half-hour slots, averaged over all days of the first month of the study. We created a dummy for every half-hour slot, which is equal to one if the solar light was used for more than 15 minutes in a row during that half hour and zero otherwise. We then calculated for each sensor the percentage of days that the light was on (as a percentage of all days that the sensor worked in August) and used this information to calculate the average across all sensors. We find that households mostly use the solar light during evening hours. The half-hour interval when most solar lights (81.94%) were switched on was between 7:30 pm and 8:00 pm, which is right after sunset in Kenya. As expected, there is also a peak, albeit a smaller one, during morning hours, in particular between 6:00 am and 6:30 am. Interestingly, between 15-20% of households also have the solar lights switched on during nighttime hours. Anecdotal evidence suggests that, among other reasons, some use the solar light as a security light during the whole night or when they get up to use the restroom or check on their cattle. As expected, use is lowest during the day — 1.05% used them during daytime (between 9:00 am and 5:00 pm) (Figure 3.2).

Figure 3.2: Use Across the Day

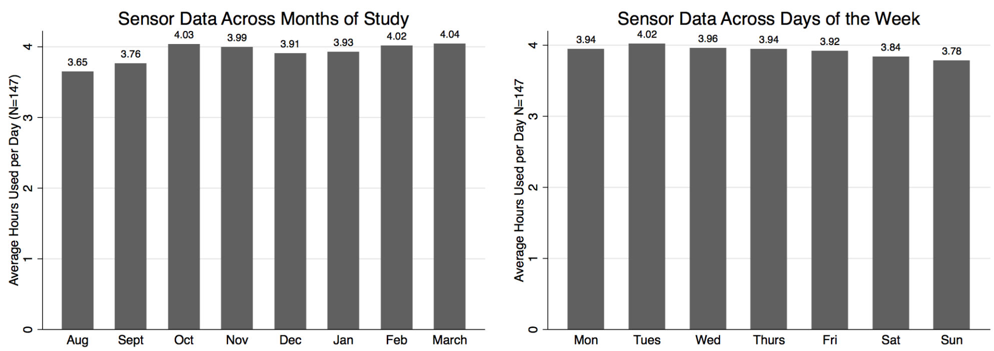


Notes: We classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors.

It is difficult to get information about use over a long period of time from survey data because conducting many survey rounds is costly and asking respondents about time periods that lie far in the past might lead to noisy and perhaps even biased results (recall bias). Thus, sensors can also be used to study changes in use over time. Households might increase use of a product as they learn about its advantages or develop a habit of using it. Households might decrease use if they discover unexpected disadvantages or if their excitement over the novelty of the product wears off over time. Use could also vary with the schooling or agricultural schedule. Figure 3.3 shows use over the eight months of the study period for the 147 solar lights for which we have data until the end of the study. Use was slightly lower in August and September, but none of the differences are statistically significant (Appendix B, Table B.1). This pattern could be linked to the fact that schools were closed in August, due to holidays, and in September, due to a teacher strike. However, as previously explained, around one third of the sensors did not survive until the end of the eight-month study and we do not know how use would have evolved amongst those households whose lights/sensors did not survive.

The second figure breaks down usage by day of the week (Figure 3.3). We observe that solar lights are used less on the weekend. This difference is statistically significant at the 5% level (Appendix B, Table B.2).
On average, households switched the light on and off 4.74 times per day (SD 3.35) with each on/off event lasting an average of 50.71 minutes (SD 93.32); 50% of all use events were shorter than 12 minutes.

Figure 3.3: Daily Use Across Months of the Study and Across Days of the Week



Notes: Sample is restricted to sensors that worked until the end of the study.

4 Comparing Survey and Sensor Data

In this section, we analyze whether estimates of technology use measured with survey data are similar to estimates of technology use measured by sensors. Moreover, we test several hypotheses that have been discussed in the literature about what drives the accuracy of survey data. Lastly, we analyze whether sensor data, which measure technology use with higher precision, allow us to detect differences across sub-groups or experimental treatments with smaller sample sizes.

4.1 Averages from Sensor and Survey Data are Similar

Comparing the three different survey answers with the sensor data, we find that the averages from the sensor data and from the survey data are relatively

Table 4.1: Mean Light Use (Hrs) per Day: Survey and Sensor Data

	(1) All Data Mean (SD)	(2) All Data Obs	(3) Exclude Missing Means (SD)	(4) Exclude Missing Obs
(1) Sens (All)	4.067	220	3.813	125
	(1.776)		(1.464)	
(2) Sens (All)- Worked until End	3.731	147	3.813	125
	(1.404)		(1.464)	
(3) Sens (Aug)	3.864	220	3.607	125
	(2.031)		(1.846)	
(4) Sens (Yest.)	3.706	147	3.777	125
	(2.132)		(2.247)	
(5) Surv (Detail)	3.388	215	3.616	125
	(1.764)		(1.625)	
(6) Surv (Detail)- Adult	3.193	215	3.152	125
	(1.377)		(1.371)	
(7) Surv (Aggr.)	3.573	161	3.492	125
	(2.073)		(2.030)	

Notes: Column 1 and 2 include all data, Column 3 and 4 only the 125 observations where we have all sensor and survey variables listed in this table (see Section 2 for further explanations). Row 1 includes all sensors no matter when they stopped working, Row 2 includes only sensors that worked until the end of the study, Row 3 includes data from all sensors for the month of August only, Row 4 includes sensor data for the day before the study, Row 5 shows survey data from the Detailed (hour by hour) Questions for adults and pupils combined, Row 6 shows the same question as Row 5, but only for adults, and Row 7 shows the Aggregated Questions where we asked about use of the entire household (see questions in Appendix D). Note that the survey question refers to the day before.

Table 4.2: Differences in Light Use (Hrs) per Day: Survey and Sensor Data

	(1) Sens (All) All	(2) Sens (All) Worked until End	(3) Sens (Aug) All	(4) Sens (Yest.)	(5) Surv (Detail)	(6) Surv (Detail) Adult	(7) Surv (Aggr.)
(1) Sens (All)	0.000	0.000	0.203 ***	0.025	0.654 ***	0.849 ***	0.513 ***
	(220)	(147)	(220)	(147)	(215)	(215)	(161)
(2) Sens (All)	-	0.000	0.229 ***	0.025	0.074	0.557 ***	0.321 *
Worked until End		(147)	(147)	(147)	(147)	(147)	(125)
(3) Sens (Aug)	-	-	0.000	-0.204	0.450 **	0.645 ***	0.288
			(220)	(147)	(215)	(215)	(161)
(4) Sens (Yest.)	-	-	-	0.000	0.049	0.532 ***	0.285
				(147)	(147)	(147)	(125)
(5) Surv (Detail)	-	-	-	-	0.000	0.195	0.085
					(215)	(215)	(161)
(6) Surv (Detail)	-	-	-	-	-	0.000	-0.430 **
Adult						(215)	(161)
(7) Surv (Aggr.)	-	-	-	-	-	-	0.000
							(161)

Notes: This table shows differences between variables in Rows and Columns. *** p<0.01, ** p<0.05, * p<0.1. Number of observations are shown in brackets. Number of observations varies since we do not have sensor data for all sensors until the end of the study and we do not have all survey measures for all observations.

similar for most measurements (Table 4.1). We also observe that the aggregated survey question (Table 4.2 Column 7, Rows 1-4) elicits answers that are more similar to the sensor data than the detailed questions (Table 4.2 Column 5, Rows 1-4). This finding will be further discussed in Section 4.4. When we asked respondents to estimate the overall use of the solar light, on average, their estimates were statistically not different from the sensor data, with the exception of one measure (Table 4.2, Column 7, Rows 1-4). The self-reported estimate is only different from the first sensor measure (i.e., across all sensors and the entire study period, Table 4.2, Column 7, Row 1) and, interestingly, the self-reported measure is smaller than the measure from the sensors. This finding stands in contrast to most of the recent literature (Thomas et al., 2013 or Wilson et al., 2016, for example) studying the use of improved cookstoves with sensor and survey data, which finds that respondents tend to overreport use on average. There is, however, an important difference between our study and previous work, namely that, in our case, adoption of solar lights was nearly universal, while adoption of improved cookstoves was typically low (see Section 3 for further discussion). We also find that households that hardly use the solar lights tend to overreport use, while households that use the solar light a lot tend to underreport use (Table 4.3, Row 1). In our case, however, there are very few households that hardly use the solar lights and hence, we do not find overreporting on average. This will be further discussed in the next section.

4.2 Frequent Users Underreport - Infrequent Users Overreport

We analyze whether certain sub-groups are more likely to under- or overreport usage. First, we test whether households that received a free solar light are more likely to overreport use of solar lights than those who paid for it. If there is a reciprocity dynamic at play, in that recipients of free lights feel more obliged to say positive things about the gift they received, then one would expect that households receiving a free light would be more likely to overreport use. Our data do not confirm this hypothesis (Table 4.3, Column 2). There is also no evidence that the gender of the respondent influences the extent to which their answers differed from the sensor measurement (Table 4.3, Column 4).

In our setup, a random sub-sample received more frequent visits and monitoring.

Table 4.3: Analysis of Under- and Overreporting Solar Light Use

VARIABLES	(1) Diff Sens-Surv (All data)	(2) Diff Sens-Surv (All data)	(3) Diff Sens-Surv (All data)	(4) Diff Sens-Surv (All data)	(5) Diff Sens-Surv (All data)	(6) Diff Sens-Surv (All data)	(7) Diff Sens-Surv (All data)
Hours Used (Sensor)	0.683***	0.682***	0.687***	0.671***	0.673***	0.677***	0.688***
	(0.123)	(0.123)	(0.123)	(0.123)	(0.168)	(0.131)	(0.124)
Free Solar Light		0.261					
		(0.347)					
Additional Visits			0.384				
			(0.320)				
Respondent Female				-0.305			
				(0.339)			
Wealth Index					-0.075		
					(0.120)		
HH Head Yrs of Schooling						0.038	
						(0.046)	
HH Size							-0.045
							(0.058)
Constant	-2.279***	-2.470***	-2.371***	2.104***	-1.809*	-2.494***	-2.003***
	(0.449)	(0.542)	(0.447)	(0.515)	(0.980)	(0.540)	(0.543)
Observations	161	161	161	161	120	152	161
R-squared	0.247	0.250	0.254	0.251	0.227	0.236	0.249
Mean	0.513	0.513	0.513	0.513	0.513	0.513	0.513

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Includes all 161 sensors for which we have the aggregated survey measure. Column 5 has fewer observations since we only collected data on assets for a subgroup. The wealth index includes information about the household's ownership of bikes, motorbikes, cars, stoves, radios, wall clocks, tin lamps, kerosene lanterns, solar lanterns, electrical lanterns, tables, beds, bed nets, chairs, sofa pieces, jembes, car batteries, animal charts, horses, cattle, goats, sheeps, chickens, pigs, mobile phones, and sim cards.

Table 4.3 shows that these additional visits did not make households more likely to overreport use, which is probably reassuring for researchers worried that frequent interactions with surveyors might alter responses (Zwane et al. 2011). We also test for a change in behavior (higher usage) among those households in Table 4.3. Households' wealth and size, as well as the household head's level of education, do not correlate with reporting differences (Table 4.3, Column 5-7). We do find, however, that underreporting strongly correlates with use. Thus, the more a household uses a solar light, the more likely they are to underreport use. The opposite also holds, whereby infrequent users are more likely to overreport (Table 4.3 and Appendix A, Figure A.2). There could be a couple of explanations for this observation. First, respondents could have a certain reference point in mind regarding reasonable light use that they report regardless of actual light use. It is also possible that underreporting occurs because respondents are not aware of other household members' use (especially in high-usage households), while respondents who hardly use the solar light overreport because they feel they are expected to use the light (social desirability bias).

4.3 Intense Monitoring Increased Use Temporarily

A random 37% of the sampled households with solar lights and sensors were exposed to more frequent visits by surveyors at the beginning of the study. More frequent visits did increase use in the first month of the study, however, this difference disappeared thereafter (Table 4.4). Different mechanisms might explain this difference: respondents might have felt more observed and used the novel product more as a result (Clasen et al. 2012; Leonard & Masatu 2006; Simons et al., 2017), the visits may have made the product more salient, i.e., reminded respondents of the product (Zwane et al., 2011; Smits & Günther, 2018), or the surveyors might have accelerated learning about the product. As we previously saw, more frequent visits did not lead to more overreporting (Table 4.3, Column 3).

4.4 Time Diary Questions vs. Aggregated Questions

In the survey, we asked about solar light use in two different ways. First, we asked adults and children to report the activities they engaged in for each one-

Table 4.4: Hawthorne Effect

VARIABLES	(1) Sensor (Hrs) First Month	(2) Sensor (Hrs) First Month	(3) Sensor (Hrs) All Months	(4) Sensor (Hrs) All Months
Frequent Visits	0.584**	0.589**	0.339	0.278
	(0.284)	(0.296)	(0.253)	(0.239)
Observations	220	147	220	147
R-squared	0.019	0.026	0.009	0.009
Mean	3.646	3.285	3.941	3.629

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Column 1 and 3 include all sensor data, while Column 2 and 4 are restricted to those that worked until the end of the study. No control variables were used.

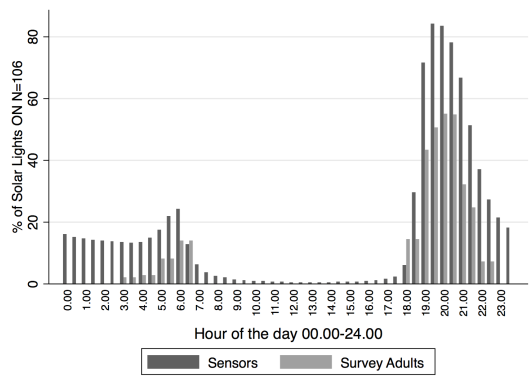
hour slot in the morning and half-hour slot in the evening (see Section 2 for more details) and then, whether they used any lighting source for each activity in each time slot. Second, we asked adults to estimate the global use of solar lights by the entire household on the previous day (see Section 2 for more details).

Using sensor data, we calculated the percentage of days that the light was used during that specific time slot for each sensor (across all days that the sensor worked), and then used this information to calculate the average across all sensors. By "used" we mean that the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot.

In Figure 4.1 and 4.2, we compare the Detailed Questions with the sensor measures. Overall, we see that the patterns of solar light usage over the course of the day match well. Note that we did not ask about use during the day and late at night, and hence these slots are, by design, empty. As expected, adults' reported use only reflects a fraction of total use. This is consistent with adults' answers to a separate question, to which over 70% responded that their school-aged children were the main users of the solar light. Figure 4.2 compares the combined answers of adults and pupils with sensor data. While the reports of usage in the evening hours seem to match the sensor data very well, some children seem to overreport use in the early morning hours.

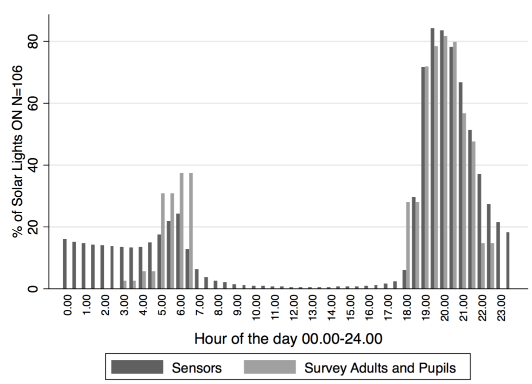
Comparing the averages of the Aggregated Question and the Detailed (or time diary) Questions, we observe that both measures provide similar results, which are both statistically indistinguishable from the sensor data for usage the day before the survey was done (Table 4.2, Column 5 and 7). The correlation coefficients in Table 4.5 (Column 5 and 7) suggest, however, that the Detailed Questions are less correlated with use than the Aggregated Question. In fact, their correlation coefficients are all below 0.2. This result might be surprising, given that asking individuals about each time slot separately (time diaries) is considered best practice to measure time use in the literature (Seymour et al., 2017). This finding might be explained by the fact that we did not ask for use during the entire 24 hours in the Detailed Questions, and that we only asked two household members and thus, do not capture usage by the entire household. Nevertheless, it is interesting that this much more lengthy and costly survey method did not correlate more with sensor data than simply asking for a global average of light use for the previous day.

Figure 4.1: Use Across the Day: Sensor vs. Survey Data (Adults Only)



Notes: For the sensor data, we classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors. In the survey we asked about activities and light use for each time-slot separately. Sample is restricted to sensors that worked until the end of the study and households where we have endline data for adults and pupils.

Figure 4.2: Use Across the Day: Sensor vs. Survey Data (Adults and Pupils)



Notes: For the sensor data, we classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors. In the survey we asked about activities and light use for each time-slot separately (in this graph we count the light as being used if either the pupil or the adult indicated that they used it). Sample is restricted to sensors that worked until the end of the study and households where we have endline data for adults and pupils.

4.5 Precision Gains with Sensor Data

While self-reported daily use of solar lights looks very similar to survey data on average (Table 4.1 and 4.2), the individual observations are not highly correlated (Table 4.5). In particular, correlation coefficients of the Detailed/Time Diary Questions (Table 4.5, Column 5 and 6) are very small, suggesting that the data are very noisy.

Table 4.5: Correlations Light Used (Hrs) per Day: Survey and Sensor Data

	(1) Sens (All) All	(2) Sens (All) Worked until End	(3) Sens (Aug) All	(4) Sens (Yest.)	(5) Surv (Detail)	(6) Surv (Detail) Adult	(7) Surv (Aggr.)
(1) Sens (All)	1.000	1.000	0.883	0.505	0.062	-0.035	0.257
	(220)	(147)	(220)	(147)	(215)	(215)	(161)
(2) Sens (All)	-	1.000	0.809	0.505	0.150	0.100	0.303
Worked until End		(147)	(147)	(147)	(147)	(147)	(125)
(3) Sens (Aug)	-	-	1.000	0.314	0.064	0.027	0.338
			(220)	(147)	(215)	(215)	(161)
(4) Sens (Yest.)	-	-	-	1.000	0.065	0.127	0.372
				(147)	(147)	(147)	(125)
(5) Surv (Detail)	-	-	-	-	1.000	0.334	0.409
					(215)	(215)	(161)
(6) Surv (Detail)	-	-	-	-	-	1.000	0.317
Adult						(215)	(161)
(7) Surv (Aggr.)	-	-	-	-	-	-	1.000
							(161)

Notes: Table shows correlations between variables in Rows and Columns. Number of observations are shown in brackets. Number of observations varies since we do not have sensor data for all sensors until the end of the study and we do not have all survey measures for all observations.

An advantage of sensor data is that it allows for more precise measurements, which enables researchers to detect smaller differences in use among sub-groups or to use smaller sample sizes than are necessary when using surveys to measure the impact of a new technology on behavior. For example, in our study we were interested in knowing whether households that received a free light used it less than households that paid for the light, in order to analyze the potential effectiveness of subsidies in increasing technology adoption. One might expect that households which purchase a solar light use it more, as households planning to use the light a lot are more likely to buy one (selection effect); simply having already paid for the solar light may also make households more likely to use it (sunk cost effect). Moreover, we were interested in whether poorer households use solar lights more. Unlike purchasing kerosene for lighting, the marginal cost of an additional hour of solar light is effectively zero. Therefore, we expect more credit-constrained households to use more light once they get access to a solar light. In Table 4.6, we show how the use of solar lights varies for different subgroups. We show survey answers for the households with sensors (N = 220), as well as survey answers from the entire sample.

Comparing household solar light usage for different types of households, we find that neither the survey nor the sensor data indicate a statistically significant difference in usage between those households which received a solar light for free and those who paid for it (Table 4.6, Columns 1-3). Our finding is in line with research on other products, such as bed nets, where the authors also did not find differences in net usage between households that paid for the nets and those that received them for free (Dupas & Cohen, 2010).

In contrast, a number of variables suggest that poorer households use the solar light more. For example, survey data and sensor data indicate that households with lower quality floors use the solar lights more (Table 4.6, Columns 4-6). This effect is more significant and the point estimates are larger when using sensor data. Moreover, the difference can only be detected with survey data if we use the entire dataset (Table 4.6, Column 6) and not only those households which also had a sensor (Table 4.6, Column 5). In line with this finding, food-insecure households also tend to use the solar light more (Table 4.6, Columns 7-9). In this case, the size of the coefficient from the survey responses from the entire dataset (Table 4.6, Column 9) and the sensor data (Table 4.6, Column 7) are very similar. There is also a negative correlation between asset ownership and solar light use (Table 4.6, Column 10). This negative correlation with wealth can only be detected with sensor data (Table 4.6, Columns 11 and 12). These results seem to confirm that more credit-constrained households tend to use the solar lights more.

Table 4.6: Correlates of Light Use

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Sensor (Hrs) First Month	Survey (Hrs) with Sensors	Survey (Hrs) All	Sensor (Hrs) First Month	Survey (Hrs) with Sensors	Survey (Hrs) All	Sensor (Hrs) First Month	Survey (Hrs) with Sensors	Survey (Hrs) All	Sensor (Hrs) First Months	Survey (Hrs) with Sensors	Survey (Hrs) All
Free Light	0.203	-0.246	-0.120									
	(0.348)	(0.362)	(0.175)									
Earth Floor				1.065***	0.364	0.354*						
				(0.325)	(0.462)	(0.202)						
Freq Cut Meal							0.246*	0.390**	0.243***			
							(0.136)	(0.162)	(0.087)			
Wealth Index										-0.177*	0.024	0.153
										(0.100)	(0.114)	(0.099)
Constant	3.712***	3.756***	3.505***	2.946***	3.261***	3.133***	3.685***	3.297***	3.250***	4.764***	3.394***	2.647***
	(0.314)	(0.306)	(0.134)	(0.288)	(0.426)	(0.178)	(0.160)	(0.185)	(0.102)	(0.508)	(0.545)	(0.466)
Observations	220	161	495	215	161	495	215	161	495	164	120	293
R-squared	0.002	0.003	0.001	0.038	0.004	0.004	0.016	0.041	0.020	0.019	0.000	0.018
Mean	3.864	3.566	3.434	3.864	3.566	3.434	3.864	3.566	3.434	3.864	3.566	3.434

5 Conclusion

There are a number of challenges with self-reported data on technology adoption, including social desirability bias, biases related to the fact that respondents feel observed, and accurate information recall. Sensors can provide more accurate, more precise data at a higher frequency than self-reported data. Hence, they can reduce the cost of analyzing behavioral change. In addition, they can help us understand biases and improve survey design, as we can test different survey techniques and compare responses to data collected with sensors. Sensor technology has the potential to transform how we measure human behavior and track the performance of policies and programs, however, there are still challenges to be overcome regarding the functionality of the technology over time. More field testing and training for social science researchers in charge of dealing with these new tools is needed (see Appendix C for more details).

While a number of studies have used sensors to measure the adoption of cookstoves (Wilson et al., 2016; Ramanathan et al., 2016; Thomas et al., 2013), this study is the first to use sensors to measure the adoption of solar lights on a large scale. Gandhi et al. (2016) used sensors to measure solar light adoption in only 37 households over less than two weeks. We were able to use sensors to collect information about solar light use in over 200 households, some of which purchased the solar light, while others received one for free.

We find that households use solar lights for around four hours per day on average and that fewer than 5% of households use the solar lights infrequently. Adoption of solar lights is much higher than what recent studies on cookstove adoption have found (Wilson et al., 2016). We also used sensor data to test what types of survey questions led to more accurate answers and whether differences between self-reported information and sensor data were particularly large for certain sub-groups.

A number of results seem especially relevant: first, as opposed to a number of papers on cookstoves (Wilson et al., 2016; Ramanathan et al., 2016; Thomas et al., 2013), and the small-scale study on solar lights (Gandhi et al., 2016), we do not find that households systematically overreport use of solar lights. However, in line with the findings of these studies, overreporting was more likely when the household used the solar light very little, which could be explained by social desirability bias. In addition, we also find that households which use the solar light a lot tend to underreport use, which, to our knowledge, has not been found before. As adoption of solar lights was nearly universal, we do not find evidence for systematic overreporting on average. In addition to the difference in adoption rates between cookstoves and solar lights, the nature of solar light usage is also very different from cookstove usage. Solar lights can be used by many household members throughout the entire day and in ways that are not visible to the respondent, whereas the use of cookstoves is typically reserved for a few household members and for a limited number of times at fixed times of day. These differences might explain why underreporting was more common in our case.

Second, while the reported hours of use per day are quite similar on average, answers from individual households correlated very little with the information we got from the sensors, suggesting that random errors are very large in survey data on technology use.

Third, we find that asking aggregated questions about use provides more accurate information than asking for each time slot separately (time diary). This result is surprising, given that time diaries are considered the gold standard in time-use data collection. However, there are still very few papers confirming the validity of this claim in developing countries (Seymour et al., 2017). The lack of correlation between the time diary survey responses and the sensor data could also be due to survey design issues, as we did not ask for every time slot throughout the day and we did not survey every household member.

Finally, we find that, as predicted by the Hawthorne effect, more frequent visits from surveyors in the beginning of the study did increase use initially. This difference disappeared once the visits stopped. Wilson et al. (2016) made a similar discovery when studying cookstoves.

We are not suggesting that sensors should replace surveys or that they should or can be used in every study of technology adoption. Many questions about adoption, and the use and impact of new technologies cannot be answered with sensors alone. In addition, sensors still require careful and time-intensive field testing, as frequent failures still pose challenges in many studies, including ours (Wilson et al., 2016). Our results, however, highlight how sensors can provide more accurate and precise information. This seems particularly relevant when social desirability is expected to be high. While it might be too early to draw general conclusions, a number of studies, including ours, suggest that the overreporting of use is mostly a problem when adoption is low, and hence that it is particularly important to receive an objective measurement in such cases. We also observe that while survey and sensor measurements were similar on average, they did not agree for individual households. Hence, sensors might be particularly relevant when researchers want to conduct sub-group analyses or use smaller sample sizes.

Finally, sensor data can help us better understand how to improve study and survey design, since they provide a credible benchmark to test different types of survey questions and interactions between surveyors and respondents.

6 Acknowledgements

We thank Kat Harrison, Professor Edward Miguel, Professor Lorenzo Casaburi, Dr. Nick Lam, and the seminar particpants at the Development Economics seminar at the University of California, Berkeley, the CSAE Conference, the PEGNet Conference, and the CEPE/CER-ETH Lunch Seminar Series for useful comments. We also thank Carol Nekesa, Lisa Schauss, Charles Amuku, Selina Obwora, Erick Bwire, and the entire field team from Innovations for Poverty Action Kenya for outstanding field work. This research was made possible with the support and cooperation of the Acumen Foundation, Bonsai Systems, Innovations for Poverty Action, SolarAid, SunnyMoney, and REMIT Kenya consulting. This work was supported by Google.org and the Development Economics Group and the NADEL Center for Development and Cooperation at ETH Zurich.

Bibliography

[1] Adair, J. G., Sharpe, D., Huynh, C.L., 1989. Hawthorne Control Procedures in Educational Experiments: A Reconsideration of Their Use and Effectiveness. Review of Educational Research, 59(2): 215–28.

[2] Angrist, J., Pischke J., 2010. The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. Journal of Economic Perspectives, 24 (2): 3-30.

[3] Arthi, V. S, Beegle, K., De Weerdt, J., Palacios-Lopez, A., 2016. Not your average job: measuring farm labor in Tanzania (English). Policy Research working paper; no. WPS 7773. Washington, D.C. : World Bank Group.

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

[5] Beaman, L., Magruder, J., Robinson, J., 2014. Minding small change among small firms in Kenya. Journal of Development Economics 2014, 108, 69.

[6] Beegle, K., Carletto, C., Himelein, K., 2012. Reliability of recall in agricultural data. Journal of Development Economics. Volume 98, Issue 1, 2012, Pages 34-41.

[7] Bertrand, M., Mullainathan, S., 2001. Do People Mean What They Say? Implications for Subjective Survey Data. American Economic Review, 91 (2): 67-72.

[8] Clasen, T., Fabini, D., Boisson, S., Taneja, J., Song, J., Aichinger, E., Bui, A., Dadashi, S., Schmidt, W.P., Burt, Z., Nelson, K.L., 2012. Making sanitation count: developing and testing a device for assessing latrine use in low-income settings. Environ Science Technology.

[9] Cohen, J., Dupas, P., 2010. Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment. The Quarterly Journal of Economics 125(1): 1-45. CRA (2013).

[10] Daniels, L., 2001. Testing alternative measures of microenterprise profits and net worth. J. Int. Dev., 13: 599-614. doi:10.1002/jid.781.

[11] Das, J., Hammer, J., Sanchez-Paramo, C., 2012. The impact of recall periods on reported morbidity and health seeking behavior. Journal of Development Economics 2012, 98, 76—88.

[12] Deaton, A., Grosh, M., 2000. Consumption. In: Grosh, Margaret, Glewwe, Paul (Eds.), De- signing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. World Bank, Washington, D.C.

[13] De Mel, S. McKenzie, D. J., Woodruff C., 2009. Measuring microenterprise profits: Must we ask how the sausage is made? Journal of Development Economics, Volume 88, Issue 1, 2009, Pages 19-31, ISSN 0304-3878

[14] Duflo, E., Glennerster, R., Kremer, M., 2007. Using randomization in development economics research: A toolkit. Handbook of development economics, 4, 3895-3962.

[15] Gandhi, A., Frey D., Lesniewski, V., 2016. Assessing Solar Lantern Usage in Uganda through Qualitative and Sensor-based Methods. IEE Global Humanitarian Technology Conference.

[16] Garn, J. V., Sclar, G. D., Freeman, M. C., Penakalapati, G., Alexander, K. T., Brooks, P., Rehfuess, E., Boisson, S., Medlicott, O., Clasen, T. F., 2017. The impact of sanitation interventions on latrine coverage and latrine use: A systematic review and meta-analysis. International Journal of Hygiene and Environmental Health, 220(2Part B), 329–340. http://doi.org/10.1016/j.ijheh.2016.10.

[17] Gautam, S., 2017. Quantifying Welfare Effects in the presence of Externalities: An Ex-ante Evaluation of a Sanitation Intervention. Working paper.

[18] Greenlight Planet, 2016. Product Sheet Sun King Eco. http://www.greenlightplanet.com (accessed: 13 October 2017).

[19] Grosh, M., Glewwe, P., 2000. Designing Household Survey Questionnaires for Developing Countries. World Bank Publications, The World Bank, number 25338.

[20] Hanna, R., Duflo, E., Greenstone, M., 2016. Up in Smoke: The Influence of Household Behavior on the LongRun Impact of Improved Cooking Stoves. American Economic Journal: Economic Policy, 8 (1): 80-114.

[21] Innovations for Poverty Action, 2016. Sensing Impacts: Remote Monitoring using Sensors. https://www.poverty-action.org/sites/default/files/publications/Goldilocks-Deep-Dive-Sensing-Impacts-Remote-Monitoring-using-Sensors_4.pdf (accessed: 12 August 2018).

[22] Leonard, K., Masatu, M.C., 2006. Outpatient process quality evaluation and the Hawthorne Effect. Social Science & Medicine, Elsevier, vol. 63(9), pages 2330-2340.

[23] Levitt, S.D., List, J.A., 2011. Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. American Economic Journal: Applied Economics, 3 (1): 224-38.

[24] Lighting Global, 2015. Product Verification Sheet. http://www.lightingglobal.org/products/glp-sunkingeco (accessed: 12 Sept 2017).

[25] Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355 (6325), 584-585.

[26] McCambridge, J., Witton, J., Elbourne, D.R., 2014. Systematic review of the Hawthorne effect: New concepts are needed to study research participation effects. Journal of Clinical Epidemiology, 67(3), 267–277. http://doi.org/10.1016/j.jclinepi.2013.08.015.

[27] Mills, E., 2003. Technical and Economic Performance Analysis of Kerosene Lamps and Alternative Approaches to Illumination in Developing Countries. Lawrence Berkeley National Laboratory.

[28] Nederhof, A.J., 1985. Methods of coping with social desirability bias: A review. Eur. J. Soc. Psychol., 15: 263–280. doi:10.1002/ejsp.2420150303.

[29] Pillarisetti, A., Allen, T., Ruiz-Mercado, I., Edwards, R., Chowdhury, Z., Garland, C., Hill, L.D., Johnson, M., Litton, C.D., Lam, N.L., Pennise, D., Smith, K.R., 2017. Small, Smart, Fast, and Cheap: Microchip-Based Sensors to Estimate Air Pollution Exposures in Rural Households. Sensors.

[30] Ramanathan T., Ramanathan N., Mohanty, J., Rehman, I., Graham, E., Ramanathan, V., 2016. Wireless sensors linked to climate financing for globally affordable clean cooking. Nature Climate Change volume 7, pages 44–47 doi:10.1038/nclimate314.

[31] Rom, A., Günther, I., 2019. Decreasing Emissions by Increasing Energy Access? Evidence from a Randomized Field Experiment on Off-Grid Solar. Working Paper.

[32] Ruiz-Mercado, I., Canuz, E., Walker, J.L., Smith, K.R., 2013. Quantitative metrics of stove adoption using Stove Use Monitors (SUMs). Biomass & Bioenergy, 57, 136–148. http://doi.org/10.1016/j.biombioe.2013.07.002.

[33] Serneels, P.M., Beegle, K.G., Dillon, A.S., 2016. Do returns to education depend on how and who you ask ? (English). Policy Research working paper; no. WPS 7747. Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/156941468868453090/Do-returns-to-education-depend-on-how-and-who-you-ask.

[34] Seymour, G., Malapit, H.J., Quisumbing, A.R., 2017. Measuring time use in development settings (English). Policy Research working paper; no. WPS 8147. Washington, D.C.

[35] Simons, G.F., Fennig, C.D., (Eds.). 2017. Ethnologue: Languages of Asia. sil International.

[36] Smits J., Günther, I., 2018. Do financial diaries affect financial outcomes? Evidence from a randomized experiment in Uganda. Development Engineering Volume 3, 2018, Pages 72-82.

[37] Thomas, E.A., Barstow, C.K., Rosa, G., Majorin, F., Clasen, T., 2013. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. Environ. Sci. Technol. 2013, 47, 13602−13610.

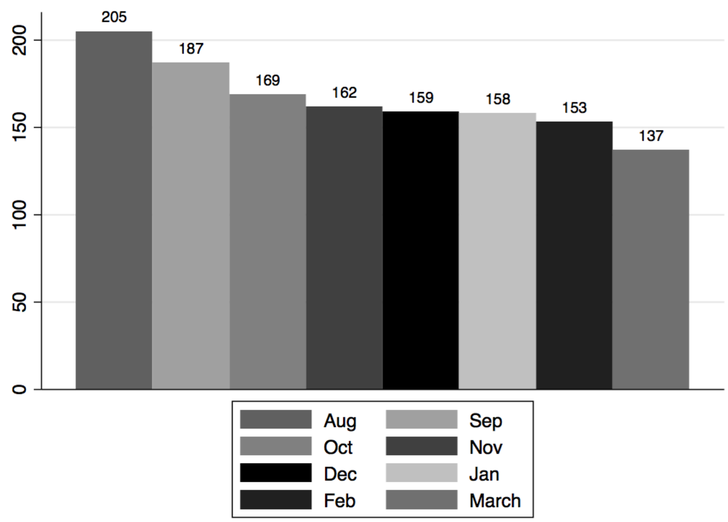
[38] Wilson, D., Coyle, J., Kirk, A., Rosa, J., Abbas, O., Adam, M., Gadgil, A., 2016. Measuring and Increasing Adoption Rates of Cookstoves in a Humanitarian Crisis. Environmental Science & Technology 50 (15), 8393-8399.

[39] Zwane, A., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D., Hornbeck, R., Gine, X., Duflo, E., Devoto, F., Crepon, B., Banerjee, A., 2011. Being Surveyed Can Change Later Behavior and Related Parameter Estimates. Proceedings of the National Academy of Sciences 108 (5): 1821-1826.

Appendix

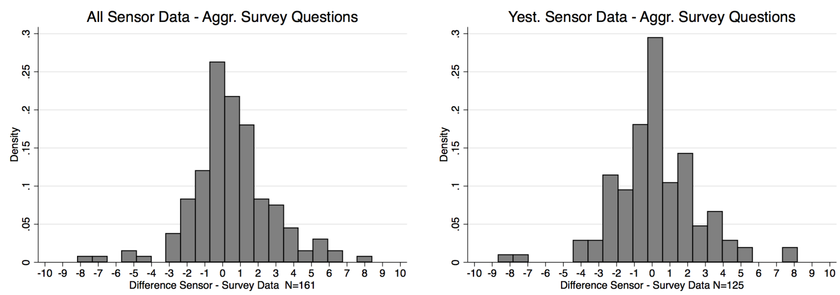
A. Figures

Figure A.1: Number of Working Sensors by the End of each Month

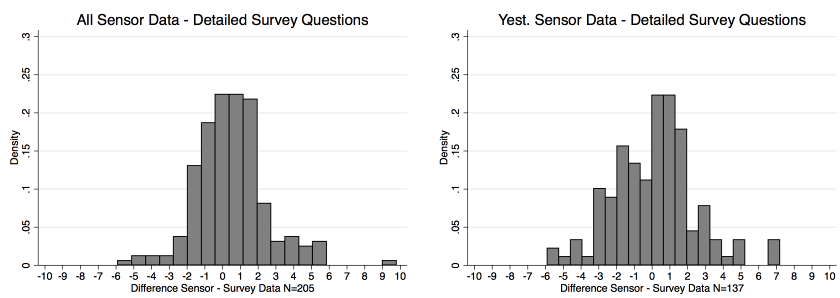


Notes: This graph shows the number of sensors that worked until the end of the indicated month.

Figure A.2: Under- and Overreporting of Use

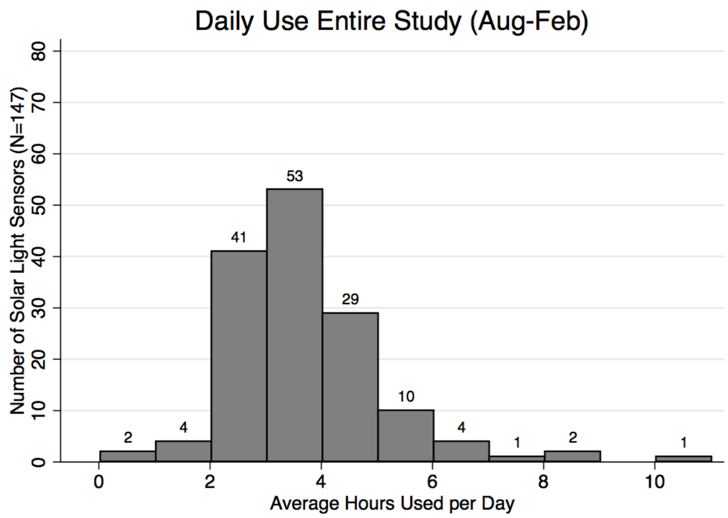


Notes: Graphs show density of difference between sensor data (where we include data from all sensors over the entire study period on the left, and only from the day before endline on the right) and survey data, where we asked the adult for a global estimate of the solar light use on the previous day (Exact question can be found in Appendix D).



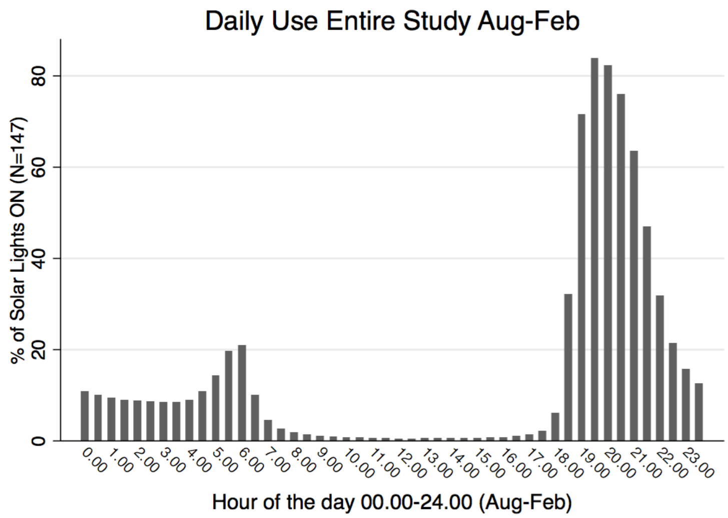
Notes: Graphs show density of difference between sensor data (where we include data from all sensors over the entire study period on the left, and only from the day before endline on the right) and survey data, where we asked the adults and pupils hour by hour about their activities and light use (time diary questions) and added the hours that either the adult or the pupil indicated that they used the solar light up (Exact question can be found in Appendix D).

Figure A.3: Average Hours Solar Lights are Used per Day (Aug-Feb)



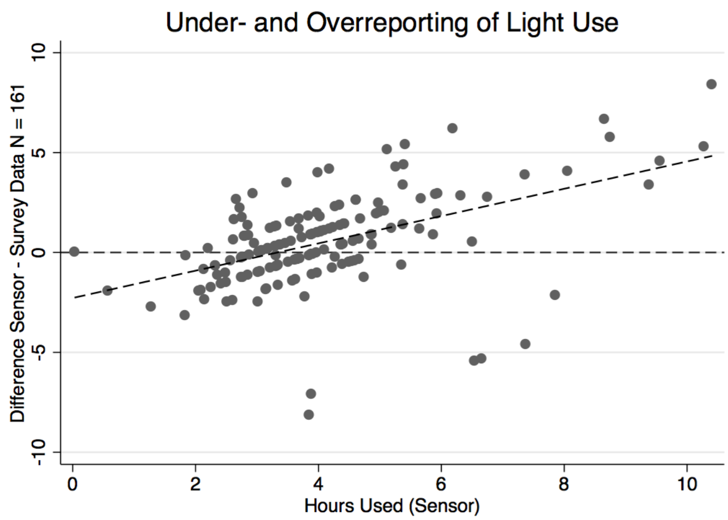
Notes: This graph shows sensor data about the average number of hours the solar lights were used per day during the entire study.

Figure A.4: Use Across the Day (Aug-Feb)



Notes: We classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors. Sample is restricted to sensors that worked until the end of the study.

Figure A.5: Under- and Overreporting of Use



Notes: This graph shows the correlation between the difference of sensor data and the survey data (Aggregated Question) and average hours used per day according to the sensor data. Positive values on the y axis indicate that respondents underreported use, while negative values suggest that they overreported use.

Figure A.6: Sun King Eco Solar Light

Figure A.7: Sun King Mobile Solar Light

B. Tables

Table B.1: Use Across Months

VARIABLES	(1) Sensor Hrs
September	0.183
	(0.254)
October	0.387
	(0.239)
November	0.384
	(0.250)
December	0.246
	(0.240)
January	0.242
	(0.233)
February	0.302
	(0.233)
March	0.325
	(0.230)
Observations	1,096
R-squared	0.004
Mean Sensor	4.053

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Left out group is August. We first calculated the average use per month per sensor. Mean use is across all months.

Table B.2: Use Across Weekdays

VARIABLES	(1) Sensor Hrs
Tuesday	0.062*
	(0.032)
Wednesday	0.016
	(0.029)
Thursday	0.016
	(0.030)
Friday	-0.008
	(0.034)
Saturday	-0.096**
	(0.046)
Sunday	-0.174***
	(0.039)
Observations	959
R-squared	0.002
Mean Use	4.022

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Left out group is Monday. We first calculated the average use per weekday per sensor. Mean use is across all weekdays.

Table B.3: Use Previous Month as Predictor for Survival

VARIABLES	(1) Sensor Hrs-Aug	(2) Sensor Hrs-Sept	(3) Sensor Hrs-Oct	(4) Sensor Hrs-Nov	(5) Sensor Hrs-Dec	(6) Sensor Hrs_Jan	(7) Sensor Hrs-Feb
Stopped Working in Sept	0.819						
	(0.538)						
Stopped Working in Oct		1.263**					
		(0.507)					
Stopped Working in Nov			-0.396				
			(0.268)				
Stopped Working in Dec				-2.237**			
				(0.971)			
Stopped Working in Jan					0.460***		
					(0.150)		
Stopped Working in Feb						-2.046**	
						(0.803)	
Stopped Working in March							0.031
							(0.482)
Constant	3.759***	3.828***	4.077***	4.095***	3.964***	4.041***	4.097***
	(0.148)	(0.160)	(0.146)	(0.161)	(0.150)	(0.143)	(0.154)
Observations	205	187	169	162	159	158	153
R-squared	0.013	0.031	0.002	0.022	0.000	0.040	0.000
N droped next Month	18	18	7	3	1	5	16

Notes: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1.

Table B.4: Comparing Survey Data and Sensor Data

VARIABLES	(1) Sensor Hrs/ All Data	(2) Sensor Hrs/ All Data	(3) Sensor Hrs/ First Month	(4) Sensor Hrs/ First Month	(5) Sensor Hrs/ Yesterday	(6) Sensor Hrs/ Yesterday
Surv (Aggr.)	0.208***		0.317***		0.412***	
	(0.062)		(0.070)		(0.093)	
Surv (Detail)		0.064		0.107		0.097
		(0.083)		(0.096)		(0.124)
Observations	161	162	161	162	125	126
R-squared	0.066	0.004	0.114	0.008	0.138	0.005
Mean Sensor	3.706	4.067	3.864	3.864	3.706	3.706

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. We include sensor data over the entire study period where we have both survey measures.

C. Lessons learned from using sensors to study technology adoption in low-income settings

First, it is critical to thoroughly pre-test sensor technology (both the sensor and the application to access the data) at a reasonably large scale in the field and to only roll out the study once all problems are solved. Often, engineering teams designing sensors are used to small sample sizes where technological challenges can be fixed along the way. It might make sense to agree in advance on a threshold of acceptable failure rates in the pilot as a commitment device. For example, we installed the sensors in a pre-existing product that was not designed to hold a sensor, thus, several sensors probably stopped working due to an imperfectly soldered connection between the sensor and the existing hardware, which also led to more light breakages. An additional challenge we had was that the application designed to access the data from the sensors initially did not work reliably and it took us quite some to determine the extent of the problem. In the meantime, our field officers had to return to the same households multiple times to ensure the data were collected, and since some of the sensors stopped working before the application was fully functioning, we lost a significant amount of data. Such issues could possibly be avoided by testing the sensors and associated technology extensively in the field and under a variety of realistic circumstances to determine vulnerabilities to contextual factors that are hard to recreate in

the lab.
Second, if the sensor is not constantly transmitting data to a central storage location throughout the study, we recommend doing a first round of sensor data collection immediately after installation and distribution (i.e., immediately after baseline) to guard against challenges linked to sensor attrition, which turned out to be a major problem in our study. Collecting data early not only ensures some data is collected from the maximum number of sensors, but can also help identify problems before they become widespread.
As a result of the two issues mentioned above, our third recommendation is to create a very detailed protocol on how to proceed if a sensor or the host technology stops working and, ideally, to include it in the pre-analysis plan. Both sensors and solar lights stopped working more often than we expected, and it was not possible to distinguish from the sensor data if the solar light broke because of the sensor or vice versa. It is therefore important to remember that both human error and technology failure are possible when building up a testing protocol. We suggest developing clear instructions about what to do if the analyzed technology or the sensor fails and to keep detailed information about replacements in order to easily account for these sensors in the analysis.
Fourth, researchers might underestimate the trade-offs between sample size and study duration on the one hand and data collection cost and management capacity on the other hand. While the data collection itself is very cheap, managing sensors and solving problems that affect many households over a long period of time is not. Researchers who plan to use sensors at a large scale should allocate considerable management and field staff time to manage them. In cases where the sensor technology has not been tested extensively in the field over long periods of time, we also recommend designing the research in such a way that most important questions can be answered even if there is a lot of sensor attrition over time. Our final recommendation is to take time to explain the sensor technology to partner organizations and the community. For example, we co-wrote a letter with the engineering team that developed the sensors explaining the functionality of the sensors to our partner organization. We also tested the acceptability of the sensors with a separate sample and developed a detailed script to explain the sensors to users. This script was written with guidance from our local partners, who are very familiar with the resident community. In addition, we provided respondents with our contact information in case of problems. We had no problems with regard to the acceptability of the sensors in the local community, but we imagine that this is highly context dependent.

D. Survey Questions

Aggregated Question:

- Do you own one or several lanterns? Options: yes/no
  - If yes: Does any of your solar lanterns still work? Options: yes/no
  - If yes: Yesterday, for how many hours did you use a solar lantern? Options: 0h-24h

Time Diary Questions:

- What did you do between XX:XX and XX:XX?
  Options:
  same as in previous time slot,
  at work (non-agricultural work)
  barber
  salon
  bathe
  dress
  brewing alcohol
  care for children / sick / elderly
  clean
  dust, sweep
  wash dishes or clothes
  ironing
  other household chores cook
  prepare food
  discuss activities of the next day with partner
  doctor/hospital
  visit
  eat
  farm work
  fetch water

firewood
fishing or hunting
funeral/wedding activities
help homework
herding animals/work with livestock
listen to radio
other religious activity (e.g., study, group)
participate in community activities/meetings/voluntary work
play sports
pray
prepare children for school
read book
repairs around/on home
rest
sewing/fixing
clothes shop for family
sleep
socialize with other household members
socialize with people outside of the household
spend time with spouse/partner
study/attend class
travel by bicycle
travel by foot
travel by motorized means
visit/ entertain friends
watch TV
Other

- What lighting source did you use for this activity, if any?
  Options:
  Electricity powered lighting
  Solar home system powered lighting
  Tin Lamp
  Kerosene lantern/Hurricane
  Fire Wood

Battery powered torch/lantern
Candle Solar lantern/solar torch
Pressurized Kerosene Lantern
Other rechargeable lantern
Cell phone light
No lighting used
Matchsticks
Other