

Productivity and Health: Physical Activity as a Measure of Effort

Oladele Akogun
Modibbo Adama University of Technology

Andrew Dillon
Northwestern University

Jed Friedman
World Bank

Ashesh Prasann
World Bank

Pieter Serneels
University of East Anglia

March 2020

Abstract: This paper examines the relationship between physical activity and individual productivity among agricultural workers paid on a piece-rate basis. In the context studied, physical activity has a clear correspondence with worker effort. Agricultural workers' physical activity is directly observed from accelerometer data and is robustly associated with their daily productivity. In addition the impact of a health intervention, which provides malaria testing and treatment, on physical activity and productivity indicates that the increased daily productivity of workers who are offered this program is explained by worker effort reallocation from low intensity to high intensity work within a fixed time period. This demonstrates, in settings when individual productivity is observed, that physical activity measures can help disentangle productivity effects due to effort. When productivity is unobserved, physical activity measures may proxy for individual productivity in physically demanding tasks. The challenges and limitations of physical activity measurement using accelerometers is discussed including their potential use for alternative contexts and the importance of field and data analysis protocols.

Keywords: labor productivity, physical activity, malaria, field experiment, accelerometers

JEL codes: I12, J22, J24, O12

Acknowledgements: Andrew Dillon (corresponding author) is a Clinical Associate Professor of Development Economics at Northwestern University, Evanston, IL USA; his email address is andrew.dillon@northwestern.edu. Oladele Akogun is a Professor of Public Health at Modibbo Adama University of Technology, Yola, Nigeria; his email address is akoguno@yahoo.com; Jed Friedman is a Senior Economist in the Development Research Group at the World Bank, Washington, DC USA; his email address is j.friedman@worldbank.org. Ashesh Prasann is an Agricultural Economist at the World Bank, Washington, DC USA; his email address is a.prasann@worldbank.org. Pieter Serneels is Professor of Economics at the University of East Anglia, Norwich UK; his email address is p.serneels@uea.ac.uk. The research for this article was financed by ESRC/DFID, the Global Poverty Research Lab (Northwestern University), Michigan State University, USDA National Institute of Food and Agriculture, University

of East Anglia and The World Bank. We would like to thank Adedoyin Adesina, Bashir Aliyou, Sani Njobdi, and the Nigerian Ministry of Health, Malaria Control Program for their collaboration and support in implementing this project. The plantation staff and sugarcane cutters welcomed us and patiently assisted us throughout the study, for which we are grateful. We thank the supervisors, enumerators, and health workers who helped implement this study. Tyler Box provided excellent research support. Eric Edmonds and Chris Udry provided helpful comments, as well as several anonymous referees and seminar participants at the University of Wisconsin-Madison, IPA Measurement meeting, the 2017 CSAE conference at Oxford University, and SEEDEC 2017 at the University of East Anglia. The study was ethically reviewed by the National Health Research Ethics Committee, Nigeria, Adamawa State Health Research Ethics Committee, and Michigan State University's Institutional Review Board. The paper was independently replicated before publication by the Global Poverty Research Lab. All data and dofiles are publicly available at <https://doi.org/10.7910/DVN/Q023HF>. A supplementary online appendix is available with this article at the World Bank Economic Review website.

1. Introduction

Productivity is central to analysis of economic activity, yet its empirical measurement remains a challenge. At the aggregate level (firm, sector, or economy wide) most analysis concentrates on production or value added (an output) per worker (input) as a base for productivity analysis. At the individual worker level, a directly observed measure of worker production (output) per unit of time (input) is often considered ideal.¹ Worker effort has received much attention as a determinant of productivity in theory (Becker 1977, Becker 1985, Gibbons 1987, Acemoglu and Pischke 1998, Lazear 2000, Fudenberg and Rayo 2019)² and from a policy perspective (World Development Report 2013)³, but much less in empirical analysis because self-reported effort may be prone to non-classical measurement error. Direct measures of physical activity offer an alternative to self-reporting when the output produced requires physical exertion. In this paper, after discussing the theoretical relationship between physical activity, effort, and productivity, a measure of physical activity using an accelerometer is validated against an incentive compatible measure of productivity (output per day in a piece rate wage setting). A second analysis demonstrates that treatment effect estimates with measures of physical activity provide a robust profile of effort substitution in response to a health intervention. As most jobs in developing countries are physical in nature⁴, this paper discusses opportunities for incorporating physical activity measures using accelerometers into empirical analysis, but also highlight current challenges and limitations of such an approach from both an analytic and implementation perspective.

¹ This type of measure, where individual output is directly observed, is often available in piece rate settings, where worker earnings provide a direct measure of a worker's output. Outside piece rate settings, earnings, while correlated with productivity, are not necessarily a good measure for individual worker productivity, as also discussed in Section 3.

² Gibbons (1987) and Lazear (2000), focusing on piece rate wage settings, consider workers as maximizing expected income net of the cost of effort each day, first by choosing whether to work or not on that day, and second by deciding how much effort to deliver when working. In the context of these models, physical activity can be thought of as a measure for overall effort, and thus a key determinant of productivity, at least for physically demanding tasks. These models build on Becker (1977, 1985) which we discuss below.

³ As motivation for the World Development Report, the authors state, "In almost every language there is a range of words related to jobs, each emphasizing a different angle. Some words hint at the nature of the specific activity being performed, evoking the skill or expertise that is required. Others refer to the volume of human inputs used in production, bringing images of *effort* and conveying a sense of physical exertion." (our italics).

⁴ See, for instance, the aforementioned World Development Report (2013).

To establish a theoretical motivation, Becker’s labor model (1977, 1985) that relates worker effort, productivity, and earnings is briefly summarized. In his model, a worker’s total earnings, Y , depend on a ‘package’ of time, t , and total effort, E . The return to effort, or wage w , is a function of effort per unit time (e) from each worker, where $e = E/t$. Workers choose levels of effort to maximize earnings where a worker’s total earnings are defined by the function, $Y = w(e)t$. The model provides a parameter of the effort intensity of work, σ , where the return to work depends on effort intensity along with human capital and other time-invariant worker characteristics. The earnings function can be specified with a Cobb-Douglas production function, $Y = \alpha e^\sigma t = \alpha E^\sigma t^{1-\sigma}$ with $\alpha = \beta h$ where h represents worker human capital and β reflects the returns to human capital conditional on effort and time, which can be thought of as a measure that captures time-invariant characteristics such as individual ability or genetic endowment.⁵ In many contexts, including agricultural labor, worker human capital, h , includes the worker’s health status, which varies over time. When the effort intensity of work, σ , is less than one equal effort will be used per hour worked as there are diminishing returns to additional effort.

Worker effort is typically unobserved in empirical studies because it is difficult to self-report, in part because of challenges with recall. Emergent technology presents new opportunities to gauge worker effort, by measuring physical activity directly through the use of accelerometers. This may generate a valid measure for worker effort especially for physically arduous tasks, which in low and middle income countries represent an important part of economic activity. Physical activity using accelerometers have been validated in multiple contexts including developing country contexts proximate to our study site. Assah et al. (2010) compare accelerometer based physical activity measurement compared to doubly-labelled water measures⁶ of physical activity. However, the validation and consistency of physical activity measures generated by accelerometer-based devices as a measure for worker productivity have not been well established.⁷

⁵ An extended version of the model could allow more explicitly for a role of capital and technology ($Y = AK^\beta \alpha e^\sigma t$). Here we focus on activities where technology is constant and the capital requirement is limited.

⁶ Doubly labelled water (DLW) method is considered the gold standard for the measurement of energy expenditure, as discussed in Section 2.

⁷ This application is more frequently used in public health and medical science. Bort-Roig et al. (2014) provides a systematic review of 26 peer reviewed, mostly small sample, studies in high income countries making use of smart phone technology to measure and analyse physical activity. None of them looks at a relationship with worker output or productivity. Table A.2 in Appendix summarises further studies making use of accelerometer.

This study is implemented at a large sugarcane plantation in rural Nigeria. Exploiting the unique feature of this setting, namely that workers are paid piece rate wages, personnel data provides a direct, observed measure of worker productivity, namely the number of sugarcane rods cut. The number of rods cut, as registered by the plantation's personnel records, linearly define workers' daily earnings. In a piece rate wage setting, an advantage over other labor contracts is that output per unit of time is both incentive compatible with effort maximization and observable.

Our empirical analysis consists of two steps. First, physical activity is validated as a substitute measure for unobserved individual worker effort by assessing its correlation with labor outcomes. The analysis demonstrates that physical activity is highly correlated with labor productivity (as well as labor supply). Each active hour is associated with increased earnings of 111 naira, or 14% of the earnings standard deviation. In a second step, our analysis estimates the validity of this measure of worker effort. Previous work, using data from an earlier harvest year, estimated that access to a workplace based malaria testing and treatment program increased worker earnings by approximately 10% over the weeks following the offer, both due to increases in productivity and labor supply (Dillon et al. forthcoming). The current study repeats the same intervention in a subsequent harvest season, re-estimates the impact on labor supply and productivity, and compares this with the impact on physical activity, the proposed alternative measure for productivity.⁸ The study estimates three sets of treatment effects: the Intent to Treat Effect (ITT), the Treatment Effect on the Treated (TOT) and the Treatment on the medically Untreated Effect (TmUT), which reflects the impact of receiving information about one's negative malaria status (without subsequent medical treatment).⁹ The estimated impacts on physical activity are found to be consistent with the estimated impacts on observed worker productivity and labor supply (and which are themselves consistent with earlier estimates from a previous round). They

⁸ One feature different from the previous year study is that switching between tasks, i.e. cane cutting and scrabbling, a less physically demanding task with a fixed wage, is no longer possible for workers in the year studied here. As a result, in this study, all workers devote all their time to cutting and none to scrabbling.

⁹ Making use of the randomized temporal order of treatment, the treatment effects compare a worker who received access to malaria testing and treatment to a counterfactual group of workers yet to receive treatment. The Intent to Treat Effect (ITT), summarizes the gains to all workers with the offer of the workplace health program. The Treatment Effect on the Treated (TOT) is estimated on the subset of malaria positive workers and reflects the impact of the bundled intervention of receiving news on positive malaria status and subsequent medical treatment. Treatment on the medically Untreated Effect (TmUT) is estimated on the malaria negative workers and reflects the impact of receiving information about one's negative malaria status (without subsequent medical treatment)

also shed further light on the productivity impacts of malaria testing and treatment. In the case of the ITT, the estimates indicate that workers who are offered this program reduce lower intensity activities in favor of moderate intensity activities, and reduce sedentary time. In the case of the TOT, workers who are tested positive and subsequently treated, reduce their daily average sedentary and ‘light’ physical activity time, while ‘fairly active’ and ‘very active’ physical activity levels increased. Total labor supply is largely unaffected, as the changes in physical activity are due to a reallocation of activity within working hours. Finally, for the Treatment Effect on the Medically Untreated (TmUT), which reflects the impact of the revelation of positive health information among workers who test malaria negative, the results are similar. Workers told they are malaria negative also exhibits an increase in the proportion of time spent in moderate activity levels as well exhibit an increase in productivity, as found in previous work.¹⁰

The direct measurement of physical activity also comes with challenges, both in terms of implementation and data analysis, which is discussed in detail. Key issues include subject non-compliance with study protocols as well as the loss of devices (a form of attrition). These issues necessitate consequential data management decisions. The analysis proceeds by bounding and trimming to assess the robustness of study results, particularly when treatment effects are the central analytic objective, in contrast with descriptive studies which often impute missing data to assess the distribution of physical activity for a given population. This highlights the importance of both careful and repeated piloting and, of course, planning for a large enough sample size.

Overall, the results indicate that physical activity can be an attractive measure for individual worker productivity in specific contexts. It allows a richer descriptive understanding of effort where this is unobserved (e.g. non-market activities), and enables empirical testing of theoretical predictions on the role of effort which has previously remained out of reach. In the conclusion potential future research topics of interest are explored. A simple cost benefit analysis demonstrates the affordability of implementing measurement of physical activity, in particular with the falling unit price of these devices, while protocols to improve compliance and reduce

¹⁰ Dillon et al. (forthcoming) who find similar results and detail how these workers switch to higher return, higher effort tasks, due to revised expectations over the perceived cost of effort. As physical activity directly proxies for effort, the effect of information on physical activity provides some evidence that the perceived cost of effort changes in response to the malaria testing and treatment program among these workers.

accelerometer loss are expected to further increase the benefits of using accelerometers in individual studies.

2. Physical Activity Measurement

Despite the theoretical attention to effort in labor economics, the measurement of physical activity has typically generated most interest in the fields of public health, medicine, and nutrition sciences. In public health and medical studies, whether assessing the health risks associated with certain lifestyles (e.g. sedentary), or estimating gains from health promotion activities, measurement efforts aim to capture the duration, frequency, intensity, or setting of physical activity (Bauman et al., 2006). Nutritionists tend to be interested in physical activity as a major component of individual energy needs. Calorie (food) deficits or surpluses are generally evaluated comparing estimated energy expenditure against a benchmark energy requirement. Physical activity is the most variable component of human daily energy expenditure and the second most important after the Basal Metabolic Rate (BMR) (FAO and WHO, 2001).

A number of tools are used for the measurement of physical activity that can be ranked according to their degree of practicality. Nutritionists and physiologists consider the doubly labeled water method (DLW)¹¹ as the ‘gold standard’ to measure energy expenditure in free-living conditions (Speakman, 1997; FAO and WHO, 2001), but this is demanding and expensive to implement. As an alternative to this objective measure, subjects are sometimes asked to keep diaries or logs of physical activity. While widely used in medical and nutrition studies, these methods can be cumbersome and costly to implement with large samples. More common are individual self-reports of physical activity. Instruments that have undergone international validation, including developing country settings, include the International Physical Activity Questionnaire (IPAQ) in its short- and long-form versions (Craig et al., 2003), the WHO-led Global Physical Activity Questionnaire (GPAQ), and a Sub-Saharan Africa Activity Questionnaire (Sobngwi et al., 2001). The instruments are designed to measure activities in different domains (work, travel to and from places, and recreational activities) and the intensity of these activities

¹¹ Doubly labelled water (DLW) method measures the average daily metabolic rate over a period of time from repeated blood, saliva or urine samples, after administering a dose of DLW.

(e.g. low, moderate or high physical activity) rather than precisely measuring energy expenditures. While validation studies for these survey instruments exist (e.g. Craig et al., 2003), they remain prone to the types of biases that may be encountered in any subjective self-report based on survey recall, and their accuracy has been called into question as it appears to often produce results that are in contradiction with objective measures, and specifically with data obtained from accelerometers (Corder and Van Slujis, 2010; Nelson et al 2019; Chaput 2019).

In the last decade, activity trackers have been developed as instruments that have the potential to provide sufficiently accurate estimates of physical activity and energy consumption at the individual level. Accelerometers are possibly the most commonly used type of device, but other types of sensors are also being used or tested (Baranowski et al., 2012; Storm et al., 2015).¹² While much work still focuses on laboratory-based studies, the use of free living protocols are increasing. Applied work making use of accelerometers has investigated the relationship between physical activity and the built environment (Sallis et al., 2016) or contrasted patterns of physical activity across countries or populations (Salvo et al., 2015). One particular challenge with this method is study subject adherence to the protocols necessary for accurate measurement – the accelerometers must be worn in a consistent fashion during waking hours for several consecutive days. Given these challenges, field studies typically drop observations, with between 9% and up to 40% of individual-day data to be found invalid, as discussed in more detail later.¹³ Despite these adherence challenges that typically lead to data censoring, affecting sample size and potentially sample composition, wearable technologies including accelerometers hold a great deal of promise due to their accuracy in real-time measurement, which is continuously advancing.

The validation and consistency of physical activity measures generated by accelerometer-based devices as a measure of work productivity have yet to be well established.¹⁴ Physical activity monitoring is not inherently advantageous for all types of jobs. Examples where physical activity

¹² Most recently, a review of accelerometer studies with sample sizes greater than 400 identified accelerometry data for more than 275,000 individuals, from 76 studies and 36 countries (Wijndaele et al, 2015). Efforts have also started to develop internationally agreed protocols for accelerometer data collection, sharing, and dissemination (Cain and Geremia, 2012; Wijndaele et al, 2015).

¹³ Trimming criteria is discussed in the next section. Appendix 2 provides an overview of trimming rules used by 13 recent studies using accelerometers.

¹⁴ Among rural populations in sub-Saharan Africa, physical activity measures have been validated against assessments of energy expenditure in at least one small sample study (Zanello et al. 2017).

monitoring might be the most analytically interesting include (i) agricultural tasks, like planting, sowing, harvesting, ploughing, herding, where technology is constant and capital requirement is limited; (ii) non-agricultural tasks with similar characteristics: like mining, quarrying, construction, and specific urban services (such as portage); (iii) non-market tasks, such as female and child household work, including for instance water carrying and wood gathering. In low and middle income countries, these types of tasks represent an important part of economic activity.¹⁵

Turning to measures of individual worker productivity, Table 1 notes examples of measures of worker productivity for specific activities and sectors used in studies published in primary publications in economics. Typically, these studies use a measure that is attributable to individual workers, easy to observe, comparable over time and across workers (conducting similar activity), and related to output for which the quality is easily monitored. The measures allow analysis of individual worker productivity within a firm, and are key when studying the causal impact of factors like the organization of work (Bandiera et al 2010, Mas and Moretti 2009), incentives and contract design (Lazear 2000, Shearer 2004, Ferraz and Finan 2009, Kaur et al 2015), training (De Grip and Sauerman 2012), or health (Zivin and Neidell 2012, Dillon et al. forthcoming), on worker productivity. Outside piece rate settings, earnings, while correlated with productivity, are not necessarily a good measure for individual worker productivity. Various institutional factors such as unions and norms, job attributes and their desirability, as well as worker characteristics have been identified as co-determinants of wages.¹⁶

3. Study Design, Context and Data

The study site is a single large (5,700 hectares) sugarcane plantation in rural Nigeria. The plantation employs 680 sugarcane cutters who work for the entire harvest season that stretches

¹⁵ In the case of Nigeria, agriculture represents close to one fourth of the economy (World Development Indicators 2016), employing approximately one third of the labor force.

¹⁶ Analysis of productivity across firms tends to focus on average worker productivity by firm, typically constructed as value added divided by number of workers, and observes large dispersion of average worker productivity across business units (see Bartelsman and Doms, 2000), across and within sectors of activity (see Gollin and Udry 2017). See Tybout (2000), Bartelsman, Haltiwanger and Scarpetta (2008), Asker et al. (2014), Syverson (2011), and Bartelsman et al. (2013) for examples of cross country micro datasets on firm productivity.

from mid-November to April (in this case the study year 2012-13) and are paid a piece rate wage. While there are other activities on the plantation, including a sugar processing facility, this study focuses solely on the sugarcane cutter labor force.

Workers are hired from local villages surrounding the plantation and are transported daily to and from the assigned work site. The cane-cutters are organized into eight work groups and each group is managed by a supervisor. Every day the supervisor and his cutters are allocated to a set of starting fields in the plantation and additional fields when the starting fields are finished. Sugarcane cutters do not work in teams to complete the rows of cane but rather work individually along a row until finished and are then assigned by the supervisor to another row to harvest. Rows of cane are typically of uniform density due to mechanized planting and the irrigated nature of sugarcane that requires fields to be encompassed with water canals. The worker's day is standardized with plantation trucks collecting workers from their village and delivering them to the field sites to be cut each day, and transporting them back to their village at the end of the day. This standardized work day ensures that for each cane cutter the number of hours of work is fixed across all workers in a work group and alleviates concerns about trade-offs between work hours and effort.

Cane cutters are paid a piece rate of 2.04 naira for every measured "rod" of cane cut, where a "rod" (approximately two meters in length) is a physical standard carried by every work group supervisor. At the end of each day, the worker's output for that day is entered on his personal 'blue card' and is signed off by both the supervisor and worker. The plantation thus keeps records of the daily output (quantity cut), the days worked, and the total earnings for each worker. Workers are paid monthly and they often keep track of their daily output by maintaining their own separate ledger. Disagreement between cutters and management over output amounts are rare. The work tends to be lucrative and an average day of cane cutting pays 1,156 naira, or approximately 7 USD. This daily wage is substantially higher than most local alternatives. With the poverty rate in the surrounding Nigerian state at 74.3% (measured at \$1 USD per day) (Nigeria National Bureau of Statistics, 2012), sugarcane cutter positions are in high demand in the local communities.

The study design is described by first providing details of the field protocol, as implementing fieldwork with accelerometers is relatively novel in empirical applications in

economics. In a second and third subsection, we describe the worker sampling and balancing tests conducted to validate the study design. In a final subsection, we document the accelerometer assignment protocol, and discuss daily compliance and attrition, drawing insights from our own as well as other studies.

Field Protocol

The objective of this field protocol was to induce exogenous variation in the timing of testing and treating the entire worker population for clinically diagnosed malaria. The intervention is described in detail in Appendix 1, and is identical to the one implemented in an earlier year and studied in Dillon et al. (forthcoming). The intervention tests workers for malaria using a blood sample drawn from each worker and sent to a microscopy laboratory, to inspect the number of malaria parasites per viewing field. Adopting the local benchmark standard to decide malaria positivity, the results were then returned to the medical team who communicated this to the worker within two days of the blood draw. Workers who tested positive, were given Artemisinin Combination Therapy treatment consisting of two pills each day for three days. The testing – and medical treatment if positive – were implemented in random order over time, and at the end of an eight-week period all cane-cutting workers employed by the plantation were tested (and treated if positive). The phased in design with random order of treatment over time allows for identification of treatment effects on the worker’s observed physical activity, labor supply and productivity (cf. Athey and Imbens (2018); Miguel and Kremer (2004)).

While labor outcomes were observed for the entire cane cutting population of 680 workers during an eight-week study period from February to April 2013, at the peak of the harvest season, the analysis in this paper is based on a subsample of workers who were equipped with activity trackers during the study period. A random sub-sample of 83 sugarcane cutters was assigned an activity tracker, as described in detail below. Each tracker was labeled with a unique code for the dedicated use of the same worker over the study period. Cane cutters clipped these activity trackers, approximately the same size and weight as a USB flash drive, to their pant hem, pocket, or belt. At the end of every work week, the data recorded by these trackers were synced to a notebook computer at the field survey office and the trackers were simultaneously charged. This

process was carefully managed and workers received back the same tracker by the start of the new workweek. Detailed checking during the fieldwork reassured that trackers were operational, although some were lost by the workers, as discussed below.¹⁷

A survey team of enumerators and health workers employed by the project collected information on worker and health characteristics from all workers. Worker information included employment history, demographic information, residential location, and household characteristics. At the end of the fieldwork the activity information collected from the trackers was linked to plantation personnel data, the worker and health information from the survey and the malaria test, producing a data set that enables the envisaged two-part analysis carried out below. In the first part we analyze the relationship of physical activity with labor supply and productivity, respectively. The second part estimates the short-run impact of malaria testing and treatment on direct measures of physical activity, as well as labor supply and earnings.

Worker sampling

The selection of workers for this study was undertaken in two stages, the second of which is most relevant to the analysis in this paper. The first stage concerns the design of the health intervention, particularly the timing of malaria testing and treatment, which was randomized over the eight-week study period among the entire worker population, stratified by work group, following the exact same procedure as the original study. The second selection step determines the subsample of workers who receive an accelerometer to wear throughout the study period, generating the sample of interest for this paper.

Selected workers for the accelerometer sample totaled 83 and comprise two sampled subgroups. Focusing on the workers who were scheduled to be tested and treated in week 1, 2, 3 and 4 according to the main sampling plan¹⁸, three to 4 workers were randomly selected for each

¹⁷ The protocol was developed after intensive pilots with a sample of other sugarcane workers off plantation in January 2013. These pilots tested practical modalities such as data registration, battery, charging and syncing time, device functioning in high temperature and water rich environment, and exploring device location to minimize loss (arm, around the neck, belt).

¹⁸ The general sampling plan covered the entire 8 weeks during which the program ran. This plan generated a detailed schedule of the identified workers that would be tested and treated every expected day. We focused on workers in

expected working day, leading to 58 workers for this first subgroup. The remaining 25 accelerometers were given to the first 25 workers who tested positive for malaria, who constitute the second subgroup.¹⁹ Malaria positive workers were oversampled because of the ex-ante concern of low power to detect the effects of malaria infection, given the positivity rate (36%) in the previous study (Dillon et al. forthcoming). As a consequence of these selection rules, a malaria positive worker was approximately 50% more likely to be selected into the study sample.²⁰ All workers who were allocated an accelerometer received the device in weeks one or two, and wore the same device until the end of the field period (or until it was lost). Workers could not change or transfer trackers during the study period as each tracker was uniquely identified.

Balancing tests

The successful randomization of the tracked subsamples is confirmed by the similarities in characteristics between the sample of tracked and non-tracked workers. Workers' characteristics including age, BMI and hemoglobin score, as well as human capital indicators including worker literacy, numeracy, school attendance, level of school completed, and plantation experience, as well as household characteristics like household composition, including number of spouses, children, total household size, and a household asset index²¹ are well balanced between the tracked and non-tracked samples (see Table S1.1 in on line Appendix). In the tracked sample, malaria positive cases were significantly higher as intended given the selection strategy. Just as in the

week 1 to 4 for the allocation of accelerometers to ensure that these workers wear the device from as early as possible and are spread across the workforce.

¹⁹ For instance, if two workers are tested and found positive on the first day, and had not already been selected for an accelerometer, they received an accelerometer; if another five workers are found positive on the second day, they received an accelerometer.

²⁰ Sample weights were calculated to reweight to the worker population based on the selection probabilities and the malaria positivity rate. Our regression results are virtually the same when the sample weights are omitted, and we proceed the analysis without sample weights.

²¹ The asset index reflects household expenditures which are not measured but rather predicted using the method suggested by Grosh and Baker (1995) and Ahmed and Bouis (2002). The questionnaire included questions on asset ownership drawn from the Nigerian Living Standard Survey 2009, a nationally representative survey, conducted by the National Bureau of Statistics (NBS), which collects detailed data on household consumption and expenditures. We run the weighted regression $Exp_i = \sum_{a=1}^p (\alpha^a D_i^a + u_i)$ on the NLSS 2010 data to obtain estimates of $\hat{\alpha}^a$, the coefficient for each asset, which we then use to predict Exp_i for our own sample. Where D_i^a represents a dummy variable indicating whether asset a is present in the household. The regression uses population weights as calculated by the NBS. Since the estimates of the coefficients are relatively sensitive to outliers, we exclude the richest 10% of households in our weighted regression on the NLSS 2010 data.

original study, the identification of the relationship between health and productivity (and in this case physical activity) relies on the successful randomization of the phased in design of the malaria testing and treatment program. To assess balance over time we test whether treatment and comparison workers who wear the tracking device but are tested for malaria at different points in the calendar differ across a range of worker and household characteristics, including work experience, education, household size, asset index, hemoglobin and body mass index. When considering the standardized mean difference in characteristics for each pairwise comparison, none of these standardized mean-differences exceed the 25 percent threshold suggested by Imbens and Wooldridge (2009) to signal possible sample imbalance (see Table S1.2 in online appendix). Differences in the mean characteristics are also not significant at conventional levels as determined by traditional t-tests of equality.

Accelerometer Assignment Protocol

Workers were assigned a Fitbit One activity tracker²² which was chosen after market research and consultation with experts, who largely emphasized reliability, immediate availability and user friendliness in the field.²³ The belt-worn accelerometer is a three-dimensional accelerometer which measures activity duration and intensity in minutes. The accelerometer reports time spent in four levels of activities: Sedentary Minutes, and Lightly, Fairly and Very Active Minutes. The proprietary algorithm assigns minute intervals to any activity if the assessed activity is more strenuous (in terms of estimated caloric expenditure) than a slow-paced walk. These assignment categories are taken as given in the analysis as the researchers do not determine the classification of activity levels to categories. The caloric expenditure threshold is normed to the BMI of the worker, with the three partitions of activity representing increasingly greater energy

²² Device specifications and data collected for this three dimensional activity tracker can be found at www.fitbit.com/au/one.

²³ Ease of use was an important consideration, and even after choosing the most user-friendly option at the time, a number of trackers were lost. One alternative approach would have been to develop one's own accelerometer. While this has the advantage over commercial products in that the underlying algorithms are known and available, discussions with experts in this field who are and have been developing their own devices indicated that the development is time consuming, costly, and beyond the scope of our study. Chen et al. (2008) provide an early review of accelerometer characteristics.

expenditures. The device tracks worker physical activity continuously throughout the day and cannot be manually turned off by the worker.

Eighty-three workers were allocated activity trackers at the beginning of the study, but 25 accelerometers were lost over the eight-week study period, resulting in 58 devices being returned at the end of the study.²⁴ This analysis utilizes the daily data generated by all trackers, until the point that they were lost. Figure 1 presents the frequency of tracker loss by the number of days until the tracker was lost by the worker. Tracker loss does not seem to occur at any particular point in time, though loss increases in the first week after assignment and during weeks 5 and 6 of the study when the plantation was unexpectedly closed and workers were monitored less closely (see below).

Because workers who lost their accelerometer may have different characteristics, thus possibly resulting in biased estimates, we assess the correlates of the probability of tracker loss.²⁵ When regressing an accelerometer loss indicator on worker characteristics and health status, no statistically significant relationships was found for any of the observed household, individual or other health characteristics, including age, experience, schooling, household composition, assets, malaria test result status, BMI, and hemoglobin (See Table S1.3 in on line Appendix). All estimated coefficients are also very close to zero in magnitude.

Daily Compliance

Studies using accelerometers vary in the magnitude of respondent non-compliance, as demonstrated by the overview of studies in Appendix 2, and Bort-Roig et al (2014). The challenge of non-compliance is addressed in the analysis using the following criteria:

²⁴ We investigated and carried out qualitative interviews to assess whether these devices were stolen, but there is no reason to believe so. Given the relatively high earnings from cane cutting, workers are keen to keep a good reputation with the plantation, and they were fully aware that the device was not useful without the complementary workstation and log in password to download and read the information, which they did not have access to.

²⁵ Selective loss may introduce a bias in later estimation results. For instance, if workers who had more physical capacity were more likely to lose their tracker due to greater physical activity, then this would likely bias activity measures downwards. If workers who were more educated or wealthier were more careful with their tracker, then this differential probability of tracker loss might also potentially bias the estimates.

1) An observation day is deleted if there was a reporting error in the accelerometer data identified because it did not sum to 24 hours in that day (for example, due to battery failure). Regardless of whether a worker wears the device for 24 hours, the device should register activity – both active and inactive - for that entire time period. The 24 hours total activity threshold therefore provides a good criterion to assess data accuracy.²⁶

2) A cut-off for a valid day of data is stipulated by establishing thresholds on total time devoted to activity (of any level). A day is deemed invalid if the activity tracker measures active time of less than three hours on work days or less than 1.5 hours on non-work days. This second rule aims to address non-adherence by dropping observations of very low activity. The fundamental reason to drop these observations stems from the fact that we are not able to distinguish between inactivity and non-compliance. The threshold used, 3 hours of registered activity for working days, and 1.5 hours for non-working days, is informed by the data. The distribution demonstrates a natural break in the data around this threshold. Figure 2 presents the histograms of sedentary time per day for both (plantation) work days and non-work days. The degree of non-adherence is readily apparent in the right tails of both histograms, and especially that for non-work days.²⁷

3) For working days the remaining days are trimmed at the 1st and 99th percentile of the rods cut per active hour distribution in order to further reduce the influence of likely measurement error in both daily productivity and daily activity measures. This follows a common approach to symmetrically dropping the top and bottom percentiles in order to reduce the role of outliers that result from measurement error. There is currently no universal standard protocol in the use of trimming for accelerometer data. Appendix 2

²⁶ In medical jargon we follow a ‘complete case’ approach, carrying out analysis on cases with complete data, rather than an ‘available case’ approach, which also considers partially complete cases, or ‘imputation’, where missing values are substituted by predicted values based on modelling of the distribution assuming underlying patterns for its parameters.

²⁷ Any classification threshold for a valid day will include some proportion of false positives and false negatives. We believe the chosen thresholds are conservative in that they likely include a relatively larger number of invalid days. However, the presented results are robust to other cut-offs in either direction at 30 minute increments, as well as wider increment bands shown in results to follow.

provides an overview of recent live free studies using accelerometer (as opposed to lab studies), and demonstrate the wide variation of data trimming rules used, depending on the purpose of the study, with between 9% and 42% of the total observations dropped from the analysis.

Appendix Table S1.4 presents summary statistics for the tracked subsample including the physical activity outcome variables sedentary, light intensity, medium intensity, and high intensity minutes of activity (labelled 'lightly', 'fairly' and 'very active', respectively), both for the full sample and after applying the above three step protocol (which is termed, going forward, Trim A). Comparing the first and second panels of Appendix Table S1.4., which describe the full sample and Trim A sample respectively, indicates that roughly 1/3 of worker-days do not meet the criteria of a valid day of observation under the above rules. Adherence to protocols is relatively worse on non-work-days as a greater proportion of worker-day observations do not meet the trimming thresholds and are therefore excluded.

Table S1.4. shows that, among all worker-days registered using the activity monitors, workers spent on an average day 18.8 hours in sedentary activity (including sleep), 2.4 hours in light intensity activity, 2.1 hours in fairly intensity activity and 0.4 hours in high intensity activity. These activity levels increase with the trimmed data (as expected) to 6.8 hours of physical activity per day, again mostly in the light or fairly active categories. Workers are certainly more active on the days when they cut cane: 7.6 hours in contrast with 5.8 hours of activity when off the plantation. The table also reports the percent of days worked among the sugarcane cutters and their conditional daily earnings, 1,154 naira, or approximately 7 USD.

Work Disruption

During the study period, an unexpected mechanical failure at the processing factory resulted in a temporary stoppage of harvest activity during study weeks 5 and 6. The nature of the production process together with the daily working order of processing machinery partially determine harvest activity on any given workday at the plantation. Since sugarcane has an optimal time to be harvested (in order to maximize sugar content), and needs to be processed on the same

workday by the sugar factory on the plantation, past planting schedule and current factory activity affect whether sugarcane cutters work on a given day.

Since all workers who had been allocated an accelerometer were already treated by this time, this does not affect the identification strategy, and still allows us to estimate treatment impacts on worker physical activity, but it does reduce the power of the study from our initial expectations, as we do not observe as many work days in the analysis as planned in the study design. It also limits the observation period over which we can estimate the impacts of malaria testing and treatment.

4. Econometric Strategy

Relationship between physical activity and labor productivity

In a first step to validate the activity measure as a proxy for productivity, the relationship between physical activity and the worker's daily output is validated by estimating the conditional association between labor outcomes and each of the measured levels of physical activity (sedentary, light, medium and high intensity minutes):

$$L_{idg} = \alpha_i + \sum_j \beta_j A_{jid} + F_{gw} + \varepsilon_{idg} \quad (1)$$

In this equation, L is daily earnings, which provide a measure of productivity at the intensive margin (as daily earnings is a linear transformation of rods cut, i.e. a direct measure of output). The analysis also considers L as labor supply (days worked), which can be thought of as a measure of productivity at the extensive margin, and as earnings unconditional on the daily work decision. A_{jid} is the activity measure for intensity j - i.e. number of minutes or hours by the intensity level of the activity (lightly active, fairly active, very active) observed for worker i in work group g on day d (note that these minutes along with sedentary minutes sum to the total minutes in a 24-hour period). The ratio of time in each of the intensity categories relative to the total time spent in activity is considered as a separate outcome to assess the distributional change of time across intensity levels. The specification includes a fixed effect, F_{gw} , for each workgroup and calendar week, w , combination to control for spatial variation across workgroups, and temporal variation within workgroups, in earnings opportunities related to work schedule and fecundity of assigned

fields.²⁸ Standard errors are clustered at the worker level.²⁹ These regression estimates serve as a validation test for the use of physical activity as a measure for labor productivity and labor supply in our study context.

Treatment effect estimation

In the second phase of analysis, the direct impact of malaria testing and treatment on physical activity and on labor productivity and earnings (as well as labor supply) is estimated. The primary econometric specification estimates the intention-to-treat effect, which compares daily labor outcomes over some observation period, t , for those workers who were tested at time t -, a period before the observation period t , with the labor outcomes for workers who are tested at $t+$, after the observation period t . The sets of workers assessed at t - and $t+$ are denoted as W_{t-} and W_{t+} . The difference in outcomes over period t represents the combined effect of testing and treating for malaria, as it compares the output of a randomly selected subsample of workers who are tested with a randomly selected subsample of worker yet to be tested. To address the potential non-random placement of workers across workgroups as well as variations in earning potential across groups and within groups over time, workgroup by week fixed effects, F_{gw} , are included in the specification. Specifically, the analysis estimates:

$$L_{igt} = \alpha + \beta T_{igt-} + F_{gw} + \varepsilon_{it}, \forall i \in W_{t-} \cup W_{t+} \quad (2)$$

where L_{igt} measures the three labor outcomes of interest: labor supply, daily productivity, and physical activity for worker i in work group g , and ε_{it} is the worker specific error term. The treatment indicator, T_{igt-} , takes a value of one if the worker receives the malaria test and, possibly, medication at time t -. As above, standard errors are clustered at the worker level.

²⁸ Alternative approaches include no fixed effect or a worker fixed effect. Including a worker fixed effect will capture time invariant worker characteristics, such as worker skill and ability that help translate physical activity and effort into output. If we recall, Becker's model discussed in the introduction motivates the effect of unobservable individual factors arising through either the transformation of human capital, β_i or human capital itself, h_i when specifying the relationship between effort and productivity. Results with both approaches are presented in Appendix Tables S1.

²⁹ As the worker is the unit of treatment – workers, stratified by workgroup, are offered the malaria testing and treatment program in a temporally randomized order – standard errors are clustered at this level (see Abadie et al., 2017).

Given the limited weeks of valid data in the study period (due to the factory break down as discussed above), the observation period t is fixed as a seven-day reference period in order to maximize the number of worker-days in the analysis. This implies that all workers assessed on day t are assigned to the treated group whereas workers assessed on day $t+7$ are assigned to the counterfactual comparison group. The observation period for the outcomes of interest extend from day t to day $t+7$. Workers are resampled throughout the study period, assigned to treatment or counterfactual groups depending on their treatment status at time t as in Dillon et al. (forthcoming).

The set of workers assessed, W , is also determined through clinical testing to be either malaria positive, P or negative, N . Disaggregating the sample of tracked workers between positive and negative malaria cases will obtain separate estimates of the impact on malaria positives and negatives – i.e., the TOT and TmUT effects as described earlier.

To estimate the TOT, Equation (4) is re-estimated but now for the subset of workers P who have tested positive, as given in Equation (5)³⁰:

$$L_{igt} = \alpha + \beta T_{igt-} + F_{gw} + \varepsilon_{it}, \forall i \in P_{t-} \cup P_{t+} \quad (3)$$

as before, L_{igt} reflects the labor outcomes of interest: productivity, and physical activity, as well as labor supply. The TOT reflects the combined effect of receiving an illness diagnosis and medical treatment for such a diagnosis.

Finally, the analysis considers the subgroup of medically untreated, who do not receive any medicine (as they are malaria “free”) but do receive information on their health status. As in previous work, this allows the estimation of a possible ‘good news’ effect³¹ for the subset of workers N following equation (6):

$$L_{igt} = \alpha + \beta T_{igt-} + F_{gw} + \varepsilon_{it}, \forall i \in N_{t-} \cup N_{t+} \quad (4)$$

³⁰ In words, the TOT is estimated by comparing labor outcomes over a common time interval for those workers who had access to treatment at time $t-$ and were treated if ill (and are therefore healthy over the period t) with the labor outcomes for workers who were not tested until time $t+$ but at that point found to be malaria positive (and thus assumed sick over the period t). This is a valid TOT estimate if those workers assessed positive on – in this case day $t+7$, are also malaria positive over the interval from day t to day $t+7$.

³¹ The TmUT compares labor outcomes over a common time interval for those workers who were tested and found negative at time $t-$ with the labor outcomes for workers who were not tested until time $t+$, but found to be negative at that point. Similar to the above, this is a valid TmUT estimate if those workers assessed negative on day $t+7$ are also malaria free over the interval from day t to day $t+7$.

The Treatment on the Medically Untreated, TmUT, tests whether ‘positive’ health information affects a worker’s productivity, as observed in previous work, and whether it impacts physical activity.³²

Robustness checks: further trimming and Manski Bounds

Robustness tests are implemented to assess the sensitivity of results to protocol noncompliance, including alternative trimming protocols and Manski bounds. Trimming methods are preferred to imputation as our analytical objective is not the descriptive characterization of the population of interest, but the estimation of the empirical relationship between physical activity and health. The analytical challenge with physical activity data is to distinguish between noncompliance and true sedentary behavior. This analysis investigates the distribution of the sedentary physical activity data and establishes conservative thresholds which assign workers as noncompliant based on this distribution. As this approach is inherently subjective, alternative trimming rules are designated as described earlier. For the main analysis, which is termed Trim A, worker day observations are included that report sedentary time for no more than 21 hours in a work day, and no more than 22.5 hours on a non-work day. The top and bottom 1% of rods cut per work hour in a given day are Winsorized. This results in a sample of 2,106 worker-day observations for the analysis, reduced from 3162 worker-day observations that were collected in the study. Two alternative less conservative trimming rules, Trims B and C, adopt more restrictive definitions of a valid data day and result in analytic samples of 1,723 and 1,283 worker-day observations respectively.³³

As a complement to trimming, Manski bounds are implemented for the estimated treatment effects related to the malaria test, and medical treatment if applicable (Manski 1990, Horowitz and

³² Dillon et al. (forthcoming) find that this group responds to the diagnosis by both increasing productivity (effort) within occupation as well as switching into sugarcane cutting from lower return occupations. That study rules out many alternative explanations for this behavioral response and finds that workers most likely to be surprised by a healthy diagnosis based on pre-test expectations respond the strongest. A key difference of this study season compared with previous seasons is that the plantation did not allow cane cutters to switch between the cutting and scrubbling tasks in this season. In contrast to the earlier work, this study therefore does not investigate impact on task or occupational switching.

³³ Trim B considers sedentary for no more than 19 hours in a work day, no more than 20.5 hours on a non-work day, and not in the top or bottom 1% of rods cut in a given day. Trim C considers sedentary for no more than 17 hours in a work day, no more than 18.5 hours on a non-work day, and not in the top or bottom 1% of rods cut in a given day.

Manski 1990). The advantage of Manski bounds is that they require no distributional assumptions and provide an upper (and lower) bound for the treatment effect based on the worst- (and best-case) scenarios when replacing missing observations with the minimum (or maximum) value in the sample. In this application, missing or trimmed observations are replaced in either the treatment or control group up to the amount that equalizes the rate of missing observations across the two groups. The Manski bounds are reported in the appendix for all treatment effects estimated in the paper, and are discussed in the results section.

5. Results

The conditional relationship between labor outcomes and physical activity are presented first to validate the use of physical activity as a measure of productivity. The results for the estimated impact of a malaria treatment and testing program on worker physical activity are presented next, and these estimates are then compared with the impact on worker productivity and labor supply.

Relationship between physical activity and individual labor outcomes

Table 2 summarizes the estimates of equation 1, reflecting the relationship between workers' labor outcomes and hours in various physical activity levels (sedentary, low, medium and high activity levels).³⁴ In the first horizontal panel, all active time is pooled into a single measure of activity, while in the second and third panels different levels of physical activity are disaggregated to demonstrate the relationship between effort intensity and daily earnings conditional on working that day, as well as daily labor supply and finally daily earnings unconditional on working.³⁵

The estimates reflect a strong relationship between productivity and physical activity, as well as between labor supply and the activity measures (i.e. a clear relationship on both the extensive and intensive margins). Column 1 of Table 2 depicts the relationship between physical activity and daily earnings (conditional on work). During workdays on the plantation, it is

³⁴ Results presented in Appendix Table S1.5 show the results without fixed effects to find similar results.

³⁵ Since 'very active' only represents a very small proportion of activity, we start by pooling 'fairly active' and 'very active' together.

especially fairly active and very active that are associated with on the job productivity – not light activity, highlighting the arduous nature of the production process. The point estimates across the three activity categories suggest that moderate activity (fairly active) is most clearly associated with daily productivity, followed by heavy activity (very active). This finding is reassuring as heavy activity captures relatively extreme physical exertion such as running, while cane-cutting, while arduous, is typically conducted at a slower pace. Lightly active time is not associated with productivity on the intensive margin.

With regard to the daily decision to supply labor to the plantation, the results in the middle column 4 of Table 2 indicate that an additional hour of (any) activity increases the likelihood of a plantation work day by 8 percentage points. Time spent in light and moderate activity has the strongest association with a work day (a marginal effect of 8 percentage points per hour for both activity categories) and then heavy activity (5 percentage points, but not precisely estimated). Clearly, workers are less physically active when not working on the plantation, and this is true not only with respect to moderate/heavy activity (the activity most associated with earnings during a day devoted to cane cutting) but also with respect to lighter physical activity. Column 7 explores unconditional earnings as a summary measure of effort with non-work days given an earnings value of zero. This is a useful measure to help us understand how physical activity projects onto both the extensive and intensive margins of productivity. Broadly speaking, 40% of the association between physical activity and earnings in our data comes from on the job effort and 60% derives from the decision to cut sugar cane on that day.³⁶

To investigate the sensitivity of the main results to data trimming decisions on protocol non-compliance (worker failure to fully wear the tracker on any given day), Table 2 also includes findings when using the less conservative trimming rules, Trims B and C, as defined earlier. The results are entirely consistent across the three alternative rules. If anything, when comparing Trims A and B, our preferred approach may be overly conservative by including many non-compliant observation days. The association between earnings conditional on work is substantially greater for Trim B's than Trim A's results. In the sample defined by Trim A, an hour of vigorous activity (fairly or very active) is associated with an additional 63 Naira while the same hour of activity in

³⁶ This reflects the estimated coefficient in the first row column 1 compared with first row column 7 (39.8/101.1).

Trim B sample is associated with an additional 96 Naira. The general conclusions regarding the association between physical activity and productivity remain in the alternative trimming approaches and, if anything, are strengthened.

Impact of malaria testing and treatment on physical activity and labor outcomes

Table 3 reports the intent to treat effect on productivity, labor supply and physical activity from the malaria testing and treatment program offered to all workers. Horizontal panel A of the table includes worker-day observations (that satisfy the Trim A assumptions) restricted to those days worked on the plantation, while includes all days observed. Outcomes include program treatment effects on labor measures (columns 1-3), physical activity hours by intensity level (columns 4-7) and the ratio of hours in each intensity level relative to the total number of active hours (columns 8-10).

The ITT effect estimated for days at work (Table 3 Panel A) indicates that the malaria program increases the daily rods cut and the earnings (Col 1-2), suggesting that malaria beliefs and malaria infection affects the productivity of this labor force.³⁷ In the week following the malaria test and offer of medication (if positive), a worker increases his daily earnings and daily output by almost one-third. When considering the impact on physical activity, the program decreases sedentary hours among workers (Col 4) and correspondingly increases time spent in more vigorous activity. After treatment, there is a shift from sedentary non-activity into moderate and heavy activity, reflecting the physical demand of improving productivity in cane cutting. Average hourly earnings for cane cutters are on average 200 Naira, so a shift of one hour from light to moderate or heavy activity represents a substantial proportional increase in output and earnings. Since the number of plantation work hours per day is fixed, we also look at the proportion

³⁷ In the overall results from an earlier harvest season, the ITT effect on labor behavior did not find a treatment impact for the 7-day reference period (Dillon et al. forthcoming). Treatment gains on daily productivity increase to .06-.09, and are precisely estimated, once a 2-3 week reference period is considered. Unfortunately the relatively few number of workers assigned accelerometers, as well as the unanticipated plantation work stoppage, prevents us from estimating physical activity effects over a longer reference period than seven days.

of active time spent at different levels of activity. Malaria testing and treatment shifts almost 10% of active time from light activity to moderate or heavy.

Panel B expands the focus to all days in the data, whether working or not. There does not appear to be a labor supply response to treatment, thereby weakening the impact of treatment on unconditioned productivity and earnings - given no labor supply response this would be the expected finding. Shifts to a more physically active day, however, are still apparent. Over all days, roughly two-thirds of an hour shift from sedentary or light activity into fairly or very active time. The proportion of active time spent in arduous activity also shifts accordingly. Given these results, it is clear that the majority of activity increase due to the malaria program occurs on days worked, with little change in activity for off days.

The TOT estimation focuses on workers who are diagnosed with malaria and yields effect sizes larger in magnitude than the ITT estimates, but now estimated imprecisely, likely due to the reduced sample size of malaria positive work days. While not significantly different from the ITT estimates, the effect sizes suggest that productivity gains from malaria treatment are larger for the malaria positive workers than for workers overall. The physical activity effects (Column 4-7) are also larger, and here precisely estimated. The fairly active category increases by 1.4 hours and very active hours increase by 0.3, with an equivalent decline in sedentary time. The increase in physical activity is nearly twice the increase estimated in the ITT estimates, suggesting that the largest gains (over seven days) of the malaria testing and treatment program accrue to the malaria positives, perhaps as to be expected given the offered efficacious treatment for a debilitating disease.

A consistent pattern also emerges when we estimate the TOT on the ratio of intensity level hours divided by the total active hours (Panel B, Columns 8-10). The estimates indicate a reallocation of 13% of total active time from light activity to either moderate or very active. When including non-work days (panel B), the gains to physical activity and unconditioned earnings are apparent but diminished from conditional results, again suggesting that the majority of changes to physical activity arise during the work period. There does not appear to be a labor supply response to treatment among the malaria positives.

The treatment effects on the medically untreated (TmUT) are reported in Table 5 and represent the effect of the worker learning that he is malaria negative. In line with earlier results

for the same setting, which identified a ‘good news’ effect on worker earnings and productivity, the estimates identify a positive effect of health information on daily productivity and earnings, with the magnitude of gains roughly half that of the ITT estimates. In an endemic area, the effect of unanticipated health information may cause workers to revise their expectations of short-run work capacity and the perceived cost of effort, consistent with previous findings in the same setting (Dillon et al. forthcoming). Correspondingly, physical activity also shifts due to the ‘good news’ effect, indicating a reduction of sedentary and light work of approximately 0.8 hours duration with corresponding increases in fairly active and very active hours. The ratios of these intensity levels relative to total active minutes indicates a significant shift from light to fairly and very active time constituting eight percent of total active time. As with the TOT and ITT, there do not appear to be changes to labor supply (Panel B), indicating that changes to activity patterns only occur during work days.

The consistency of the ITT, TOT, and TmUT impact estimates on physical activity, along with those on productivity, provide additional support for physical activity as a valid measure for worker effort and productivity in this setting. To assess the robustness of the above results we re-estimated all models, now using different trimming rules, i.e. Trim B and Trim C, as well as the corresponding Manski bound estimates. These estimation results are reported in Appendix Tables S3-S5 for the ITT, TOT and TmUT effects respectively. In general, the estimation results tell the same story with the less conservative trimming rules B and C as with Trim A. The consistency seen in the Manski bounds estimates further confirm the robustness of these estimates.

6. Conclusion

The empirical measurement of labor productivity at the micro level is a key to improving our understanding of determinants of worker productivity. Randomized controlled trials that analyze the causal impact of important factors such as the organization of work, incentives, contract design, training or health care on worker productivity have brought about a new interest in the measurement of individual worker level productivity. Task specific measures of worker output can provide a useful measure, when observed. A growing literature makes creative use of output measures, including the kilogram amount of fruit picked per day (Bandiera et al 2010), the

number of trees planted per day (Shearer 2004), the amount of data entered per day (Kaur et al 2015), the number of windshields installed per day (Lazear 2000), or the number of items the cashier scanned per second (Mas and Moretti 2009) (see, again, Table 1).

Nevertheless, measuring individual worker productivity remains a severe challenge in settings where individual worker output is unobserved or not clearly attributable to a unit of time. The traditional approach has been to revert to daily or hourly earnings as a measure for worker productivity, but in light of weak contemporaneous correlations between earnings and productivity in a number of settings, measures of worker effort may be of intrinsic interest. While wages can be correlated with productivity, they are co-determined by a range of other factors, including the institutional environment, job attributes, and worker individual characteristics, which are subject of a vast (theoretical and empirical) literature. Effort is also central to a number of theoretical debates and has attracted policy interest (Becker 1977, 1985, Gibbons 1987, Acemoglu and Pischke 1998, Lazear 2000, Fudenberg and Rayo 2019). This study investigates whether a direct measure of physical activity using accelerometers, can draw valid inference of worker effort in order to approximate individual worker productivity. The specific context is that of an arduous physical task – the cutting of cane sugar at a large plantation in Nigeria, and we expect the application to be generalizable to numerous settings as discussed below.

This paper first explores the association of physical activity with labor productivity (and labor supply) among agricultural workers in a piece rate wage setting, and finds a strong association. Each active hour is associated with increased (unconditional) earnings of 101 Naira which is 13% of the earnings standard deviation. To further validate physical activity as a measure of worker productivity, the study estimates the impact of a malaria testing and treatment program on both productivity and physical activity, building on previous work. If physical activity is a good measure for worker productivity in this setting we expect to see an impact of the health treatment on activity, in line with earlier results that find an effect of the malaria testing and treatment program on worker output. The results demonstrate an overall consistency between the estimated impacts on physical activity with the estimated impacts on observed productivity. Moreover, the physical activity impact estimates shed further light on the productivity impacts from earlier work, as they show where shifts in effort occur. Workers daily average sedentary time and light physical

activity are reduced after treatment, while fairly active and very active physical activity levels increased. The shifts in effort levels is larger among the malaria positives (TOT). A directionally similar but smaller change is observed among the malaria negatives (TmUT), reflecting the impact of malaria information. This shift is in line with results from a previous round but is imprecisely estimated. All told, the results indicate that physical activity can be a useful input for studying the determinants of productivity, and may reveal important mechanisms in analysis of productivity change.

The study also highlights challenges that come with the measurement of physical activity. First, implementation requires well-designed field protocols based on piloting to assess where devices are worn by the workers for most accurate measurement and to minimize loss. This may depend on the specific task. Further work validating the location of the accelerometer as well as whether a single or multiple accelerometer might provide the most accurate measure of productivity, conditional on specific tasks, would be a welcome advance to this field of study. The study also revealed challenges related to subject compliance and accelerometer loss, in line with other work using accelerometers to measuring physical activity. Alternative monitoring or incentive mechanisms during fieldwork may increase compliance.

A related question is whether the device should be used on its own or in combination with other measures of worker productivity. This assuredly relates to the nature of the sector and activity considered. For any activity, where physical activity has not been validated as a measure of individual worker productivity, it seems best to use both in tandem, at least in a pilot phase.³⁸

A second challenge is the need to address any imperfect compliance in the data that remains after efforts to minimize non-compliance are taken. This paper reports trimming criteria in detail, and use robustness checks to assess the sensitivity of the results to the criteria used. As shown in the overview of studies in Appendix 2, the trimming criteria adopted can vary considerably across

³⁸ In cases where worker productivity is directly observed, this allows for a direct assessment of the measure's validity, potentially with personnel data or a labor diary. In cases where worker productivity is not individually observed, parallel collection of information on labor outcomes, through observation or recall, seems recommended. Once physical activity has been shown to be correlated with worker output, parallel data collection on worker output may be deemed unnecessary.

recent studies, in the absence of a harmonized protocol.³⁹ One important consequence of the need for trimming is that it reduces the sample of analysis and this may pose a challenge, especially when additional unforeseen circumstances arise (in our case a company closure of two weeks). Ex-ante statistical power calculations should take this into account.

Future applications of physical activity measurement in development economics

What is the future use and usefulness of physical activity measurement in development economics research? A range of potential applications exist at both the descriptive level as well as for testing of theoretical predictions.

When productivity is *observed*, studying its relationship to physical effort can yield additional insights on the channels through which productivity is influenced, distinguishing, for instance, possible complements or substitutes to effort in the production process. The estimated treatment effects of a malaria program demonstrate this mechanism. Workers are not only more productive in response to treatment, but we identify the margin on which productivity change occurs. In the results, productivity does not come from more efficient time use (which is fixed in our context, but could be observed in other contexts), but rather stems from increased effort, which is measured. When productivity is *unobserved*, the assessment of effort can provide a measure for productivity given assumptions on the physical nature of work and its incentive compatibility. This is of particular interest when there is no payment transaction or market wage, such as in the cases of non-market household tasks (e.g. fetching water) or subsistence agriculture (e.g. land preparation, planting, weeding and harvesting). The same holds for non-agricultural work tasks with similar characteristics (certain manufacturing jobs, mining, quarrying, construction, and select urban services (e.g. portage)).

When effort and productivity are closely related, considering the impact on effort provides little additional information if output per unit of time itself is observed – although it may still be

³⁹ The desirability of a harmonized protocol is topic of current discussion, as illustrated by a recent Delphi study with 16 physical activity experts, which holds a plea for such harmonization. 10 were in favor of a more centralised harmonization, 2 for a decentralised harmonization, 1 for mixed approach and 1 had no opinion (See Wijnendaele et al 2015). At the time of that study, most studies were still concentrating on lab contexts rather than free living environment. Despite its increasing use, experience with the technology remains relatively young and data collected in specific applications and in a free living environment may require their own protocol.

useful when the aim is to test specific theoretical predictions on the relationship between effort and productivity, and how ability may mediate this relationship. Indeed, incorporating this new technology into a broad range of studies can also allow improved testing of theoretical predictions that have remained largely out of reach so far. We provide four examples.

- *The role of health for labor*

Though this paper is one empirical example, the relationship between health and productivity is fundamental to understanding the returns to health investment, as health affects worker's physical capacity and the provision of effort.⁴⁰ For physically demanding tasks, physical activity provides an attractive measure of effort, and allows for the study of the role of health in setting a workers' 'physical activity frontier'.⁴¹ Aggarwal et al. (2019) estimates the effect of program design incentives on patient compliance with diabetes treatment using physical activity as an outcome. Emerging studies on the role of sleep and productivity are a further example where accelerometers have been used to measure sleeping time (Rao et al. 2019).

- *Effort and productivity in non-market activities*

A range of work activities take place outside the market. While shadow wages may provide a proxy for worker productivity in some settings, they do not in other settings. Physical activity can provide a welcome measure of effort, and possibly productivity, for physically demanding non-market tasks. One set of examples are household activities, often carried out by women and children, like fetching water, collecting firewood, etc. Another set of examples arise in subsistence farming, particularly household labor in agriculture. For specific activities like land preparation, ploughing, planting, weeding, and harvesting physical effort can be measured as a likely important input to the production process alongside more easily recalled inputs (such as fertilizer use). When

⁴⁰ From a human capital perspective it is remarkable how little is known about the impact of health on adult labor outcomes, in contrast to the vast literature analysing the role of education in determining labor outcomes. While one can think of a number of good reasons why this is the case (including the multidimensionality of health, the concentration of education early on in life, addressing at least some of the endogeneity, and the relative ease with which to measure standard education constructs) it also means that much labor analysis does not take health into account.

⁴¹ A recent example provides an illustration. Blattman and Dercon (2018) using a RCT find that workers have a strong preference to work as self-employed when they are given the choice between working in self-employment or a manufacturing firm in Ethiopia. An important reason, they argue, is that people do not keep manufacturing jobs because they are physically too demanding and negatively impact on their health (In recent follow up work they find that these effects disappear in the long term.) Studies of this nature would benefit from direct measurement of physical activity.

linked with time diaries, accelerometers have the potential to also facilitate the study of time use, as the amount of time in any particular activity can be mapped and classified ex-post by a survey respondent.

- *Behavioral responses to incentive structures*

The notion that a piece rate is a superior contract when individual worker output is easy to measure rests on the assumption that workers are fully rational and optimizing in their choices. Recent evidence demonstrates that workers deliver less effort than they plan and want to provide (Kaur 2015), suggesting that both employers and workers have an incentive to provide commitment contracts that address failures in self-discipline and that increase worker effort. Direct measurement of effort can help build a better understanding of the patterns of effort over time, its correlates with observables (e.g. time preferences), and its responsiveness to incentives.

- *Labor supply to maximize or to target income?*

The classic approach to labor supply assumes that workers maximize income conditional to time and effort capacity constraints. An alternative approach argues that workers target income levels - perhaps best reflected in the papers on taxi drivers.⁴² Dupas et al. (2018) provides direct evidence for earned income targeting among bicycle taxi drivers in Kenya. Little is known about the role of effort in this decision making. For instance, for physically arduous tasks, income targeting may be more attractive for those with lower physical capacity, for whom physical capacity forms a binding constraint and prohibits maximizing income (while it is not yet binding for those with higher physical capacity)

A final question relates to the affordability of this type of research. A brief cost calculation shows its affordability, in particular with the falling unit price of physical activity monitors. The cost estimates sum to a total cost of USD 3,890, or USD 46,9 per worker which corresponds to USD 1.56 per work day observation.⁴³

⁴² This debate is reflected in a series of papers, with Camerer et. al (1997) arguing that taxi drivers target income and adapt their labor supply accordingly, while Farber (2005) argues against this, and Crawford and Meng (2011) find that taxi drivers engage in hours *and* income targeting.

⁴³ In this calculation, each worker has 6 weeks x 5 days = 30 days of observations. Due to the disruption in weeks 5 and 6 the 8-week planned study period was reduced to 6 weeks actual work in the field. An accelerometer with the same functions as the one used in this study can now be purchased for USD 35 a piece (at the time of writing - and prices are falling further). The hardware expenses of the study would thus today amount to USD 35 * 83 workers = USD 2,905. The fieldwork preparation, pilot and implementation costs were standard and will vary across differing

Overall, this study suggests that physical activity can be a useful measure of effort and can help to describe key mechanisms through which worker productivity changes for specific activities. This is especially attractive when other measures of output or effort are unavailable or costly to observe. Challenges accompanying the use of this technology can be overcome through well-prepared implementation and analysis.

survey contexts. The additional costs associated with implementing the accelerometer protocol include additional field time to collect, charge and redistribute the accelerometers, estimated to require an additional day per week for the four enumerators and one supervisor who implemented the study (6 additional days of the field team costs USD 471). There is also data management costs which we estimated to occupy 20% of a data manager's time in the field (USD 514). This calculation of marginal cost assumes that the enumerator teams will already stay in the field for the duration of data collection for other survey-based reasons.

References

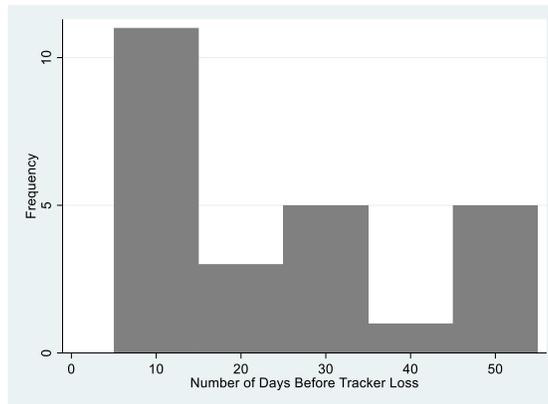
- Abadie, A et al. 2017. “When Should You Adjust Standard Errors for Clustering?” *NBER Working Paper*. doi: 10.3386/w24003.
- Acemoglu, D. and Pischke, J. (1998). “Why do Firms Train? Theory and Evidence,” *The Quarterly Journal of Economics* 113 (1): 79-119.
- Ahmed, A. U., and Bouis, H. E. 2002. “Weighing What’s Practical: Proxy Means Tests for Targeting Food Subsidies in Egypt.” *Food Policy*, 27(5), 519-540.
- Asker, J., Collard-Wexler, A., and De Loecker, J. 2014. “Dynamic Inputs and Resource (Mis) allocation.” *Journal of Political Economy*, 122(5), 1013-1063.
- Baird, S., Gong, E., McIntosh, C., and Ozler, B. 2014. “The heterogeneous effects of HIV testing”. *Journal of Health Economics*, 37(1): 98-112.
- Bandiera, O, Barankay, I., and Rasul, I. 2010. “Social Incentives in the Workplace.” *Review of Economic Studies*, 77(2): 417-458.
- Baranowski, T., Abdelsamad, D., Baranowski, J., O’Connor, T. M., Thompson, D., Barnett, A., Cerin, E., and Chen, T. A. 2012. “Impact of an Active Video Game on Healthy Children’s Physical Activity.” *Pediatrics*, 129(3), e636-e642.
- Bartelsman, E. J., & Doms, M. 2000. “Understanding Productivity: Lessons from Longitudinal Microdata.” *Journal of Economic Literature*, 569-594.
- Bartelsman, E., Haltiwanger, J., and Scarpetta, S. 2013. “Cross-Country Differences in Productivity: The Role of Allocation and Selection.” *The American Economic Review*, 103(1), 305-334.
- Bauman, A., Phongsavan, P., Schoeppe, S. and Owen, N., 2006. Physical activity measurement—a primer for health promotion. *Promotion & education*, 13(2), pp.92-103.
- Becker, G. 1985. “Human Capital, Effort, and the Sexual Division of Labor.” *Journal of Labor Economics* 3(1): Part 2, S33-S58.
- Becker, G. 1977. “A Theory of the Production and Allocation of Effort.” NBER Working Paper Series, No. 184.
- Bellemare, C, and Shearer, B. “Gift giving and worker productivity: Evidence from a firm-level experiment.” 2009. *Games and Economic Behavior*, 67(1), 233-244.
- Bort-Roig, J, Gilson, ND., Puig-Ribera, A., Contreras, RS., and Trost, SG. 2014. “Measuring and Influencing Physical Activity with Smartphone Technology: A Systematic Review.” *Sports Medicine*, 44(5): 671-686.
- Chaput J-P, 2019, *Lancet Global Health*, Vol 7 February 2019
- Chen, K. et al. (2008) “Comparing the Performance of Three Generations of ActiGraph Accelerometers” *Journal of Applied Physiology* 105 (4). 1091-1097.
- Corder, K., and Van Sluijs, E. M. 2010. “Invited Commentary: Comparing Physical Activity across Countries—Current Strengths and Weaknesses.” *American Journal of Epidemiology*, kwq068.
- Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F. and Oja, P. 2003. “International Physical Activity Questionnaire (IPAQ): 12-Country Reliability and Validity.” *Medicine & Science in Sports & Exercise*, 35, 1381-1395.
- Cropper, M. L., Haile, M., Lampietti, J., Poulos, C., and Whittington, D. 2004. “The Demand for

- a Malaria Vaccine: Evidence from Ethiopia.” *Journal of Development Economics* 75: 303-318.
- dalla Martha R.C., Tada M.S., Ferreira R.G., da Silva L.H., Wunderlich G. 2007. “Microsatellite Characterization of *Plasmodium falciparum* from Symptomatic and Non-symptomatic Infections from the Western Amazon Reveals the Existence of Non-symptomatic Infection-associated Genotypes.” *Memórias do Instituto Oswaldo Cruz* 102:293-298.
- Deressa, T. 2007. “Measuring the Economic Impact of Climate Change on Ethiopian Agriculture: Ricardian approach”. World Bank Policy Research Paper No. 4342, World Bank, Washington DC.
- Dillon, A., Friedman, J., and Serneels, P. “Health Information, Treatment, and Worker Productivity”. *Journal of European Economic Association*, forthcoming.
- Dupas, P. 2011. “Health Behavior in Developing Countries.” *Annual Review of Economics*, 3(1), 425-449.
- FAO and WHO. 2001. Human energy requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. FAO: Rome.
- Fudenberg, D. and Rayo, L. (2019). “Training and Effort Dynamics in Apprenticeship,” *American Economic Review* 109 (11): 3780-2812.
- Gibbons, R. 1987. "Piece-Rate Incentive Schemes." *Journal of Labor Economics*: 413-429.
- Gollin D., C. Udry, 2017, Heterogeneity, Measurement Error, and Misallocation: Evidence from African Agriculture. NBER Working Paper, w25440. Doi: 10.3386/w25440
- Gong, E. 2015. “HIV Testing and Risky Sexual Behavior.” *The Economic Journal*, Vol 125, Issue 582, p 32-60.
- Government of Nigeria. 2011. “National Guidelines for Diagnosis and Treatment of Malaria.” Federal Ministry of Health, March, mimeo.
- Grosh, M., and Baker, J. L. 1995. “Proxy Means Tests for Targeting Social Programs.” *Simulations and Speculations*, 118.
- Herrmann, M. A., & Rockoff, J. E. 2012. “Worker Absence and Productivity: Evidence from Teaching” *Journal of Labor Economics*, 30(4), 749-782.
- Hicks, J.H., M. Kleemans, N. Y. Li, E. Miguel, 2017, Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata, NBER Working Paper. doi: 10.3386/w23253
- Imbens G., and J. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1): 5–86.
- Jalan, J., and Somanathan, E. 2008. “The Importance of Being Informed: Experimental Evidence on Demand for Environmental Quality.” *Journal of Development Economics*, 87(1), 14-28.
- Lazear, E. P. 2000. "Performance Pay And Productivity" *American Economic Review*, v90 (5,Dec), 1346-1361.
- Laishram, D. D., Sutton, P. L., Nanda, N., Sharma, V. L., Sobti, R. C., Carlton, J. M., and Joshi, H. 2012. “The Complexities of Malaria Disease Manifestations with a Focus on Asymptomatic Malaria.” *Malaria Journal*, 11(1): 29.
- Mas, A. and Moretti, E. 2009. “Peers at Work” *American Economic Review* 99 (1), 112-145.
- Miller, L. H., Baruch, D. I., Marsh, K., and Doumbo, O. K. 2002. “The Pathogenic Basis of Malaria.” *Nature*, 415(6872), 673-679.
- National Bureau of Statistics, 2012, Nigeria Poverty Profile.

- Najera, J. A. and Hempel, J. 1996. "The Burden of Malaria." World Health Organization /MAL/96.10.
- Nelson M. C., K. Taylor, C. A. Vella, 2019, Comparison of Self-Reported and Objectively Measured Sedentary Behavior and Physical Activity in Undergraduate Students, *Measurement in Physical Education and Exercise Science* 2019, VOL. 23, NO. 3, 237–248
- Rottmann, M., Lavstsen, T., Mugasa, J. P., Kaestli, M., Jensen, A. T. R., Müller, D., Theander, T. and Beck, H-P. 2006. "Differential Expression of Var Gene Groups is Associated with Morbidity Caused by *Plasmodium falciparum* Infection in Tanzanian Children." *Infection and Immunity* 74 (7): 3904-3911
- Sallis, J.F., Cerin, E., Conway, T.L., Adams, M.A., Frank, L.D., Pratt, M., Salvo, D., Schipperijn, J., Smith, G., Cain, K.L. and Davey, R., 2016. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. *The Lancet*, 387(10034), pp.2207-2217.
- Salvo, D., Torres, C., Villa, U., Rivera, J.A., Sarmiento, O.L., Reis, R.S. and Pratt, M. 2015. "Accelerometer-based physical activity levels among Mexican adults and their relation with sociodemographic characteristics and BMI: a cross-sectional study." *International Journal of Behavioral Nutrition and Physical Activity*, 12(1), p.79-90.
- Sauerman J., 2016, Performance measures and worker productivity, IZA World of Labor, 260
- Sherar, L.B., Griew, P., Esliger, D.W., Cooper, A.R., Ekelund, U., Judge, K. and Riddoch, C. 2011. "International children's accelerometry database (ICAD): design and methods". *BMC Public Health*, 11(1), p.485-498.
- Sobngwi, E., Mbanya, J. C. N., Unwin, N. C., Aspray, T. J., and Alberti, K. G. M. M. 2001. "Development and Validation of a Questionnaire for the Assessment of Physical Activity in Epidemiological Studies in Sub-Saharan Africa." *International Journal of Epidemiology*, 30(6), 1361-1368.
- Sowunmi, A., Gbotosho, G. O., Happi, C. T., Adedeji, A. A., Fehintola, F. A., Folarin, O. A., Tambo E., and Fateye, B. A. 2007. "Therapeutic Efficacy and Effects of Artemether-lumefantrine and Amodiaquine-sulfalene-pyrimethamine on Gametocyte Carriage in Children with Uncomplicated *Plasmodium falciparum* malaria in Southwestern Nigeria." *The American journal of tropical medicine and hygiene*, 77(2), 235-241.
- Speakman, J., 1997. *Doubly labelled water: theory and practice*. Springer Science & Business Media.
- Storm, F. A., Heller, B. W., and Mazzà, C. 2015. "Step Detection and Activity Recognition Accuracy of Seven Physical Activity Monitors." *PloS One*, 10(3), e0118723.
- Strauss, J. and Thomas, D. 1998. "Health, Nutrition, and Economic Development." *Journal of Economic Literature*, 36 (2):766-817
- Syverson, C. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2), 326-365.
- Thornton, R. L. 2008. "The Demand for, and Impact of, Learning HIV Status." *American Economic Review*, 98(5), 1829.
- Toure, F.S., Bisseye, C. and Mavoungou, E. 2006. "Imbalanced Distribution of *Plasmodium falciparum* EBA-175 Genotypes Related to Clinical Status in Children from Bakoumba, Gabon." *Clinical Medicine and Research*, 4:7-11.
- White, N.J. 2005. "Intermittent Presumptive Treatment for Malaria". *PLoS Medicine* 2: 1-63.

- Wijndaele K., et al., 2015, Utilization and Harmonization of Adult Accelerometry Data: Review and Expert Consensus, *Medicine & Science in Sports & Exercise*, 47, p2129-39
- WHO. 2010, Guidelines for the Treatment of Malaria (2e). 2nd Edition. Geneva: World Health Organization.
- WHO. 2012. *World Malaria Report 2012*. Geneva: World Health Organisation.
- Woodring, J. V., Ogutu, B. Schnabel, D., Waitumbi, J. N., Olsen, C. H., Walsh, D. S., Heppner Jr., D. G. and Polhemus, M. E. 2010. "Evaluation of Recurrent Parasitemia after Artemether-Lumefantrine Treatment for Uncomplicated Malaria in Children in Western Kenya." *American Journal of Tropical Medicine and Hygiene* 83: 458-464.
- Zanello, G., Srinivasan, C. S., and Nkegbe, P. 2017. "Piloting the Use of Accelerometry Devices to Capture Energy Expenditure in Agricultural and Rural Livelihoods: Protocols and Findings from Northern Ghana." *Development Engineering*. 2: 114-131.

Figure 1.
Days Before Tracker Loss

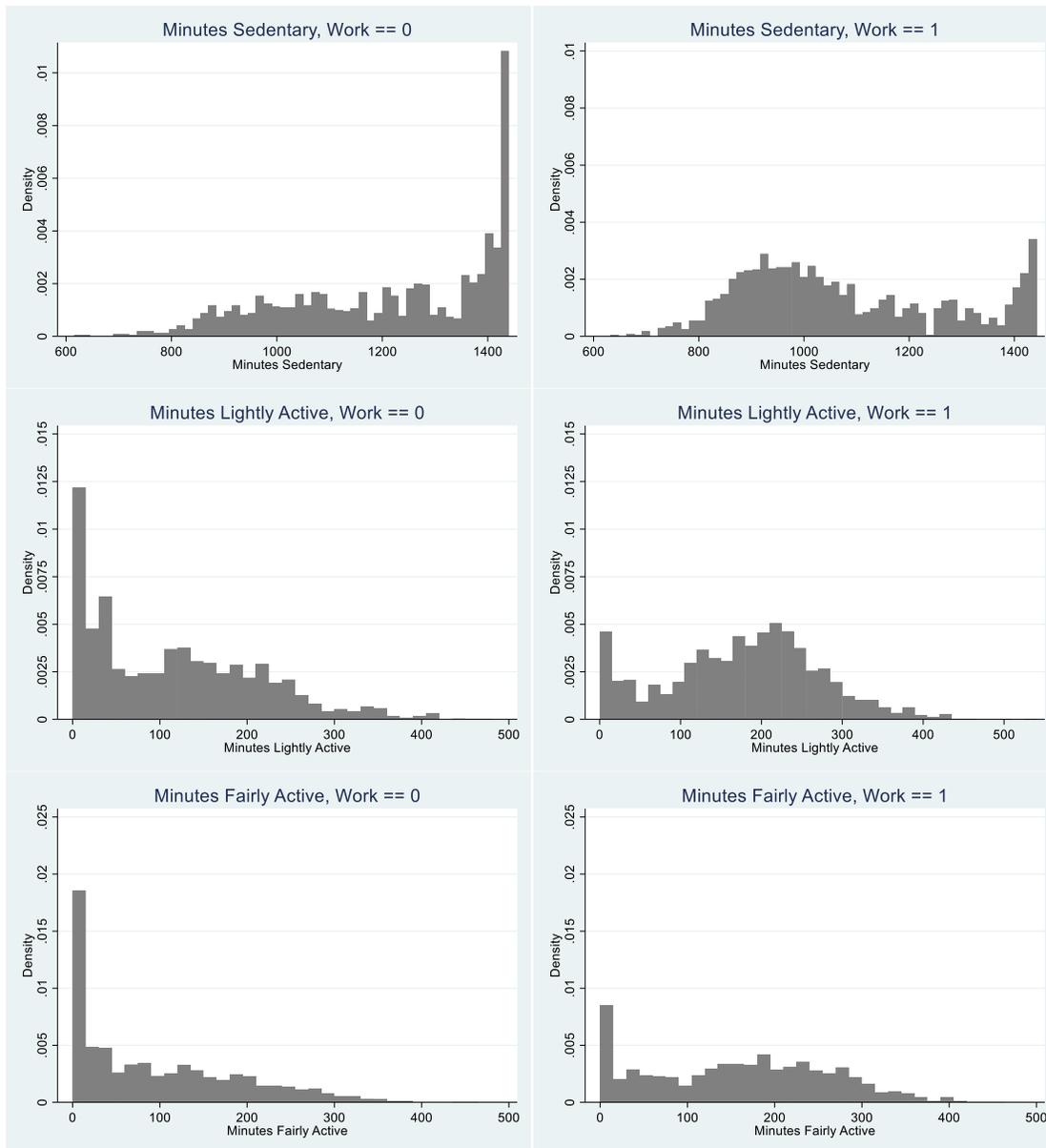


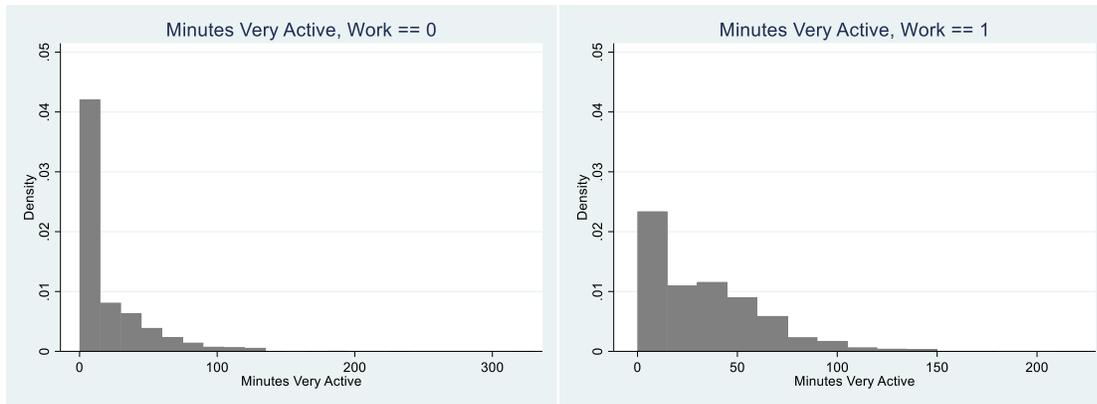
Source: Author's calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.

Notes: The figure shows when respondents lost their accelerometer, given that their instrument stopped recording data before the completion of the study. They were grouped into bins where the number of days was assigned to the 10-day bin if the accelerometer was lost before the 15th day, the 20-day bin if it was lost between the 16th and 25th day, the 30-day bin if it was lost between the 26th and 35th day, the 40-day bin if it was lost between the 36th and 45th day, and the 50-day bin if it was lost after the 45th day **but** before the completion of the study (harvest season).

Figures 2.

Minutes Sedentary, Lightly Active, Fairly Active and Very Active when non-working versus working





Notes: The above figures depict the frequency and distribution of the activity data at different intensity levels as well as between working and non-working days. The level of intensity was defined by Fitbit’s proprietary algorithm, and workdays were established with personnel data provided by the plantation.

Source: Author’s calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.

Table 1.

Measures of worker productivity across sectors and activities

Agriculture	Fruit pickers	Kg fruits picked per day	Bandiera, O., Barankay, I., Rasul, I. "Social incentives in the workplace" <i>Review of Economic Studies</i> 77:2 (2010): 417–458.
	farm workers	Quantity of fruit harvested per day	Zivin G., J. M. Neidell. 2012. "The Impact of Pollution on Worker Productivity." <i>American Economic Review</i> , 102(7):3652-73.
	Sugar cane cutters	Amount of cane (number of rods) cut per day	Dillon et al. (forthcomng) "Health Information, Treatment and Worker Productivity: Experimental Evidence from Malaria Testing and Treatment among Nigerian Sugarcane Cutters," World Bank Policy Research Working paper 7120.
	Tree planters	Number of trees planted per day	Shearer B., Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment, <i>The Review of Economic Studies</i> , Vol. 71, No. 2 (Apr., 2004), pp. 513-534.
Services	Data entry	Data entry per day	Kaur S., M. Kremer, and S. Mullainathan, Self-Control at Work, <i>Journal of Political Economy</i> 2015 123:6, 1227-1277
	Windshield installers	Number of windshields installed per day	Lazear, E. "Performance pay and productivity" <i>American Economic Review</i> 90:5 (2000): 1346–1361.
	Supermarket cashiers	Number of items scanned per second	Shaw, K., Lazear, E. "Tenure and output" <i>Labour Economics</i> 15:4 (2008): 704–723.
	Call agents	Average length of call	Mas, A., Moretti, E. "Peers at work" <i>American Economic Review</i> 99:1 (2009): 112–145.
	lawyers	Number of hours billed	De Grip, A., Sauermann, J. "The effects of training on own and co-worker productivity: Evidence from a field experiment" <i>The Economic Journal</i> 122:560 (2012): 376–399.
	Academic scientists	PhD thesis gets published in top journal, number of citations	Azmat, G., Ferrer, R. Gender Gaps in Performance: Evidence from Young Lawyers IZA Discussion Paper No.9417, 2015.
	Academic scientists:	impact factor weighted publications	Waldinger, F. "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany" <i>Journal of Political Economy</i> 118:4 (2010): 787–831.
	Teachers	Teacher value added	Azoulay, P., Graff Zivin, J., Wang, J. "Superstar extinction" <i>Quarterly Journal of Economics</i> 125:2 (2010): 549–589.
	Paramedics	Transport times, number of procedures	Chetty, R., Friedman, J. N., Rockoff, J. E. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood" <i>The American Economic Review</i> 104:9 (2014): 2633–2679. Jackson, C. K., Bruegemann, E. "Teaching students and teaching each other: The importance of peer learning for teachers" <i>American Economic Journal: Applied Economics</i> 1:4 (2009): 85–108. Brachet, T., David, G., Drechsler, A. M. "The effect of shift structure on performance" <i>American Economic Journal: Applied Economics</i> 4:2 (2012): 219–246.

Politicians	Number of submitted bills	Ferraz, C., Finan, F. Motivating Politicians: The Impacts of Monetary Incentives on Quality and Performance NBER Working Papers No.14906, 2009.
Professional service firms (law, banking, accounting)	Associate performance in up-or-out contests	Barlevy, G. and Neal, D. "Allocating effort and talent in professional labor markets," Journal of Labor Economics 37 (1): 187-246.
Navy recruiters	Number and quality of results	Asch, B. J. "Do incentives matter? The case of navy recruiters" Industrial and Labor Relations Review 43:3 (1990): 89S–106S.

Manufacturing

Apprenticeship		Fudenberg, D. and Rayo, L. "Training and Effort Dynamics in Apprenticeship." American Economic Review 109: 11 (2019).
Factory workers	Fish processing	Park, S. "Socializing at work: Evidence from a field experiment with manufacturing workers," American Economic Journal: Applied Economics 11: 3 (2019).

Sources: Author's own summary and Sauerman (2016).

Table 2.

Relationship between Labor Outcomes and Physical Activity (All Trims; Workgroup by Week Fixed Effects)

	Daily earnings (conditional on working)			Work Day			Daily earnings (unconditional on working)		
	Trim A	Trim B	Trim C	Trim A	Trim B	Trim C	Trim A	Trim B	Trim C
Hours active	39.756*** (10.091)	69.882*** (14.578)	58.850* (23.902)	0.075*** (0.006)	0.091*** (0.008)	0.116*** (0.014)	101.143*** (8.985)	132.676*** (13.941)	162.967*** (2.286)
Hours lightly active	-1.311 (20.931)	22.723 (23.924)	18.135 (31.504)	0.081*** (0.012)	0.082*** (0.013)	0.104*** (0.019)	81.191*** (18.431)	94.036*** (21.089)	124.522*** (28.128)
Hours fairly or very active	60.661*** (13.122)	96.499*** (16.059)	83.440*** (23.875)	0.072*** (0.008)	0.095*** (0.009)	0.123*** (0.014)	113.254*** (13.849)	155.924*** (16.378)	185.706*** (23.511)
Hours lightly active	-1.960 (21.671)	22.778 (24.467)	20.472 (31.583)	0.079*** (0.012)	0.079*** (0.014)	0.099*** (0.018)	78.208*** (19.114)	89.626*** (21.578)	119.500*** (28.292)
Hours fairly active	63.170*** (20.357)	96.309*** (23.150)	76.470** (29.489)	0.079*** (0.010)	0.107*** (0.011)	0.138*** (0.015)	124.989*** (18.053)	170.813*** (20.153)	201.149*** (25.735)
Hours very active	51.007 (56.046)	97.257* (57.489)	112.287* (62.667)	0.047 (0.032)	0.055* (0.033)	0.069* (0.035)	73.630 (53.179)	104.328* (53.624)	131.621** (59.797)
Number of worker-days	1190	1024	752	2106	1723	1238	2106	1723	1238

Notes: *** p< 0.01, **p<0.05, *p<0.1. Trim levels are based on the following criteria – for trim A observations were trimmed (replaced as missing) when the number of sedentary hours exceeded 21 on a work day or exceeded 22.5 on a non-workday. In trim B observations were trimmed when the number of sedentary hours were between

19 and 21 on a workday or between 20.5 and 22.4 on a non-workday. In trim C observations were trimmed when the number of sedentary hours were between 17 and 19 on a workday or between 18.5 and 20.5 on a non-workday. Standard errors are clustered at the worker-level, and workgroup-week fixed effects are incorporated. Work days are included in this analysis when they correspond with personnel data acquired from the plantation and the total number of hours of accelerometer data in that day was equal to 24.

Source: Author's calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.

Table 3.
Intent to Treat Effects (ITT) on Labor Outcomes and Physical Activity

Panel A: Work-Day Observations where Labor Supply==1										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira) (1)	Daily Rods Cut (2)	Daily Labor Supply (1=Worked) (3)	Sedentary Hours (4)	Light Work Hours (5)	Fairly Active Hours (6)	Very Active Hours (7)	Light/ Active Hours (8)	Fair/ Active Hours (9)	Very/ Active Hours (10)
Program Offer (1=Yes)	456.225* (234.451)	203.592* (104.663)		-0.946** (0.365)	-0.120 (0.155)	0.848*** (0.236)	0.217*** (0.071)	-0.094*** (0.022)	0.069*** (0.017)	0.025*** (0.008)
Constant	1,386.202*** (239.116)	618.887*** (106.744)		18.187*** (0.579)	3.614*** (0.390)	1.842*** (0.270)	0.356*** (0.122)	0.645*** (0.049)	0.306*** (0.035)	0.049*** (0.018)
Number of Worker-Day Observations	559			503				503		
Panel B: Full Sample										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira)	Daily Rods Cut	Daily Labor Supply (1=Worked)	Sedentary Hours	Light Work Hours	Fairly Active Hours	Very Active Hours	Light/ Active Hours	Fair/ Active Hours	Very/ Active Hours
Program Offer (1=Yes)	184.546 (158.508)	82.344 (70.762)	-0.038 (0.064)	-0.566* (0.299)	-0.102 (0.136)	0.528*** (0.190)	0.140** (0.060)	-0.060*** (0.018)	0.042*** (0.015)	0.018** (0.007)
Constant	-1,160.923*** (289.250)	-518.331*** (129.124)	-0.770*** (0.124)	13.198*** (0.957)	5.525*** (0.603)	3.194*** (0.476)	2.083*** (0.219)	0.563*** (0.054)	0.257*** (0.044)	0.180*** (0.022)
Number of Worker-Day Observations	333			333				333		

Note: *** p< 0.01, **p<0.05, *p<0.1. Trim levels are based on the following criteria – for trim A observations were trimmed (replaced as missing) when the number of sedentary hours exceeded 21 on a work day or exceeded 22.5 on a non-workday. In trim B observations were trimmed when the number of sedentary hours were between 19 and 21 on a workday or between 20.5 and 22.4 on a non-workday. In trim C observations were trimmed when the number of sedentary hours were between 17 and 19 on a workday or between 18.5 and 20.5 on a non-workday. Standard errors are clustered at the worker-level, and workgroup-week fixed effects are incorporated. All

outcomes are for one-week reference periods. All valid worker-days are included in this analysis, and were considered valid when they correspond with personnel data acquired from the plantation and the total number of hours of accelerometer data in that day was equal to 24.

Source: Author's calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.

Table 4.

Treatment on the Treated (TOT) Effects on Labor Outcomes and Physical Activity

Panel A: Work-Day Observations where Labor Supply==1										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira) (1)	Daily Rods Cut (2)	Daily Labor Supply (1=Worked) (3)	Sedentary Hours (4)	Light Work Hours (5)	Fairly Active Hours (6)	Very Active Hours (7)	Light/Active Hours (8)	Fair/Active Hours (9)	Very/Active Hours (10)
Program Offer (1=Yes)	1,079.748 (680.832)	481.734 (304.001)		-1.694* (0.821)	-0.033 (0.316)	1.424*** (0.481)	0.303* (0.158)	-0.134*** (0.037)	0.103*** (0.026)	0.031* (0.017)
Constant	-1,673.724 (1,512.932)	-746.546 (675.542)		22.204*** (2.070)	1.201 (0.945)	1.175 (1.394)	-0.579 (0.925)	0.652*** (0.113)	0.398*** (0.097)	-0.050 (0.109)
Number of Worker-Day Observations	165			146				146		
Panel B: Full Sample										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira) (1)	Daily Rods Cut (2)	Daily Labor Supply (1=Worked) (3)	Sedentary Hours (4)	Light Work Hours (5)	Fairly Active Hours (6)	Very Active Hours (7)	Light/Active Hours (8)	Fair/Active Hours (9)	Very/Active Hours (10)
Program Offer (1=Yes)	300.060 (472.523)	133.792 (210.956)	-0.212 (0.132)	-0.468 (0.762)	-0.341 (0.259)	0.637 (0.505)	0.172 (0.155)	-0.090** (0.040)	0.064* (0.034)	0.026 (0.019)
Constant	-1,476.328* (803.368)	-658.803* (358.661)	-0.198 (0.241)	14.756*** (1.511)	3.149*** (0.411)	4.692*** (1.216)	1.403*** (0.476)	0.403*** (0.087)	0.471*** (0.103)	0.126** (0.054)
Number of Worker-Day Observations	107			107				107		

Note: *** p< 0.01, **p<0.05, *p<0.1. Trim levels are based on the following criteria – for trim A observations were trimmed (replaced as missing) when the number of sedentary hours exceeded 21 on a work day or exceeded 22.5 on a non-workday. In trim B observations were trimmed when the number of sedentary hours were between 19 and 21 on a workday or between 20.5 and 22.4 on a non-workday. In trim C observations were trimmed when the number of sedentary

hours were between 17 and 19 on a workday or between 18.5 and 20.5 on a non-workday. Standard errors are clustered at the worker-level, and workgroup-week fixed effects are incorporated. All outcomes are for one-week reference periods. All valid worker-days are included in this analysis, and were considered valid when they correspond with personnel data acquired from the plantation and the total number of hours of accelerometer data in that day was equal to 24.

Source: Author's calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.

Table 5.

Treatment on the Medically Untreated (TmUT) on Labor Outcomes and Physical Activity

Panel A: Work-Day Observations where Labor Supply==1										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira)	Daily Rods Cut	Daily Labor Supply (1=Worked)	Sedentary Hours	Light Work Hours	Fairly Active Hours	Very Active Hours	Light/Active Hours	Fair/Active Hours	Very/Active Hours
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Program Offer (1=Yes)	192.081* (113.609)	85.748* (50.718)		-0.653* (0.382)	-0.161 (0.165)	0.634** (0.265)	0.180** (0.075)	-0.080*** (0.027)	0.058** (0.022)	0.023** (0.009)
Constant	1,541.581* (197.704)	688.207*** (88.261)		18.014*** (0.589)	3.639*** (0.396)	1.968*** (0.295)	0.378*** (0.125)	0.637*** (0.052)	0.313*** (0.038)	0.051*** (0.018)
Number of Worker-Day Observations	394			357				357		
Panel B: Full Sample										
	<i>Labor Outcomes</i>			<i>Physical Activity</i>				<i>Distribution of Activity</i>		
	Daily Earnings (Naira)	Daily Rods Cut	Daily Labor Supply (1=Worked)	Sedentary Hours	Light Work Hours	Fairly Active Hours	Very Active Hours	Light/Active Hours	Fair/Active Hours	Very/Active Hours
Program Offer (1=Yes)	109.784 (114.618)	49.009 (51.169)	0.011 (0.071)	-0.598* (0.311)	-0.015 (0.157)	0.488** (0.196)	0.125** (0.060)	-0.049** (0.021)	0.034** (0.017)	0.015** (0.007)
Constant	909.350*** (201.223)	405.960*** (89.832)	0.580*** (0.084)	18.351*** (0.514)	3.397*** (0.328)	1.911*** (0.219)	0.341*** (0.115)	0.626*** (0.035)	0.328*** (0.026)	0.047*** (0.015)
Number of Worker-Day Observations	226			226				226		

Note: *** p< 0.01, **p<0.05, *p<0.1. Trim levels are based on the following criteria – for trim A observations were trimmed (replaced as missing) when the number of sedentary hours exceeded 21 on a work day or exceeded 22.5 on a non-workday. In trim B observations were trimmed when the number of sedentary hours were between 19 and 21 on a workday or between 20.5 and 22.4 on a non-workday. In trim C observations were trimmed when the number of sedentary hours were between 17 and 19 on a workday or between 18.5 and 20.5 on a non-workday. Standard errors are clustered at the worker-level, and workgroup-week

fixed effects are incorporated. All outcomes are for one-week reference periods. All valid worker-days are included in this analysis, and were considered valid when they correspond with personnel data acquired from the plantation and the total number of hours of accelerometer data in that day was equal to 24.

Source: Author's calculations are based on data for the study which comes from surveys on a sugarcane plantation in rural Nigeria from February to April 2013. Survey was conducted by the Modibbo Adama University of Technology.