NBER WORKING PAPER SERIES

UNDERSTANDING RURAL HOUSEHOLDS' TIME USE IN A DEVELOPING SETTING: VALIDATING A LOW-COST TIME USE MODULE

Erica M. Field
Rohini Pande
Natalia Rigol
Simone G. Schaner
Elena M. Stacy
Charity M. Troyer Moore

Working Paper 29671
http://www.nber.org/papers/w29671

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2022

Thanks to Nikita Kohli, Vidushi Dhawan, Sitaram Mukherjee, Neeraj Trivedi, and Rashi Sabherwal for excellent assistance in the data collection, data analysis, field management and overall project management. Thanks to Diva Dhar for input on the method design and to Madhulika Khanna and Kate Vyborny for comments on the paper. We are grateful to the Bill and Melinda Gates Foundation for generous funding for this research. This project was approved by Harvard University IRB (approval number IRB13-1934) and IFMR India IRB (approval number IRB00007107). All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Understanding Rural Households' Time Use in a Developing Setting: Validating a Low-Cost Time Use Module
Erica M. Field, Rohini Pande, Natalia Rigol, Simone G. Schaner, Elena M. Stacy, and Charity M. Troyer Moore
NBER Working Paper No. 29671
January 2022
JEL No. D13,J22,O1

ABSTRACT

Time use data facilitate deeper understanding of individual labor supply choices, especially for women, who are more likely to engage in unpaid care and home production. However, traditional time use data collection methods are time-consuming, expensive and susceptible to significant attrition. To address these concerns, we develop an abbreviated, low-cost time use survey module designed for low-literacy populations. It captures contextually-determined broad time use categories of interest to researchers - in our case, time allocations across market work, household labor, and leisure. Using survey experiments in the field, we show that, relative to the widely-used assisted diary approach, the new module is lower cost and relatively more accurate in capturing individuals' average time use. Its primary shortcomings are its limited ability to capture short duration activities and simultaneous activities. Using the example of passive childcare, we show how module design can provide accurate information on multitasking for an identified category of interest.

Erica M. Field
Department of Economics
Duke University
Social Sciences Building
Box 90097
Durham, NC 27708
and CEPR
and also NBER
emf23@duke.edu

Rohini Pande
Yale University
Department of Economics
27 Hillhouse Avenue New
Haven, CT 06520 and
CEPR
and also NBER
rohini.pande@yale.edu

Natalia Rigol
Harvard Business School
Soldiers Field
Boston, MA 02163
and NBER
nrigol@hbs.edu

Simone G. Schaner
Department of Economics and
Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA 90089
and NBER
schaner@usc.edu

Elena M. Stacy
University of California, Berkeley
estacy@berkeley.edu

Charity M. Troyer Moore
Director for South Asia Economics Research
MacMillan Center
Yale University
34 Hillhouse Ave
New Haven, CT 06520
charity.moore@yale.edu

A data appendix is available at http://www.nber.org/data-appendix/w29671

1 Introduction

Understanding how people spend their time can provide critical insight into how individual activities, and particularly gender roles, change both inside and outside the household over the course of economic development. However, reliable time use data are notoriously difficult and expensive to collect, and thus rarely available in nationally representative surveys, especially in low-income settings (Hirway, 2010). In our focus country of India, no large-scale time use data were collected over a 20-year period, during which time female labor force participation fell from 30% to 21% (World Bank Group, 2021). While standard household surveys document this decline alongside significant increases in women's education levels, they are silent on whether women reallocated their time to leisure, unpaid work (within or outside the home) or childcare, alternatives with very different implications for household dynamics and welfare.

Collecting time use data in lower-income settings is particularly challenging. Low literacy rates make self-administered time diaries infeasible. Detailed enumerator-administered modules take up significant survey time, and require enumerator expertise frequently beyond that readily available. Moreover, time use data collection in low-income settings frequently needs to account for cultural context, a higher incidence of passive caregiving, multitasking, and potential classification errors arising when households undertake the same activities for both productive and consumptive purposes (Charmes, 2006; Seymour et al., 2020).

In this paper we propose a short, low-cost time use data collection method, which we term the "stylized diary hybrid", or "Hybrid" approach. Our method is so named because it combines elements of the assisted time diary approach, in which an enumerator helps the respondent fill a time diary by reconstructing a reference day, and stylized survey questions, which collect self-reported time spent in broad activity categories. Respondents participate in a token-based time allocation exercise, in which they relay time use for the previous day to the enumerator, who allocates one token per hour across cards with photos depicting the activity category. The tokens provide a visual representation of time spent on different activities, which respondents review and refine with enumerators after narrating their day.

We validate the Hybrid module with a rural population in Northern India. Specifically, we compare data quality and module performance with two other methods: the survey-based assisted diary method used to collect time use data by India's National Sample Survey (the "Traditional" approach), and a resource-intensive "Gold Standard" method which uses high-frequency visits to respondents' households. In our validation exercise, we randomize the method deployed within and across subjects over multiple visits.

The Hybrid module performs well across the spectrum. Average survey responses align with those collected using the Gold Standard approach and align *more* closely than the Traditional approach. The module performs similarly well for the demographic strata we sampled, spanning the life cycle for rural men and women, with some deviations for young, unmarried men. The module is also substantially less expensive to utilize, requiring less enumerator training and taking 33% (5 minutes) less time than the Traditional module. We estimate that the Hybrid approach yields a savings of about 40% compared to the Traditional module, in addition to reducing potential enumerator fatigue. Finally, the module's simplicity makes it an appealing option for respondents in low-income settings, especially women: the module is easily understood by low-literacy respondents, and its brevity can help minimize measurement error and attrition arising from respondent fatigue.

By design, the Hybrid module covers a limited set of activities and lacks detail compared to typical time diary data. The majority of activities, for example, are collected in hours rather than minutes, which the use of tokens necessitates. While the module captures average time use well, our analysis suggests it performs worse than the Traditional module at recording low duration activities. The module structure also limits its ability to comprehensively capture multitasking. Using the example of passive caregiving, we show how this shortcoming can be addressed for specific short duration activities which are particularly prone to multitasking and are of interest to researchers.

A central design choice for the Hybrid module is category selection, as appropriate categories depend on research questions and local context. Thus we advise researchers interested in utilizing the Hybrid approach to consider conducting qualitative work to inform their list of time use categories and the exact wording used to describe those. This sort of pre-survey qualitative investment is in line with that typically required when developing data collection protocols for concepts that require knowledge of local context to inform instrument design.

Beyond developing a high-performing, easily-implemented time use module targeting low-income populations, our work makes a methodological contribution, demonstrating how researchers can use experimental techniques to assess the quality of data collection methods. The relative strengths and weaknesses of different time use data collection approaches have been difficult to quantify, in large part because different methods are rarely deployed across statistically equivalent populations, complicating comparisons (Kan and Pudney, 2008).[1] We validate the Hybrid module by randomizing the method used to collect time use data during repeated visits to respondents. This allows us to compare responses across alternative time use collection methods, and assess concerns of potential priming effects within the experimental design.

The remainder of the paper proceeds as follows: Section 2 reviews time use data collection techniques, our approach to developing the Hybrid module, and the different time use methods we evaluate. Section 3 describes our study population and validation experiments. Section 4 presents our main results, discussing the module's overall performance in terms of data quality and ease and cost of implementation. Finally, Section 5 concludes.

2 Background: Time Use Data Collection

2.1 Approaches to Time Use Data Collection

While data on time allocation is relevant to research and policy, collecting time use data is complicated and relatively expensive. Table 1 summarizes commonly used time use data collection methods. Self-administered diaries – where the respondent fills a time diary either in real time or retrospectively – are often used in higher income settings as they do not require enumerator time, yield rich data, and can account for multitasking.

Since this method is infeasible for limited literacy populations, time use data collection in low-income countries has often employed stylized survey questions, which ask individuals to aggregate time over a given reference period (e.g., "How much time did you spend cooking yesterday?"). Research on this method suggests it results in higher measurement error compared to detailed diaries, and errors vary systematically by respondent characteristics, such as gender or number of working hours (Kan and Pudney, 2008). Higher errors for female respondents could, in part, reflect how time spent in household work and home-based

[1] A recent paper by Seymour et al. (2020) makes significant improvements on previous approaches to quantify the value of different time use methods, testing approaches across multiple study settings, but lack of experimental variation precludes the authors from drawing the types of conclusions that we can here.

production are not explicitly linked to hours in the way market work is, e.g., through hourly wages or extended contiguous blocks of work during the day. Behavioral issues such as telescoping, availability heuristics, and social desirability bias are also likely to affect stylized questions (Kahneman et al., 2004), and this approach is often more cognitively burdensome for respondents than appreciated (Seymour et al., 2020).

One solution is the assisted retrospective diary method – where an enumerator works with the respondent to fill a time diary covering a reference day. While commonly used in low-income settings, this approach remains cognitively taxing, time consuming, and requires skilled survey personnel (Seymour et al., 2020). Another possibility is observation-based time use surveys, which require enumerators to monitor individual respondents throughout an entire day. However, this approach is much more expensive and may be susceptible to Hawthorne effects. For both these reasons, the latter method is rarely used.

Reflecting these challenges, a recent review of time use data collected by lower and middle-income countries by Hirway (2010) found that approximately half of documented time use data collection exercises were simply pilots or small-scale surveys, and just a handful of countries have collected national time use data more than once. Moreover, little work has been done to quantify measurement error or assess the relative accuracy of different methods of time use data collection.

2.2 Qualitative Research to Inform Module Development

To map individual activities into contextually relevant broad time use groupings in our study population, we first conducted open-ended, semi-structured conversations in which we asked respondents to explain what they had done during the previous day. Our aims were twofold: (i) identify major categories of time use among both men and women, and (ii) identify categories that would enable us to quantify the extent to which women engage in paid and unpaid work, with a focus on how much production took place inside versus outside the home (to better understand mobility). Key focus areas therefore included *where* respondents reported undertaking activities and *for what purpose* activities were undertaken – for production, consumption, or both.

To give one example of how this informed our category construction, we found that women under-reported income-generating activities, describing tasks like caring for livestock purely as household chores, even though the household sometimes earned money by selling output like milk or eggs. We, therefore, developed separate time use categories for pure home production versus home-based work that generated income.

Based on our qualitative findings, we finalized the following broad categories for the Hybrid module: sleeping, working on the household's field, income generating work for regular/daily wage, self-employed income generating activity, household chores/unpaid work outside the house, household chores/unpaid work inside the house, active child or elder care, and relaxing/leisure. These activities could readily be represented using a set of context-specific photos, which helped anchor categories for respondents.[2]

Another important issue this exercise highlighted was the risk of underestimating one important repeated activity for women - caring for children while undertaking other activities. When discussing their time use, women often did not report or even recognize that they had undertaken this "passive" childcare. Given the centrality of caregiving in many women's lives (Ironmonger, 2014) and based on these insights, we ensured

[2]Note that an important consideration in employing photos for this purpose is that the images do not overly restrict respondents' interpretation of the category components.

Table 1: Time Use Survey Methods

Method	How administered	Frequently used in	Advantages	Disadvantages	Survey examples
Observation-based	Enumerator shadowing or observation	– Used infrequently due to cost and complexity	– Avoids retrospective reporting biases – Can be used in populations with less sense of time and low literacy	– Potential for Hawthorne effects – One enumerator per respondent per day – Costly and time consuming enumerator training	– Bangladesh Bureau of Economic Research Survey of Intra-Household Distribution and Poverty Incidence (2004)[1]
Time diaries	Self-reported (real-time or retrospective) or with enumerator assistance (retrospective)	– Stand-alone national time use surveys – Higher-income/ education populations	– Can gather very detailed and comprehensive information – Can account for simultaneous activities – Can reduce measurement error since reported hours must add up to 24	– Detailed versions are time-intensive – Cannot be self-administered by low-literacy populations – Can be prone to social desirability bias[2]	– American Time Use Surveys – Eurostat Time Use Surveys
Stylized questionnaire	Via enumerator	– Module within national household surveys – Low-income settings (e.g., used in women's empowerment in agriculture index[3])	– Easier to administer to populations with less sense of time – Can be tailored to specific types of time use – Can fit into larger household survey	– Cognitive burden can increase time needed to administer[4] – Recall bias; telescoping, social desirability bias, may affect responses – Does not account for simultaneous activities or time of day/chronological order – May over- or under-count time that should add up (e.g., 24 hours of the day)[5]	– Argentina 2001 Survey of Living Conditions – 2005 Bangladesh Household Income and Expenditure Survey – 1998-99 Nicaragua Living Standards Measurement Survey –2002 Mexican Family Life Survey –2016 Young Lives Survey
Experiential sampling methods	Respondents contacted at random intervals and asked to report their activity	– Behavioral surveys – Higher income populations	– Avoids retrospective reporting biases – Can gather measures of subjective well-being alongside time use – Can cover relatively longer time periods than most approaches – Nature of short responses can be less burdensome	– Systematic non-response (by individuals or activities) – Tends to focus on specific episodes rather than paint full picture of time use, or is time-consuming and generates respondent fatigue	German Socio-Economic Panel [6]

Authors synthesis utilizing the following source documents: National Research Council (2000); United Nations Development Programme (2018); Charmes (2015).
[1] As described in Khondker (2006).
[2] Chenu and Lesnard (2006)
[3] Alkire et al. (2013)
[4] Seymour et al. (2020)
[5] Masuda et al. (2014)
[6] Anusic et al. (2017)

that all three time use modules considered included guidelines to capture passive care through enumerator probing and explicit questioning related to this category.

2.3 The Time Use Modules

Our validation exercise compares three time use data collection methods: two survey-based – the Traditional and Hybrid methods – and one observation based, "Gold Standard" method.

Traditional Assisted Diary This approach was modelled on the 1998/99 Indian Time Use Survey conducted by India's Ministry of Statistics and Planning.[3] Enumerators interviewed the assigned respondent about the activities s/he undertook, in chronological order, during the previous day. Respondents could report on up to six separate activities completed within any given hour, allowing for detailed capture of activity duration in minute increments. Enumerators also recorded whether the respondent performed passive childcare during each 15 minute increment,[4] and the location (inside/outside the household) and nature (paid/unpaid) of each activity. Enumerators used 152 categories to classify time use.

Gold Standard Method This approach used a modified version of in-person observations of actual time allocation, essentially amounting to short, high-frequency interviews, to record respondents' time use on the day of the visit. The enumerator visited the assigned respondent every hour throughout the reference day. During each visit, the enumerator asked the respondent what activities s/he had undertaken since the enumerator last visited. Activities, coded using the same 152 categories as above, were coded in minutes, and up to 6 activities could be recorded in a given hour, with passive care available as a cross-cutting simultaneous activity, in line with the Traditional method. The location (inside/outside the household) and nature (paid/unpaid, along with method of payment) of each activity were recorded as well.

On the last visit of the day, enumerators collected information on the respondent's planned rest-of-day activities and associated timing prior to departing for the day.[5] They also collected retrospective data on this period during a subsequent visit the next day. As responses across the prospective and retrospective reports were nearly identical, we use the prospective reports through the rest of the paper.

Each visit throughout the day was brief, averaging 2-3 minutes after the initial introductory visit and prior to the end of day data collection.[6] This approach aimed to reduce measurement error due to recall while limiting disruption in households' schedules and minimizing the possibility that participants altered activities due to the presence of a stranger in the household (Hawthorne effects), giving us as close a measure as possible to their natural time allocation.[7]

Stylized Diary Hybrid Here, respondents narrated the activities they undertook in the previous day chronologically to the enumerator. Enumerators combined and allocated one-hour tokens to final activities

[3] This was the last time such data were widely collected by the government prior to our experiment; the government collected time use data again in 2019, after our experiment was completed.

[4] We did not differentiate whether the passive childcare continued throughout the entirety of any given time use interval.

[5] Enumerators did not conduct observation between the hours of 6 pm and 6 am.

[6] One concern is that respondents may adapt their activities not only due to Hawthorne effects, but also because they anticipate additional visits during the day. Our Gold Standard protocols aimed to minimize this possibility by ensuring enumerators told respondents they should go about whatever activities they intended to do, regardless of location, through the day.

[7] For respondents that left home during the day, we collected information about activities upon their return to ensure we could reconstruct time use for the entire day.

after the day's description was complete. The enumerator allocated 24 tokens to 8 major activities, represented by pictures, which can be found in the Online Appendix.[8] Since each activity was illustrated by a picture placed in front of the respondent, illiterate respondents could easily participate in the exercise.[9] At the end of the exercise, respondents decided whether the token allocations accurately captured activities of the previous day.

In this way, the Hybrid approach shortened module duration by converting respondents' narratives into stylized time use categories, while easing respondent burdens, since the enumerator conducted the classification and kept track of partial hours. Importantly, the module did not require respondents to aggregate time spent on different activities throughout the day or be familiar with standard clock time. The enumerator kept track of activities that took less than a complete hour on a separate notepad, then aggregated and rounded these inputs to activity hours, as applicable, once the respondent had finished their narration. To ensure multitasking where individuals were caring for family members was captured, the enumerator also asked about, aggregated, and reported the total time devoted to passive caregiving for each of the eight activity categories[10]

In sum, all methods captured time use over a 24-hour time period that spanned 6 am to 6 am, although the Traditional and Hybrid methods collected next day retrospective reports, while the Gold Standard collected data on time spent on activities on the day they occurred.

3 Study Sample and Experimental Design

3.1 Study Sample

Our validation study took place in northern Madhya Pradesh, an area characterized by high levels of rural poverty and conservative gender norms. Our respondent sample is drawn from the sampling frame of a randomized controlled trial spanning 197 cluster of villages, known as gram panchayats (GPs), described in Field et al. (2021). Inclusion criteria for the sample frame were that the household must have appeared on India's public workfare program payroll (the Mahatma Gandhi National Rural Employment Guarantee Scheme, MGNREGS), reported work for MGNREGS in the previous year, and had at least one married, unbanked woman. For the time use study, we enrolled households from 13 GPs in Gwalior district.

We recruited respondents using household roster information. As women's roles in the household and society change markedly upon marriage and with advancing age during marriage, we created six strata based on demographic characteristics: unmarried male respondents, unmarried female respondents, male respondents with wives under age 30, married female respondents under age 30, married male respondents with wives over age 30, and married female respondents over age 30. In total we sampled 515 respondents in 212 unique households from these six groups.

[8] The categories were sleeping, working on the household's field, income generating work for regular/daily wage, self-employed income generating activity, household chores/unpaid work outside the house, household chores/unpaid work inside the house, active child or elder care, and relaxing/leisure.

[9] Photos and physical token allocation have been used elsewhere: Masuda et al. (2014) used pictures to help respondents allocate time in short intervals when testing a time use module in Ethiopia, and some country surveys in the 2016 Young Lives survey relied on physical allocation of 24 pebbles to activity categories, but did not use pictures and asked respondents to allocate pebbles to categories themselves.

[10] In practice, this amounted to enumerators asking respondents if they had responsibility for a dependent that no one else was caring for after each described activity.

Demographic characteristics in Table 2 reflect our stratified enrollment protocol: half the sample is female, just over two-thirds are married, and the average age is 31. Overall, sampled individuals are poor, with limited human capital and predominantly belonging to disadvantaged castes. Households have an average of 5.6 members, and reported an average household income of Rs. 6,780 in the past 30 days.[11] Education levels are relatively low, with an average of 6 years. Just over one-third of respondents (two thirds of women) considered themselves housewives, and therefore not in the labor force. Another 16% had some type of household enterprise, and 31% worked as day laborers. While just 11% of respondents were studying, nearly 4 in 10 reported being unemployed. Appendix Table A1 reports demographic characteristics by stratum.

Table 2: Respondent Demographic Characteristics

	Mean	SD	N
Female	0.495	0.500	499
Married	0.673	0.469	499
Age	30.886	12.802	499
Scheduled Caste household	0.432	0.496	474
Scheduled Tribe household	0.057	0.232	474
Other Backward Caste household	0.489	0.500	474
Num. household members	5.551	2.205	497
Num. household members < age 16	1.772	1.448	499
Num. household members < age 7	0.651	0.935	499
Household Income (Past 30 Days)	6778.556	7531.934	450
Years of Education	5.964	4.348	496
Housewife	0.342	0.475	488
Self-Employed	0.156	0.363	488
Day Laborer	0.314	0.464	488
Wage Worker	0.031	0.173	488
Unemployed	0.389	0.488	488
Student	0.111	0.314	488

Source: 2018 Madhya Pradesh time use survey.

Figure 1 illustrates how time allocation among women in our sample (measured by the Traditional module) compared to national averages for rural women (per the 2019 India Time Use Survey, collected one year after our exercise). Overall, patterns are broadly similar, with women spending most of their non-sleeping time on either leisure (5-7 hours) or chores (also 5-7 hours). Compared to national data, our sampled women spend less time on leisure, more time on indoor chores, more time on self employment, and less time sleeping.

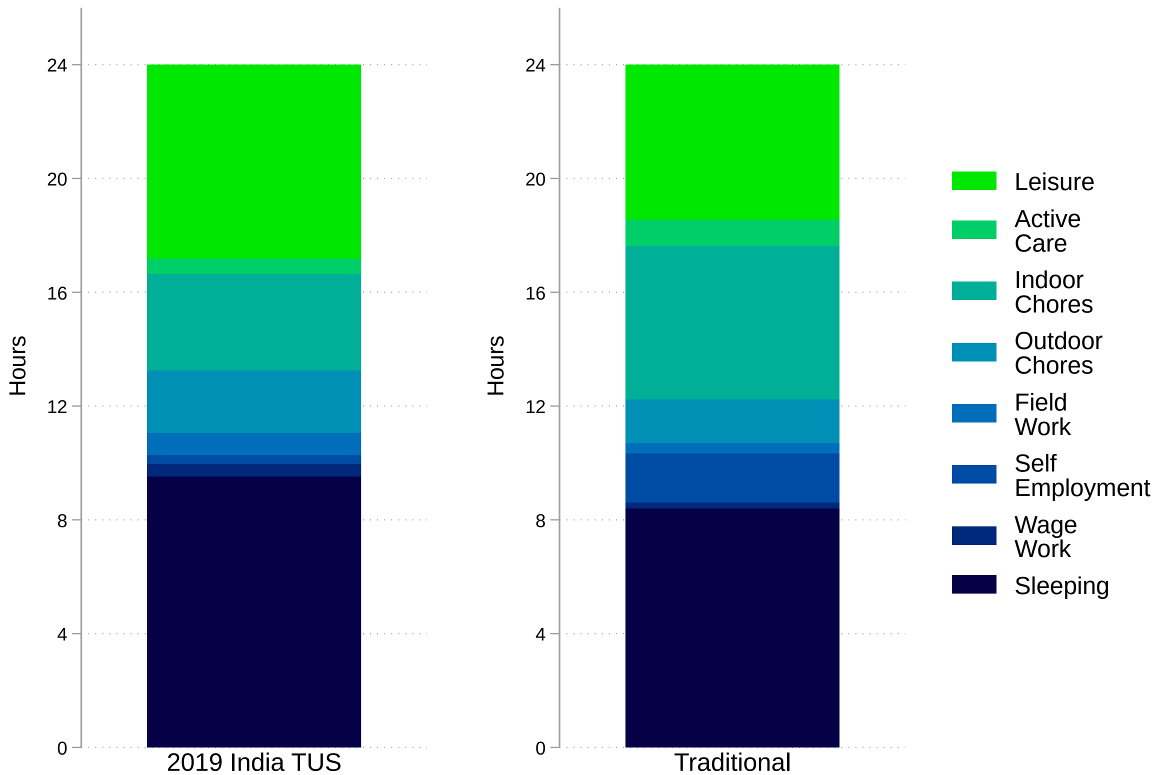
3.2 Study Design

Our validation study was conducted in January and February 2018. All study participants were visited thrice using the protocol described below:

- **Visit 1-Reference Day 1:** Respondent time use was recorded using the Gold Standard approach.

[11]This amount has purchasing power of approximately $321 USD respectively at the time of the survey in 2018, given a PPP exchange rate of Rs. 21 per USD.

Figure 1: Average Time Use of Rural Females: 2019 India Time Use Survey and Traditional Method Data



Data Sources: 2019 India Time Use Survey, and 2018 Madhya Pradesh time use survey, traditional method. Data restricted to rural women.

- **Visit 2-Reference Day 1:** The next day, an enumerator visited the respondent and administered either the Traditional or Hybrid module, based on random assignment, which was stratified by community (GP) and demographic group (unmarried women, married women under 30, married women over 30, unmarried men, men married to a woman under 30, men married to a woman over 30). Respondents reported on the previous day's (reference day 1) activities, enabling within-subject comparison of the retrospective data from visit 2 compared to reference day 1 Gold Standard data. During visit 2, we also collected basic demographic information.[12]

- **Visit 3-Reference Day 2:** This visit occurred at least one week after visit 2. The enumerator administered one of the three time use methods, again based on random assignment, stratified on community, demographic group, and the method assigned in visit 2. An important motivation for this visit was to account for the possibility that "priming" subjects with the Gold Standard method on visit 1 improved recall during visit 2, which could influence measured performance differences between the Traditional and Hybrid approaches.[13]

Respondents were paid small incentives (less than \$1) after the second and third visits, and attrition was low. Of the 515 individuals contacted, 499 completed the first two visits, while 497 completed visit 3, resulting in an attrition rate of less than 1%.

To compare data across time use methods, we assigned each of the 152 unique activity codes to one of the 8 Hybrid module categories. The detailed mapping of categories is found in Table 2 in the Online Appendix. Appendix Table A2 reports differences in demographic characteristics by random assignment of data collection methods for both visit 2 and visit 3. Overall, the randomization achieved balance on demographic characteristics, as intended.

4 Comparing Time Use Modules

4.1 Activity measurement

A. Average effects

We measure time use on reference day 1 for individual i through two visits (denoted as v). Visit $v = 1$ used the Gold Standard for all respondents and forms the baseline against which we compare the performance of the module (either Traditional and Hybrid) administered on visit 2. To make within-person comparisons, we estimate:

$$y_{i,v} = \beta_1 Trad_{i,v} + \beta_2 Hybrid_{i,v} + \delta_i + \epsilon_{i,v} \tag{1}$$

where $y_{i,v}$ is the outcome of interest (i.e. hours spent on a particular activity), $Trad_{i,v}$ and $Hybrid_{i,v}$ are dummy variables for whether data was collected using either the Traditional or Hybrid module (these variables are always equal to zero on visit 1), δ_i are individual fixed effects, and ϵ_{i_v} is an error term. We cluster standard errors at the individual level.

[12]These data were not collected on the first visit since the time required to collect them would have directly interfered with respondents' time allocations.

[13]Details on sample sizes in strata×time use method×visit cells are in Appendix Table A3.

Figure 2 graphs reference day 1 time spent on the 8 Hybrid activity categories and passive care, as measured by the three methods. The first bar in each panel graphs mean time recorded per the Gold Standard. The next two bars show regression-adjusted means for the Traditional and Hybrid modules respectively, using coefficients from specification 1. Table A4 in the Appendix reports point estimates and standard errors. Since all methods captured time use for the same day, small, insignificant differences relative to the Gold Standard indicate good performance.

The Hybrid module performs well – differences relative to the Gold Standard are generally small in magnitude, and none are significantly different from zero. The Traditional module also does reasonably well in all categories of time use except for passive care. Relative to the Gold Standard method, the Traditional module significantly overestimates time spent on passive care by 0.55 hours, which is substantial relative to the mean of 0.78 hours per the Gold Standard method. As it also under-estimates leisure by 0.40 hours (a modest error relative to the Gold Standard mean of 6.3 hours), it appears that the Traditional module tends to classify some pure leisure erroneously as passive care. The slightly better performance of the Hybrid module is especially striking given that the Traditional and Gold Standard survey instruments used the same time use categories and coding.

B. Heterogeneity by Demographic group

Module performance may vary by demographic group, however, especially since we designed the categories for the Hybrid module with women in mind. To investigate this, we re-estimate equation 1 within each of our six demographic strata.

Tables A5 and A6 in the Appendix report within-person deviations from the Gold Standard for the Hybrid and Traditional modules by demographic strata. There are no significant differences between Hybrid and Gold Standard reports for any of the female strata. Point estimates are also small in magnitude relative to Gold Standard means, with the exception of some estimates for unmarried women (Table A5). The Traditional module performs well across groups – the only significant difference (at the 10 percent level) is for time spent sleeping among unmarried women. These results also hold for passive care, which would normally be classified as multitasking, for married women under 30, those with the highest levels of this activity.[14] Table A6 shows that both modules also perform well for married males.

However, the Hybrid method misclassifies chores as leisure among unmarried men. Chores in the home are under-estimated by 1.6 hours, which is substantial compared to a mean of 2.2 hours per the Gold Standard, while leisure is overestimated by 2.1 hours relative to a Gold Standard mean of 7.5 hours. This result could reflect social desirability bias, e.g., if young men failed to mention chores when describing their day in broad strokes, or if they pushed back against the number of tokens initially allocated to household work.

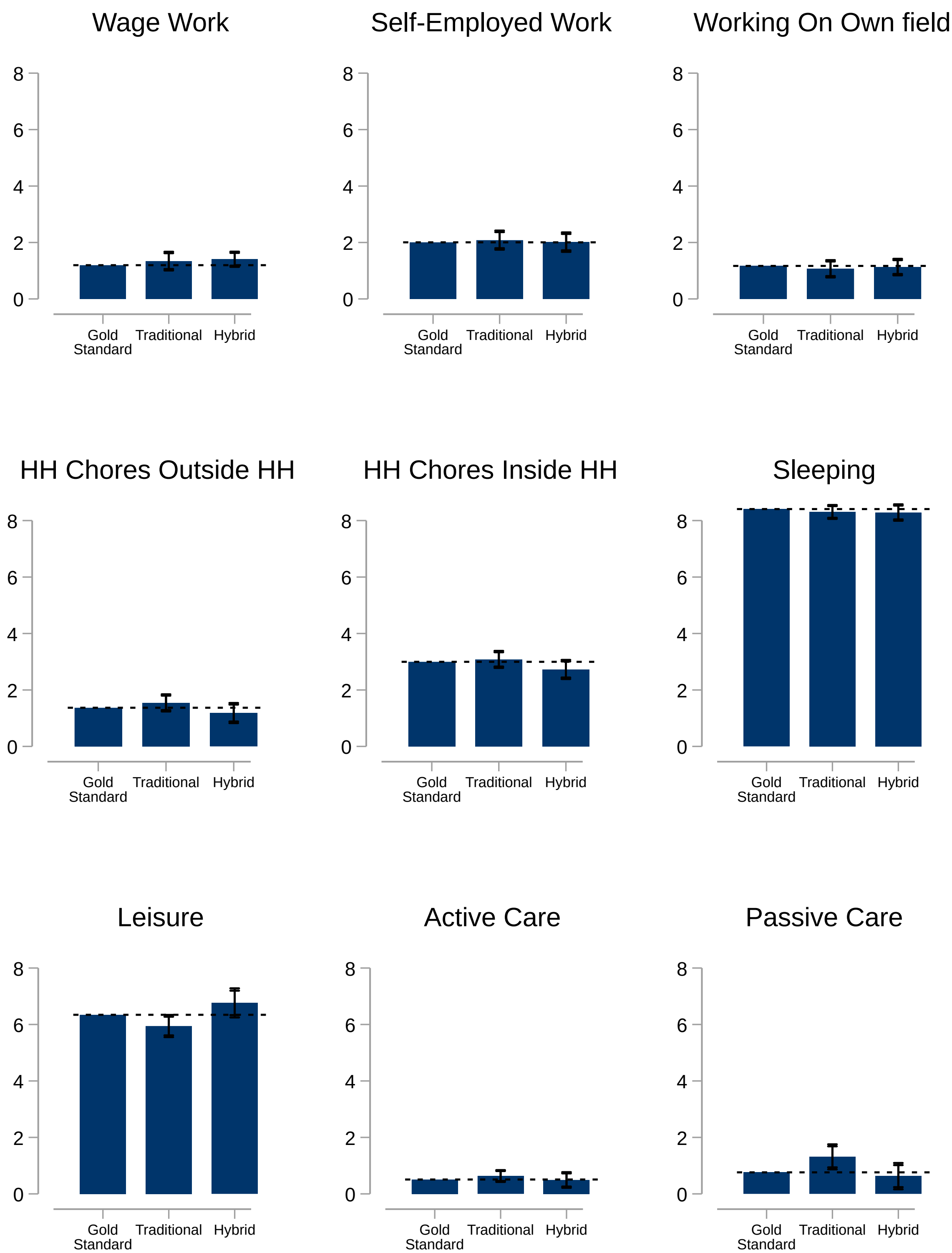
C. Performance by Activity Duration

By recording time use in hourly intervals, the Hybrid module may fail to capture shorter duration activities. Among our respondents, activities most likely to fall in this short-duration category[15] were active care (where

[14] Interestingly, we find less evidence of multitasking as a female-specific issue overall than is popularly thought: The average respondent reported undertaking a total of 31.5 activities in the Gold Standard first visit, and females reported a statistically insignificant .47 more activities that day (p-value = .140). Of course, women's lack of market engagement could depress the amount of activities undertaken in a day.

[15] Based on terciles, excluding responses of zero.

Figure 2: Reference Day 1 Module Comparison



Outcomes reported in hours. The first bar in each panel represents the Gold Standard mean for the indicated outcome, and the following bars are regression adjusted means using coefficients from equation 1. Sample restricted to Reference Day 1. Whiskers display 90 and 95 percent confidence intervals based on robust standard errors clustered at the individual level.

84% of these reports were classified as low duration), and outdoor household chores (at 71%). Note that the use of tokens in the Hybrid approach is what constrains reporting to larger intervals, as allocating tokens representing smaller time intervals was deemed too complex a protocol.

To investigate this concern, we limit attention to reference day 1 and create three activity-level measures of method performance *relative* to the Gold Standard. First, the difference in time recorded for activity a using method $m \in \{\text{Hybrid}, \text{Traditional}\}$ relative to the Gold Standard. Second, to measure overreporting we consider activities for which time spent per the Gold Standard is zero and construct a dummy equal to one when method m records positive time spent on a. Finally, to measure underreporting, we consider activities in which the Gold Standard records positive time spent and construct a dummy variable equal to one when method m fails to record time on a. We stack the data at the individual $i \times$ activity a level and estimate:

$$y_{i,a} = \alpha Hybrid_i + \zeta_a + \psi_s + \epsilon_{i,a} \tag{2}$$

where $y_{i,a}$ is the outcome of interest, $Hybrid_i$ is an indicator variable for whether individual i was administered the Hybrid method on visit 2, ζ_a are activity category fixed effects, ψ_s are strata fixed effects, and $\epsilon_{i,a}$ is an error term clustered at the individual level. The omitted group here is the set of respondents randomly assigned to receive the Traditional module during visit 2, so α should be interpreted as the difference in recording error relative to the Traditional module.

Results are reported in Table 3. In column (1) we limit the sample to zero time activities per the Gold Standard and examine the size of reporting differences from the Gold Standard benchmark. The Hybrid module is slightly more accurate here - it reports a marginally significant 0.1 hours less than the Traditional module for categories *not* recorded in the Gold Standard surveys. Columns (2)-(4) consider the sample of activities with positive Gold Standard time reported, and divide the sample into low, medium and high duration activities (based on terciles of Gold Standard duration, excluding zero responses). The Hybrid module intensive margin difference from the Gold Standard duration does not differ statistically from the Traditional module across any of the three terciles.

Column (5) shows that the Traditional approach reported positive time in a category even when zero time was reported in the Gold Standard for reference day 1 in 14% of individual$\times$activity categories. The coefficient on the Hybrid module shows that this method was 5 percentage points less likely to overreport an activity on the extensive margin, consistent with column (1), again highlighting the module is less likely to overreport relative to the Traditional module.

In contrast, columns (6)-(8) highlight that the Hybrid approach is more likely to miss low and medium time intensity activities. The outcome variable here equals 1 if the Visit 2 report captured no time for the category. The magnitude of underreporting is practically large for the low and medium duration Gold Standard-reported activities, increasing underrerporting from 22% in the Traditional module to just over 35% of relevant categories with positive, but low, time reported. Underreports for the middle duration tercile are still significant and sizable compared to the Traditional approach underreports, amounting to 13% of these activities being underreported overall, compared to 5% of activities being underreported by the Traditional methods. These relatively high rates of underreporting could reflect recall difficulties on the part of our sample, or – for low-duration activities – correct classification for activities of very short duration that never round up to one hour (token) using the Hybrid approach. Underreports for high intensity activities are, by contrast, rare.

In sum, underreporting in the Hybrid module largely occurs on the extensive margin. In contrast, the Hybrid approach performs slightly better than the Traditional module at avoiding overreporting activity categories not captured in the Gold Standard visit. Given these results, researchers should carefully consider the extent to which low duration activities are important to their aims when determining whether to utilize an approach similar to the Hybrid – i.e., with broad category groupings and larger time increments – trialled here. Moreover, future field testing could explore the use of tokens representing 15- or 30-minute intervals in place of the 60-minute tokens used in this study when low-duration activities are of particular interest.

Table 3: Hybrid Performance by Duration of Activities

	Gold Standard Difference				Overreport	Underreport		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$GS = 0$	$GS =$ Low	$GS =$ Mid	$GS =$ High	$GS = 0$	$GS =$ Low	$GS =$ Mid	$GS =$ High
α: Hybrid Module	-0.095*	0.072	-0.037	0.134	-0.051***	0.137***	0.081***	0.000
	(0.052)	(0.108)	(0.128)	(0.156)	(0.015)	(0.031)	(0.018)	(0.010)
Traditional Mean	0.329	0.236	-0.182	-0.787	0.141	0.217	0.052	0.016
N	1672	776	800	744	1672	776	800	744

The top column headers denote variable outcomes, and the second level specifies the sample included for that regression, based on Gold Standard Visit 1 reports *(GS)*. Outcomes in columns 1 through 4 are reported in hours. Gold Standard sample restriction in columns 2 - 4 are terciles after excluding reports for that respondent-activity category equal to 0. Sample restricted to Reference Day 1. All regressions include strata and activity category fixed effects. Standard errors clustered at individual level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01.

4.2 Is respondent priming a threat to our validation method?

If participating in the Gold Standard exercise improved respondents' ability to recall activities the next day, then reference day 1 within-person differences between the Gold Standard and other approaches may be biased towards zero. To test for such a priming effect, we examine the relative performance of the three methods on reference day 1 as compared to reference day 2. With priming, we would expect reference day 2 differences from the Gold Standard to be larger than reference day 1 differences.

Specifically, we compare method-specific reports *relative* to the Gold Standard for reference day 1 - when the Gold Standard was conducted the day prior to the Traditional or Hybrid data collection - and reference day 2, where comparisons are across person. Using data from all three visits, we estimate an augmented version of equation 1:

$$\begin{aligned} y_{i,v} = {} & \gamma_1 Trad_{i,v} + \gamma_2 Trad_{i,v} \times Ref Day\, 2_v + \gamma_3 Hybrid_{i,v} + \\ & \gamma_4 Hybrid_{i,v} \times Ref Day\, 2_{i,v} + \gamma_5 Ref Day\, 2_{i,v} + \xi\, Typical\ Day_{i,v} + \delta_i + \epsilon_{i,v} \end{aligned} \tag{3}$$

Here, $Ref Day\, 2_v$ is an indicator variable for reference day 2 and $Typical\ Day_{i,v}$ is a control for whether the respondent reported the day of interest was a typical day. δ_i are individual fixed effects, and ϵ_{i_v} is an error term. We again cluster standard errors at the individual level. The sign and significance of the interaction terms lets us assess whether the quality of reporting relative to the Gold Standard deteriorated for reference day 2.

Table 4 reports results. Coefficients on the reference day interaction terms are largely insignificant, with the exception of leisure for the Traditional method. The coefficients on the Hybrid×reference day 2 interaction are all statistically insignificant and generally small compared to the Gold Standard mean, suggesting our within-person comparisons are not biased by priming or recall effects from the Day 1 Gold Standard visit.

Table 4: All Visits Pooled Comparisons

	(1) Wage Work	(2) Self Employed	(3) Working Own Field	(4) HH Chores Outside HH	(5) HH Chores Inside HH	(6) Sleeping	(7) Leisure	(8) Active Care	(9) Passive Care
γ_1: Traditional Module	0.250*	0.086	-0.129	0.178	0.116	-0.144	-0.530***	0.173*	0.420**
	(0.147)	(0.145)	(0.124)	(0.138)	(0.131)	(0.108)	(0.179)	(0.097)	(0.196)
γ_2: Traditional Module × Day 2	-0.561	0.158	0.221	0.390	0.234	0.234	-0.862*	0.186	0.508
	(0.437)	(0.329)	(0.305)	(0.295)	(0.304)	(0.255)	(0.474)	(0.243)	(0.368)
γ_3: Hybrid Module	0.130	0.004	-0.016	-0.199	-0.291**	-0.067	0.504**	-0.064	-0.002
	(0.137)	(0.150)	(0.138)	(0.155)	(0.148)	(0.120)	(0.220)	(0.107)	(0.189)
γ_4: Hybrid Module × Day 2	0.017	0.096	0.214	-0.353	-0.145	0.102	-0.168	0.237	0.589
	(0.356)	(0.339)	(0.357)	(0.318)	(0.317)	(0.279)	(0.518)	(0.230)	(0.366)
γ_5: Day 2	0.472*	-0.163	-0.149	-0.048	-0.342*	-0.216	0.588*	-0.141	-0.129
	(0.273)	(0.219)	(0.241)	(0.207)	(0.197)	(0.190)	(0.340)	(0.177)	(0.185)
Gold Standard Mean - Day 1	1.098	1.973	1.150	1.387	3.105	8.463	6.309	0.514	0.781
N	1495	1495	1495	1495	1495	1495	1495	1495	1495

Column headers denote variable outcomes, reported in hours. Sample includes both Reference Day 1 and 2. All regressions are as specified in equation 3, including individual fixed effects and an indicator for whether it was a typical day. Standard errors clustered at individual level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01.

4.3 Comparing module implementation

It is also important to compare the cost and ease of implementation of the Hybrid and Traditional modules. We are less interested in comparisons relative to the Gold Standard, as this method is costly and poorly suited to wide-scale implementation.

The Hybrid method performs better than the Traditional module in terms of the total cost per survey, which is calculated inclusive of training costs and materials: The Hybrid approach cost less than $6 per survey, compared to $10 per survey for the Traditional method. Particularly when conducted with large samples, this 40% cost savings is substantial.

Lower cost was partially reflected in the Hybrid module's shorter duration, which also limited respondent fatigue and cognitive burden. Figure 3 plots survey duration for reference day 2 for the Traditional and Hybrid modules. Average completion time for the Traditional module was 14 minutes, while the Hybrid method took 9 minutes. Time differences at the right end of the distribution are higher: for example, at the 90th percentile, the difference is 6 minutes, and it is 8 minutes at the 95th percentile. While the mean difference is a modest absolute time savings of 5 minutes, this amounts to a 33 percent reduction in module time, and could provide a valuable time reduction in a long, multi-module survey.

Figure 3: Distribution of Survey Duration by Method

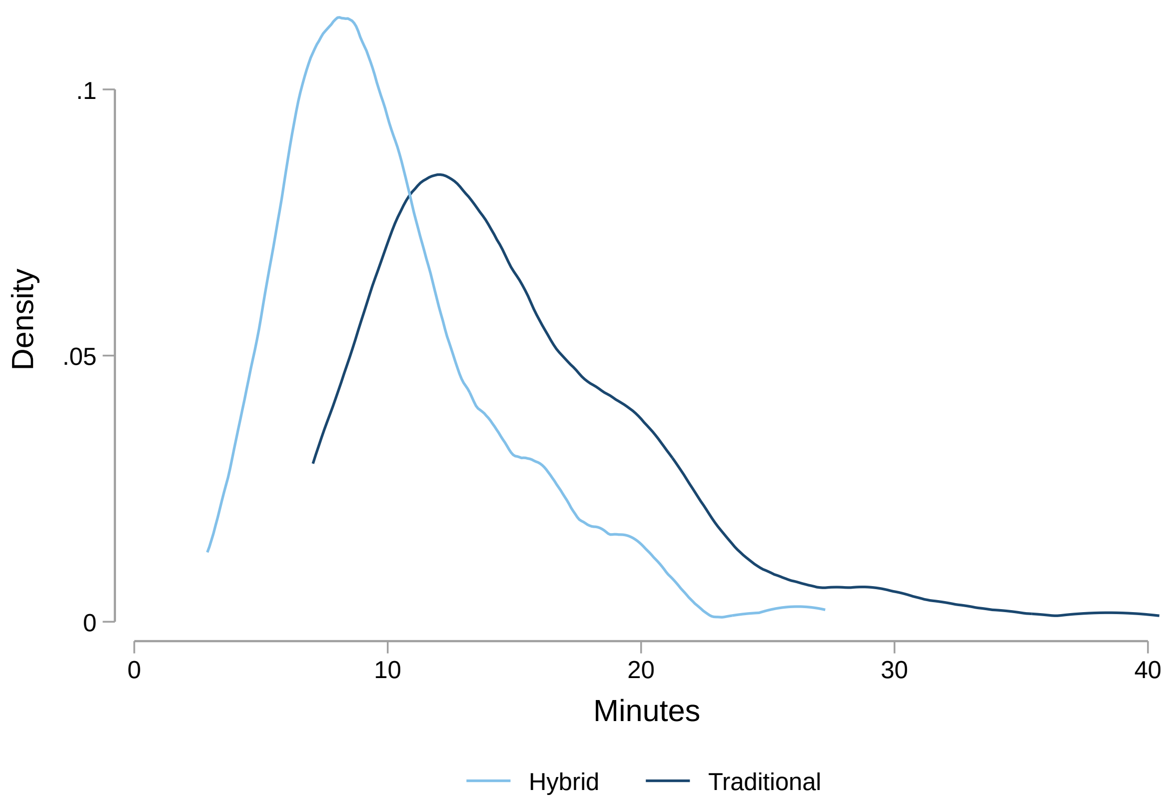


Figure shows kernel density plot of survey duration. Sample restricted to Reference Day 2.

Another important consideration is the time needed to train enumerators to administer the module.

Preparation for the Traditional method included detailed training to ensure enumerators could categorize time use in 152 distinct activity categories. We needed three full training days to prepare enumerators to classify these activities accurately. The Hybrid method, in contrast, only required one day of training to ensure enumerators could correctly categorize the eight activity codes with ease.

Crucially, the Hybrid module was the easier method to administer. The cards with photos and tokens helped respondents engage with the method, and the lack of extensive probing and overall module brevity ensured respondents and enumerators were not fatigued by the exercise. In contrast, the detailed probing required to classify activities sometimes taxed respondents administered the Traditional (and Gold Standard) modules. For enumerators, the most complicated part of the Hybrid method was calculating "left over" minutes. If respondents reported undertaking an activity for 30 minutes in the morning, for example, and then 20 minutes later in the day, the enumerator needed to set aside these minutes and then sum them at the end of the exercise to determine whether to allocate an hour to that activity or not.

5 Discussion

Our novel Hybrid time use module, in which respondents narrate their days and enumerators assist respondents in allocating time to a limited number of stylized time use categories, accurately captures less educated respondents' time use in a poor, rural setting. Compared to a widely-used assisted retrospective diary approach, the Hybrid method is also less expensive and requires less time to train enumerators, while limiting respondent fatigue and cognitive burden.

The Hybrid module requires the researcher to identify a concise set of activity categories and does not capture *when* in the reference period activities occur. Given this, more traditional approaches would be better suited for research that requires substantial detail on how respondents spend their days. And while the module does a good job capturing average time use across activity categories, it is relatively less accurate at registering short duration activities on the extensive margin. On the other hand, the Hybrid module is less likely to generate "false positives", i.e., record time in a high-level activity the respondent did not actually undertake. While it is not designed to capture simultaneous activities, it can accurately gather information on multitasking relevant to a particular activity of interest – in our case, passive care.

To minimize underreporting and maximize data quality, it is therefore important for the researcher to carefully select activity categories. Since time use can vary across populations and demographic groups, we recommend interested researchers first conduct targeted qualitative work, similar to that described here, to understand how their study population spends its time, and use this information to construct Hybrid categories.

Additional testing of the Hybrid approach in other settings would help build understanding of the extent to which our results are externally valid. Another important avenue for extension would be to combine a version of the Hybrid method with approaches from psychology designed to understand respondents' perceptions of well-being as they undertake specific activities, in line with experiential time use approaches, as in Kahneman et al. (2004).

Finally, our randomized validation design is informative for researchers interested in trialling novel approaches to collecting time use data in resource-constrained settings. Systematically incorporating more high-quality time use data collection into major national surveys in emerging economies would be valuable: Doing so, for example, would significantly improve researchers' ability to understand how households' labor allocation – particularly that of women – evolves over the course of economic development, and how this is shaped by contextual factors like institutions and social norms.

References

Alkire, S., R. Meinzen-Dick, A. Peterman, A. Quisumbing, G. Seymour, and A. Vaz (2013). The womens empowerment in agriculture index. *World Development 52*, 71–91.

Anusic, I., R. E. Lucas, and M. B. Donnellan (2017). The validity of the day reconstruction method in the german socio-economic panel study.

Charmes, J. (2006). *A Review of Empirical Evidence on Time Use in Africa from UN-Sponsored Surveys.*, Chapter 3, pp. 39–74. World Bank.

Charmes, J. (2015). Time use across the world: Findings of a world compilation of time use surveys. *UNDP Human Development Report Office*.

Chenu, A. and L. Lesnard (2006). Time use surveys: A review of their aims, methods, and results. *Archives européennes de sociologie. European journal of sociology. 47*(3), 335–359.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021, July). On her account: Can strengthening women's financial control boost female labor supply? *American Economic Review*.

Hirway, I. (2010). *Time-Use Surveys in Developing Countries: An Assessment*, pp. 252–324. London: Palgrave Macmillan UK.

Ironmonger, D. (2014, September). There are only 24 hours in a day! solving the problematic of simultaneous time. Presented at the 25th IATUR Conference on Time Use Research.

Kahneman, D., A. Krueger, D. Schkade, N. Schwarz, and A. Stone (2004). A survey method for characterizing daily life experience: the day reconstruction method. *Science 306*(5702), 17761780.

Kan, M. Y. and S. Pudney (2008). 2. measurement error in stylized and diary data on time use. *Sociological Methodology 38*(1), 101–132.

Khondker, B. H. (2006, September). Capturing intra-household distribution and poverty incidence: A study on bangladesh. Technical report, University of Dhaka.

Masuda, Y. J., L. Fortmann, M. K. Gugerty, M. Smith-Nilson, and J. Cooka (2014). Pictorial approaches for measuring time use in rural ethiopia. *Social indicators research 115*(1), 467–482.

National Research Council (2000). Time-use measurement and research: Report of a workshop. *The National Academies Press*.

Seymour, G., H. Malapit, and A. Quisumbing (2020). Measuring time use in developing country agriculture: Evidence from bangladesh and uganda. *Feminist Economics 26*(3), 169–199.

United Nations Development Programme, I. (2018). *Time-use surveys and statistics in Asia and the Pacific: A review of challenges and future directions*. ILO Regional Office for Asia and the Pacific.

World Bank Group (2021). Labor force participation rate, female (% of female population ages 15+) (Modeled ILO estimate) - India. World Bank Data. `https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?end=2019&locations=IN&start=1998&view=chart`, Accessed on 2021-07-12.

A Appendix Tables and Figures

Table A1: Respondent Demographic Characteristics by Gender

	Unmarried 16-30			Married Under 30			Married Over 30		
	(1) Female Mean	(2) Male Mean	(3) Difference	(4) Female Mean	(5) Male Mean	(6) Difference	(7) Female Mean	(8) Male Mean	(9) Difference
Married†	0.000	0.038	-0.038*	0.943	0.954	-0.011	0.965	0.988	-0.023
			(0.022)			(0.034)			(0.023)
Age	18.054	19.329	-1.275***	25.908	29.138	-3.230***	42.070	48.163	-6.093***
			(0.367)			(0.808)			(1.402)
Scheduled Caste household	0.493	0.405	0.087	0.390	0.390	-0.000	0.463	0.463	-0.000
			(0.083)			(0.077)			(0.078)
Scheduled Tribe household	0.030	0.063	-0.033	0.073	0.073	0.000	0.049	0.049	0.000
			(0.035)			(0.041)			(0.034)
Other Backward Caste household	0.448	0.506	-0.059	0.512	0.512	0.000	0.476	0.476	0.000
			(0.083)			(0.079)			(0.078)
Num. household members	5.865	4.392	1.472***	6.184	6.471	-0.287	5.119	5.198	-0.079
			(0.279)			(0.368)			(0.313)
Num. household members < age 16	1.703	0.835	0.867***	2.471	2.437	0.034	1.523	1.558	-0.035
			(0.195)			(0.217)			(0.207)
Num. household members < age 7	0.392	0.089	0.303***	1.218	1.195	0.023	0.465	0.453	0.012
			(0.093)			(0.165)			(0.113)
Household Income	5872.917	7657.971	-1785.054	6904.762	7084.756	-179.994	6604.573	6604.573	-0.000
			(1295.090)			(1314.248)			(1132.885)
Years of Education	8.054	10.215	-2.161***	4.563	7.989	-3.425***	0.855	4.558	-3.703***
			(0.432)			(0.555)			(0.435)
Housewife	0.479	0.000	0.479***	0.839	0.012	0.827***	0.707	0.012	0.695***
			(0.060)			(0.041)			(0.052)
Self-Employed	0.085	0.141	-0.057	0.057	0.198	-0.140***	0.122	0.321	-0.199***
			(0.052)			(0.050)			(0.063)
Day Laborer	0.014	0.179	-0.165***	0.103	0.733	-0.629***	0.159	0.631	-0.472***
			(0.046)			(0.058)			(0.067)
Wage Worker	0.000	0.115	-0.115***	0.000	0.035	-0.035*	0.012	0.024	-0.012
			(0.036)			(0.020)			(0.021)
Unemployed	0.732	0.038	0.694***	0.839	0.023	0.816***	0.707	0.024	0.684***
			(0.057)			(0.043)			(0.053)
Student	0.169	0.526	-0.357***	0.000	0.012	-0.012	0.000	0.000	0.000
			(0.072)			(0.012)			(0.000)
N	74	79		87	87		87	87	

Robust standard errors clustered at the individual level in parentheses. * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$. † Marital status here refers to marital status collected at the time of the time use data collection exercise, while column headers refer to baseline marital status from the RCT that collected this information that was used as a demographic strata for our sample.

Table A2: Demographic Balance

	Visit 2 Subsamples				Visit 3 Subsamples						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Traditional Method Mean	Hybrid Method Mean	P-value: Traditional vs. Hybrid	N	Gold Standard Method Mean	Traditional Method Mean	Hybrid Method Mean	P-Value: Gold Standard vs. Traditional	P-Value: Gold Standard vs. Hybrid	P-Value: Traditional vs. Hybrid	N
Female	0.492	0.498	0.894	499	0.518	0.473	0.500	0.411	0.743	0.621	497
	(0.032)	(0.032)			(0.039)	(0.039)	(0.039)				
Married	0.684	0.663	0.612	499	0.651	0.685	0.681	0.510	0.562	0.936	497
	(0.029)	(0.030)			(0.037)	(0.036)	(0.036)				
Age	31.100	30.671	0.708	499	30.151	31.236	31.283	0.446	0.406	0.974	497
	(0.842)	(0.779)			(0.946)	(1.064)	(0.978)				
Scheduled Caste household	0.445	0.419	0.570	474	0.399	0.456	0.449	0.308	0.372	0.902	472
	(0.032)	(0.032)			(0.039)	(0.040)	(0.040)				
Scheduled Tribe household	0.071	0.042	0.173	474	0.082	0.063	0.026	0.517	0.026**	0.106	472
	(0.017)	(0.013)			(0.022)	(0.019)	(0.013)				
Other Backward Caste household	0.458	0.521	0.169	474	0.481	0.456	0.526	0.653	0.431	0.216	472
	(0.032)	(0.033)			(0.040)	(0.040)	(0.040)				
Num. household members	5.440	5.664	0.258	497	5.409	5.582	5.590	0.452	0.448	0.972	495
	(0.130)	(0.150)			(0.160)	(0.166)	(0.178)				
Num. household members < age 16	1.752	1.791	0.763	499	1.596	1.903	1.795	0.060*	0.186	0.509	497
	(0.089)	(0.095)			(0.106)	(0.123)	(0.107)				
Num. household members < age 7	0.676	0.627	0.555	499	0.524	0.721	0.699	0.051*	0.085*	0.832	497
	(0.061)	(0.058)			(0.068)	(0.074)	(0.075)				
Household Income	6459.389	7086.572	0.377	450	6469.205	6637.586	7249.671	0.845	0.372	0.491	448
	(490.617)	(512.701)			(599.511)	(619.491)	(634.891)				
Years of Education	5.904	6.024	0.758	496	6.521	5.788	5.572	0.128	0.047**	0.653	494
	(0.274)	(0.278)			(0.339)	(0.342)	(0.336)				
Housewife	0.332	0.352	0.634	488	0.348	0.370	0.307	0.674	0.432	0.227	486
	(0.030)	(0.031)			(0.038)	(0.038)	(0.036)				
Self-Employed	0.156	0.156	1.000	488	0.137	0.136	0.196	0.982	0.150	0.143	486
	(0.023)	(0.023)			(0.027)	(0.027)	(0.031)				
Day Laborer	0.320	0.307	0.770	488	0.329	0.290	0.319	0.449	0.845	0.573	486
	(0.030)	(0.030)			(0.037)	(0.036)	(0.037)				
Wage Worker	0.020	0.041	0.191	488	0.012	0.037	0.043	0.155	0.094*	0.787	486
	(0.009)	(0.013)			(0.009)	(0.015)	(0.016)				
Unemployed	0.389	0.389	1.000	488	0.391	0.414	0.362	0.684	0.587	0.341	486
	(0.031)	(0.031)			(0.039)	(0.039)	(0.038)				
Student	0.115	0.107	0.773	488	0.130	0.123	0.080	0.851	0.138	0.194	486
	(0.020)	(0.020)			(0.027)	(0.026)	(0.021)				

Robust standard errors in parentheses. * $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A3: Sample Selection

Category	**N**		
Married Women 16-30	87		
Married Women 30+	86		
Unmarried Women 16-30	74		
Man married to woman 16-30	87		
Man married to Woman 30+	86		
Unmarried Man 16-30	79		
Visit 2	**Traditional**	**Hybrid**	
Married Women 16-30	42	45	
Married Women 30+	44	42	
Unmarried Women 16-30	37	37	
Man married to woman 16-30	45	42	
Man married to Woman 30+	43	43	
Unmarried Man 16-30	39	40	
Visit 3	**Traditional**	**Hybrid**	**Gold Standard**
Married Women 16-30	31	27	29
Married Women 30+	26	28	32
Unmarried Women 16-30	21	31	22
Man married to woman 16-30	28	27	30
Man married to Woman 30+	29	29	28
Unmarried Man 16-30	30	24	25

Table A4: Reference Day 1 Comparisons

	(1) Wage Work	(2) Self Employed	(3) Working Own Field	(4) HH Chores Outside HH	(5) HH Chores Inside HH	(6) Sleeping	(7) Leisure	(8) Active Care	(9) Passive Care
β_1: Traditional Module	0.145	0.077	-0.102	0.172	0.085	-0.103	-0.399**	0.124	0.549**
	(0.170)	(0.173)	(0.157)	(0.155)	(0.154)	(0.126)	(0.201)	(0.107)	(0.226)
β_2: Hybrid Module	0.210	0.007	-0.041	-0.188	-0.271	-0.122	0.422	-0.016	-0.134
	(0.138)	(0.176)	(0.149)	(0.182)	(0.173)	(0.149)	(0.263)	(0.141)	(0.237)
P-Value: $\beta_1 = \beta_2$	0.767	0.776	0.779	0.134	0.124	0.922	0.013**	0.430	0.038**
Gold Standard Mean	1.098	1.973	1.150	1.387	3.105	8.463	6.309	0.514	0.781
N	998	998	998	998	998	998	998	998	998

Column headers denote variable outcomes, reported in hours. Sample restricted to Reference Day 1. All regressions are as specified in equation 1, including individual fixed effects. Standard errors clustered at individual level in parentheses. * p< 0.10, ** p< 0.05, *** p< 0.01.

Table A5: Reference Day 1 Comparisons (By Demographic Group)

	(1) Wage Work	(2) Self Employed	(3) Working Own Field	(4) HH Chores Outside HH	(5) HH Chores Inside HH	(6) Sleeping	(7) Leisure	(8) Active Care	(9) Passive Care
Unmarried daughter									
β_1: Traditional Module	-0.128	0.077	0.047	0.135	-0.086	-0.383*	0.189	0.149	0.885
	(0.327)	(0.501)	(0.296)	(0.474)	(0.484)	(0.218)	(0.454)	(0.140)	(0.661)
β_2: Hybrid Module	0.500	-0.529	0.135	-0.561	0.131	-0.095	0.412	0.007	-0.084
	(0.421)	(0.472)	(0.112)	(0.390)	(0.427)	(0.269)	(0.592)	(0.146)	(0.208)
Gold Standard Mean	0.233	1.323	0.105	1.724	4.965	8.977	6.459	0.213	0.217
N	148	148	148	148	148	148	148	148	148
Married woman under 30									
β_1: Traditional Module	0.107	0.163	-0.131	0.246	0.516	0.133	-0.688	-0.345	1.526
	(0.152)	(0.183)	(0.254)	(0.305)	(0.522)	(0.440)	(0.578)	(0.326)	(1.011)
β_2: Hybrid Module	-0.098	0.211	-0.052	0.333	-0.356	-0.344	0.219	0.087	-1.759
	(0.182)	(0.258)	(0.058)	(0.235)	(0.380)	(0.282)	(0.526)	(0.274)	(1.120)
Gold Standard Mean	0.200	1.174	0.500	0.763	6.398	8.318	5.148	1.499	3.350
N	174	174	174	174	174	174	174	174	174
Married woman over 30									
β_1: Traditional Module	0.000	0.076	-0.023	-0.055	-0.058	-0.091	-0.336	0.487	0.178
	(0.046)	(0.372)	(0.335)	(0.380)	(0.337)	(0.211)	(0.497)	(0.386)	(0.498)
β_2: Hybrid Module	-0.262	-0.034	0.091	-0.126	0.554	-0.159	0.277	-0.341	0.317
	(0.231)	(0.433)	(0.271)	(0.462)	(0.457)	(0.346)	(0.520)	(0.631)	(0.421)
Gold Standard Mean	0.500	2.427	0.765	1.317	4.309	8.514	5.588	0.580	0.516
N	172	172	172	172	172	172	172	172	172

Column headers denote variable outcomes, reported in hours. Row titles in italics indicate demographic group sample restriction. Sample restricted to Reference Day 1. All regressions are as specified in equation 1, including individual fixed effects. Standard errors clustered at individual level in parentheses. * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.

Table A6: Reference Day 1 Comparisons (By Demographic Group)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Wage Work	Self Employed	Working Own Field	HH Chores Outside HH	HH Chores Inside HH	Sleeping	Leisure	Active Care	Passive Care
Unmarried son									
β_1: Traditional Module	-0.041	0.150	-0.011	-0.026	0.143	-0.002	-0.325	0.111	0.004
	(0.420)	(0.363)	(0.243)	(0.308)	(0.422)	(0.229)	(0.471)	(0.199)	(0.027)
β_2: Hybrid Module	0.504	-0.050	-0.292	-0.802	-1.633***	-0.163	2.146**	0.290	0.175
	(0.436)	(0.538)	(0.261)	(0.525)	(0.552)	(0.302)	(0.914)	(0.399)	(0.158)
Gold Standard Mean	1.136	1.620	0.824	2.195	2.152	8.437	7.486	0.150	0.006
N	158	158	158	158	158	158	158	158	158
Man married to woman under 30									
β_1: Traditional Module	0.941	-0.346	-0.844	0.400	0.257	-0.069	-0.591	0.252	0.485
	(0.632)	(0.560)	(0.561)	(0.458)	(0.242)	(0.339)	(0.376)	(0.222)	(0.319)
β_2: Hybrid Module	0.526	-0.141	-0.270	0.012	-0.228	-0.036	0.288	-0.151	0.526
	(0.413)	(0.393)	(0.452)	(0.524)	(0.354)	(0.395)	(0.595)	(0.245)	(0.387)
Gold Standard Mean	2.329	2.544	2.057	1.284	0.609	8.490	6.247	0.440	0.310
N	174	174	174	174	174	174	174	174	174
Man married to woman over 30									
β_1: Traditional Module	-0.100	0.372	0.414	0.305	-0.275	-0.231	-0.550	0.066	0.244
	(0.521)	(0.457)	(0.417)	(0.341)	(0.198)	(0.324)	(0.553)	(0.138)	(0.167)
β_2: Hybrid Module	0.159	0.492	0.147	-0.095	-0.110	0.078	-0.690	0.019	0.147
	(0.281)	(0.475)	(0.647)	(0.475)	(0.172)	(0.521)	(0.549)	(0.064)	(0.160)
Gold Standard Mean	2.070	2.634	2.475	1.159	0.373	8.110	7.058	0.120	0.120
N	172	172	172	172	172	172	172	172	172

Column headers denote variable outcomes, reported in hours. Row titles in italics indicate demographic group sample restriction. Sample restricted to Reference Day 1. All regressions are as specified in equation 1, including individual fixed effects. Standard errors clustered at individual level in parentheses. * $p< 0.10$, ** $p< 0.05$, *** $p< 0.01$.