Piloting the Measurement of Time Use, Supervisory Care and Women's Agency in Indonesia

PROSPERA
Australia Indonesia Partnership for Economic Development

INVESTING IN WOMEN
SMART ECONOMICS
AN INITIATIVE OF THE AUSTRALIAN GOVERNMENT

FEB

Lembaga Demografi

Australian Government

With support from the International Labour Organization, UN Women and Statistics Indonesia (BPS)

Piloting the Measurement of Time Use, Supervisory Care and Women's Agency in Indonesia

11 April 2023

Collaborators:
Prospera
Demographic Institute, University of Indonesia
Investing in Women

With support from:
International Labour Organization
UN Women
Statistics Indonesia (BPS)

Table of Contents

Acknowledgements

The authors of this report are Amy Damon, Diahhadi Setyonaluri, Calvin Aryaputra, Usha Adelina Riyanto and Bimbika Sijapati Basnett, with substantive inputs from Samantha Watson. We would like to express our gratitude to Zarmina Nasir from Investing in Women and Ariane Utomo from the University of Melbourne for their constructive feedback throughout the study and on earlier versions of this report. We also thank Sara Duerto Valero, Cecilia Tinonin and Dwi Yuliawati-Faiz from UN Women for collaborating with us on this project. We are also grateful to Melissa Wells for her feedback, as well as Rolando Montecalvo and Ben Jackson for their valuable editorial assistance.

This report is a collaborative effort between Prospera, Investing in Women and the University of Indonesia. We would like to acknowledge and thank the International Labour Organization, UN Women and Statistics Indonesia (BPS) for their support.

We also wish to acknowledge the invaluable contributions of the team of enumerators and research supervisors whose dedication and hard work made this study possible.

We express our sincere gratitude to Bapak Ali Said, Director of Population and Employment Statistics at BPS, Ibu Mariet Tetty Nuryetty, Expert Statistician at the Directorate of Population and Employment Statistics in BPS, and their team for their invaluable support and keen interest in this collaborative study, as well as for generously sharing Indonesia's past experiences in collecting time use data. We hope that this study will benefit BPS in collecting accurate and timely time use statistics in the future. Moreover, we hope that other government agencies and development programs will be able to access time use data and leverage it to inform policies and programs to advance gender equality and women's empowerment in Indonesia and beyond.

Executive Summary

Despite significant changes to the Indonesian economy over the past three decades and increasing access to education for girls, there has been limited change in Indonesia's female labour force participation (FLFP). In August 2022, Indonesia's FLFP was 52%, compared to 83% for men. Women's participation in the labour market declines significantly following marriage and childbirth. When women do join the labour force, they are overrepresented in informal and low-quality jobs, with gender wage gaps prevailing (Cameron et al., 2018; Schaner & Das, 2016). Women in Indonesia are also underrepresented in leadership and decision-making roles, in both public (Prospera et al., 2022) and private sectors (IDX et al., 2022). The Covid-19 pandemic exacerbated gender inequality in labour participation because it increased women's care responsibilities, pushed women of reproductive age out of employment, and forced younger (15–19) and older (55+) women to take up low-quality jobs to cope with financial precarity (World Bank, forthcoming; Sijapati Basnett et al., 2022).

Globally, data on women's time use suggests an inverse relationship between their participation in paid work and the time they devote to unpaid care work. It shows that women perform 3.2 times more unpaid care work than men (Addati et al., 2018), and this inequality is associated with lower participation in the paid labour market (World Bank, 2018), less access to formal employment and decent work, and less involvement in the wider public sphere (UN Women, 2021). During the Covid-19 pandemic in Indonesia, almost 90% of women cited domestic and care work as the main reason for their discontinued employment between 2019 and 2022. The proportion of mothers who helped their children during school closures was also 2.7 times larger than that of fathers (UNICEF et al., 2022).

As the saying goes, we cannot improve what we do not measure. The recognition and valuation of unpaid care and domestic work are crucial to achieving gender equality, driving productivity and building an inclusive economy. The UN's Sustainable Development Goal (SDG) 5.4 calls for investment in public services, infrastructure, social protection policies and the promotion of shared responsibilities within the household and family. Achievement of this goal will positively influence progress towards other relevant SDGs, such as SDG 8 on decent work and SDG 10 on inequality, among others. However, to improve the situation, reliable and timely data is needed in Indonesia to better understand how individuals use their time, so that policymakers can prioritise investments and assess the country's progress on gender equality relative to others.

Time use surveys collect data on the amount of time individuals spend on certain tasks over a specified period, typically 24 hours. They reveal details of a person's daily life that are not otherwise captured in standard household surveys, including the amount of unpaid domestic and care work people do (which may entail caring for children and others) and total work time (time spent in all paid and unpaid productive activities). The results can reveal inequities in the distribution of work, rendering visible the extent of unpaid work and the time constraints that act as barriers to engaging in activities that may expand economic

opportunities (e.g. going to school, studying, seeking employment, etc.). There are several informative examples of countries and development efforts leveraging such data to inform their programs and policies (Buvinic & King, 2018).

Previous attempts to collect time use data in Indonesia have been challenging, but the development of improved and innovative instruments could aid future efforts to gather such data. Independent time use surveys, usually in a diary format where a person fills out a detailed record of their activities, have been the preferred survey approach to generate data on unpaid work. But a full time use survey can be time-consuming, costly and may not necessarily fully capture the demands of unpaid care work, with the time spent 'being on call for care' particularly prone to underestimation. Statistics Indonesia (BPS) trialled time use surveys three times between 1998 and 2005,[1] encountering challenges familiar to those seen in other emerging economies. Varying levels of education and literacy, informal and seasonal work, complex family arrangements and traditional gender roles complicate data collection on time use.

However, recent studies and experiences have shown that careful survey design, technological advancement and iterative dialogue between data producers and users can help overcome these challenges and generate internationally comparable and reliable time use data. On their own, time use surveys do not reveal the extent to which women and men are able to make choices on how to use their time and act on their decisions, without being constrained by socio-cultural and economic barriers (agency). This necessities use of other methods to complement time use surveys.

In November 2022, Prospera, Investing in Women, Lembaga Demografi (LD) Universitas Indonesia, the International Labour Organization (ILO) and UN Women kickstarted a time use and women's agency pilot study with two work packages. The partners trialled a new pilot modular light diary tool developed by ILO to assess how men and women allocate their time. The study also tries to shed light on women's ability to make choices about time use, and what factors influence them, by testing a novel 'lab-in-the-field' experiment and fielding attitudinal questions.

Work Package 1 piloted a new modular approach to time use survey developed by ILO to collect data on how individuals in two-parent households spend their time across different activities. This study piloted a 'light diary' tool that included the fullest coverage of simultaneous activities to measure all dimensions of unpaid care, including supervisory care. Unpaid supervisory care refers to the time a person is 'available and in close proximity to provide active care for a child or a dependent adult should the need arise' (UNSD, 2022). This may occur at any location where the care recipient is present and in close proximity to the care provider. The survey approach is termed 'light' because it records 41 types of pre-coded activities, rather than open responses (recorded verbatim in the respondents own

[1] The first pilot was done in 1998, when the time use module was embedded in the Survey of 100 Villages (Survei Seratus Desa). The second pilot was done in 2004 in Greater Jakarta, and the third survey in 2005 in four provinces: West Sumatra, Bali, Central Java and North Sumatra.

words), for one 24-hour day and included a novel sequence of recovery questions to target underreporting of supervisory care. The instrument is designed to be attached to a parent survey to take advantage of existing infrastructure of a larger national survey. It is administered face-to-face through an electronic table using CSPro software, which pre-codes activities and uses fixed time episodes to ensure faster data collection while minimizing errors through features like data entry format restrictions and automated updates.

This pilot aims to contribute to the development of international guidelines for collecting standardised time use statistics to be able to quantify women's unpaid work, and to test and refine a set of instruments that can eventually be included in Indonesia's National Labour Force Survey (Sakernas), which is conducted each six months. This initiative will assist BPS in the development of methods, tools and guidelines to produce time use data more efficiently and report against the SDG indicator 5.4.1, 'time spent on unpaid care work by sex, age and location' and related SDGs (e.g. SDG 8 on decent work).

Work Package 2 trialled a novel methodology to gauge women's agency over time use, or the ability to define time use goals and pursue them. Our study employed an attitudinal questionnaire to elicit respondents' agency over time use across four main dimensions:

- Critical consciousness: awareness that men and women may have unequal time allocations, preferences and time poverty. As an example, how strong do you agree/disagree with the statement: women can work even as the main earner?
- Self-efficacy: confidence in one's ability to make decisions about how to spend one's time). As an example: how strongly do you agree/disagree that you can change you daily schedule?
- Voice: the ability to talk about time use). As an example, how strongly do you agree/disagree that you can ask a household member to take care of child or another family member?
- Decision-making: The extent to which one can decide the amount of time to spend on activities. As an example, how decides how much time you can spend on paid work?

Couples then participated in a lab-in-the-field experiment to test how much time women and men would allocate to attend a hypothetical training program to improve their income earning potential, and to understand how people's time allocations may change or adjust when information is private, public or negotiated. Lab-in-the field experiment combines standardized lab methodology with a field experiment conducted in a natural setting.

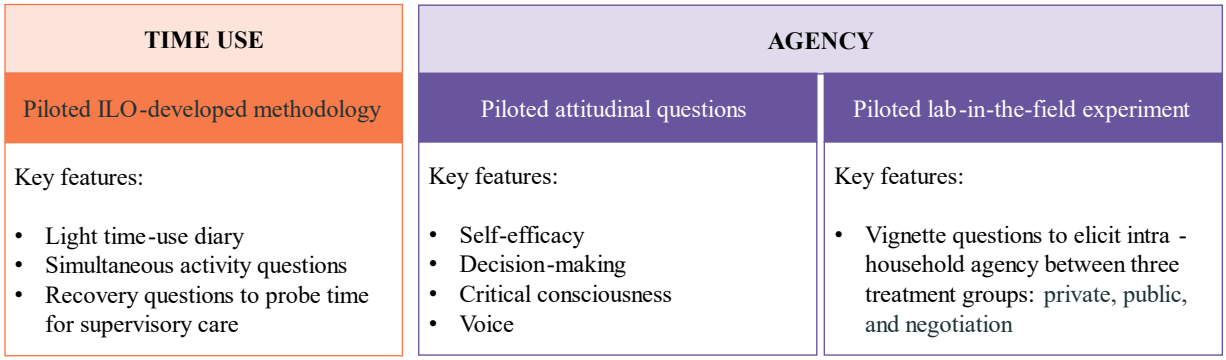
To this end, each couple was randomly assigned to one of three treatment groups:

- Private: each spouse decides the time allocated to training on their own,
- Public: each spouse reveals the time they allocated to training to the other spouse
- Negotiated: spouses negotiate with one another about the allocation of time for training).

Each participant was then asked to respond to four vignettes (or stories) posing the following questions:(1) how many hours they would attend if their spouse was at work all day, (2) how many hours they would attend if their spouse was at home, (3) how many hours a neighbour

should attend, and (4) how many hours their spouse should attend. The diagram below provides a visual summary of the components of the two work packages.

While the attitudinal survey questions are adapted from Sinharoy et al.'s (2021)work, the lab-in-the field experiment used for this study is novel and has not been previously attempted.

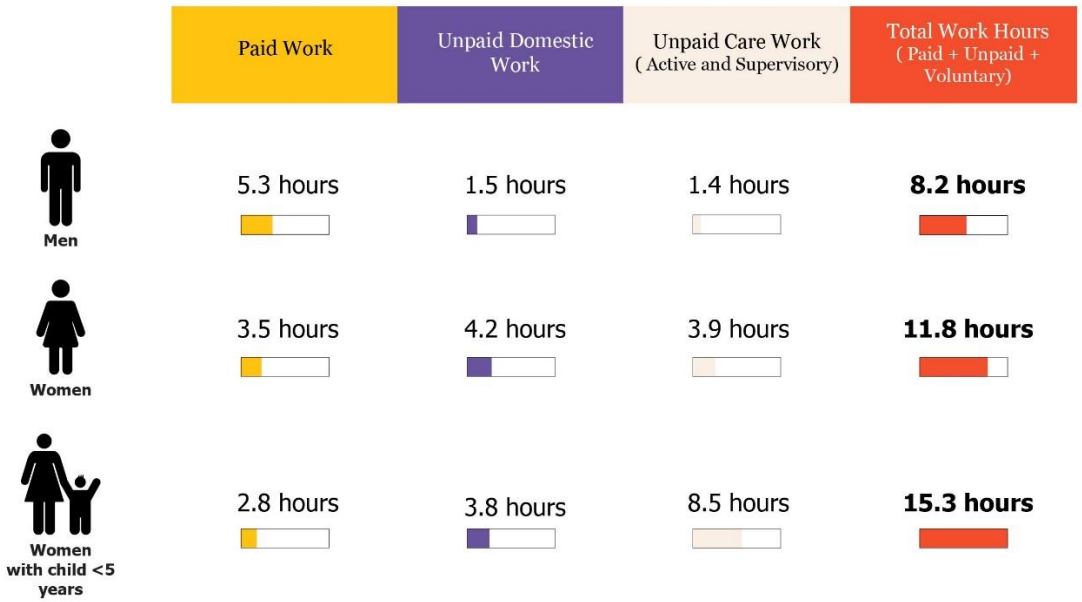


After proof of concept and cognitive testing of the instruments (i.e. assessment of respondents' understanding of the questions and the terms used), 226 randomly selected couples, or 452 individuals, in Greater Jakarta and Greater Surabaya participated in the study between October and November 2022. In each research site, three cities and two sub-districts within each city were selected randomly. The days of the week when each two-parent household would be interviewed were also selected randomly. Information on how the couples spent their day was collected directly from each spouse. This ensured that the twin requirements for pre-assignment of reporting day and no proxy reporting were met, thus minimising the scope for systematic differences in time use by day of the week and proxy response error to bias estimates.

Three separate types of empirical data – time use data, attitudes to agency, and experimental data – were collected in the two work packages, and a combination of descriptive analysis and multivariate regression was used to analyze them. Descriptive analysis (mean, standard deviation and other relevant comparisons, including data visualisations) was used to examine distribution of time allocation for various activities across demographic characteristics and care responsibilities. Responses to attitudinal questions (on a scale of 1 to 5, strongly disagree to strongly agree) were grouped according to four different dimensions of agency; responses to each individual category were analysed and aggregated scores were calculated. A higher score indicates a higher degree of agency. For the experimental outcomes, the average responses to each vignette by treatment group were compared; whether people responded differently to the vignettes depending on their treatment assignment was also examined. The differences between each spouse's preferred hours of attendance at the training were compared with the other spouse's preference for them. Finally, multivariate analysis was used to investigate how time use, attitudes towards agency and experimental data are related, and how different levels of agency affect time use outcomes.

Although this pilot study drew on a small urban sample, some of the findings show broader patterns that could be tested in a larger study. The findings are summarised below in tabular form.

Time use	
Key finding 1: On average, men spend 1.5 times more time than women on paid employment. Meanwhile, women spend almost three times more time than men on unpaid domestic work.	1. Both men and women spend most of their time on self-care and maintenance activities (including sleeping), with no discernible gender differences. 2. Men spend more time on paid employment activities ($\bar{x}$ = 5.3 hours) than women, while women spend 2.8 times as much time as men on unpaid domestic work ($\bar{x}$ = 4.2 hours) and engage in more unpaid active caregiving services ($\bar{x}$ = 1.7 hours).



Average number of hours spent on selected activities, for men, women, and women with children aged up to 5 years old (please note, volunteering hours are not shown separately but are included in total).

Time use

Key finding 2: Women, particularly mothers with young children, spend more time on supervisory care and work longer hours overall than men. Increased hours in unpaid domestic, care, and supervisory tasks all negatively all reduce the probability of participation in paid work, with supervisory care having the negative effect.	1. The number of hours allocated to supervisory care is significantly higher among women, who spend twice as much time ($\bar{x}$ = 2.2 hours) as men on it. The difference is even greater for mothers with children under 5 years of age ($\bar{x}$ = 5.2 hours). 2. When considering the total amount of time spent on all types of work (paid employment, unpaid domestic work, unpaid care and volunteer/trainee work), women work longer hours than men ($\bar{x}$ = 11.8 hours for women, $\bar{x}$ = 8.2 hours for men). 3. A larger proportion of women with children provide supervisory care while doing tasks that require less attention, such as resting or eating, compared to women caring for adults. Meanwhile, women caring for adults in need provide supervisory care while doing more intensive tasks, such as cooking and cleaning. This finding highlights how demanding supervisory care of young children can be. It suggests that women with young children may have greater restrictions when combining caregiving with other activities. 4. There is a negative relationship between amount of unpaid work and the probability of both women and men. An increase in 1 hour of unpaid work decreases the probability of employment by 3% for men and 4% for women, with supervisory care having the largest negative effect on women. For female respondents, each additional hour of supervisory care they provide is associated with a 7% decrease in their probability of engaging in paid work.

Agency – Attitudinal questions	
Key finding 4 (self-efficacy): Women and men feel a high degree of confidence in their ability to decide how to use their time. However, fewer women can reach out for help from others for domestic and caregiving responsibilities.	1. The means of the aggregate self-efficacy scores of men and women are similar, but there are clusters of men with very high and very low self-efficacy scores. 2. A larger share of women than men feel they can change their daily schedule, but more men are able to ask for help with domestic and childcare duties. 3. These results likely reflect the fact that women have either internalised the view that domestic and caregiving activities are their responsibility or have less access to alternative care arrangements.
Key finding 5 (decision-making): Compared to men, women have more agency in deciding the amount of time they spend on domestic and caregiving activities, but less agency in deciding how much time to allocate to work for pay.	1. Overall, males report lower decision-making agency than women, especially with regard to domestic and caregiving activities. 2. Compared to men, women seem to have more agency in deciding the amount of time spent on domestic work/caregiving or leisure, but less agency in deciding how much to work for pay. This is particularly prevalent among less educated women who tend to adhere to gendered work allocations in the household. 3. Men's low level of decision-making agency may be attributed to their engagement in paid employment, the schedule of which is externally established.
Key finding 6 (critical consciousness): The norm of men being the breadwinners of the family is entrenched. Women in general and more educated men are more open to the idea of women working as breadwinners.	1. There is broad agreement that men should provide for the family and women are responsible for domestic work. However, a larger share of women than men disagree or strongly disagree with this view, and a larger share of women also think that women can work as the main earners. 2. Favourable male attitudes toward women working increase as men's education increases. Men with a higher level of education also report a higher critical consciousness about women working more and sleeping less.

Agency – Attitudinal questions	
Key finding 7 (voice): Men and women can change their schedules with ease, and this is more apparent among those who work for pay.	1. The aggregated scores for men and women have similar means – both find it easy to change their schedules without consulting their spouse. 2. However, a larger share of men than women feel they can access support for household duties and childcare.
Key finding 8: There is a correlation between time devoted to paid and unpaid work and their level of agency. Women with higher agency in terms of decision-making and voice tend to participate more in paid work.	1. High self-efficacy among women is associated with higher unpaid domestic work, but not caregiving work. 2. Women with higher decision-making and voice agency are associated with allocating more hours to paid work. 3. Women with strong opinions about who should work are able to ask for help and spend less time providing supervisory care or caregiving services for their family.

Agency – Experimental	
Key finding 8: In general, negotiation between respondents and their spouses increases both men and women's preferred hours of training.	1. Women reported the lowest hours of attendance (2.7 hours out of a possible 8 hours) in the private treatment group, that is, when their husbands were unaware of their answers. When they negotiated with their spouse or disclosed their preference to their spouse, women reported a higher preferred number of training hours (3.1 or 3 hours, respectively). 2. The reported number of hours that women would like their husbands to attend was highest in the private treatment group (5.2 hours) but lowest when they negotiated with their husbands (4.8 hours). 3. Similarly, men reported the highest preferred number of hours of training when they negotiated with their wives (3.7 hours), but they wished for their spouse to attend for longer (4.1 hours). 4. In the negotiation treatment group, men were 'bargained up' for their own attendance and their wives'. This result is robust across all types of households and indicates that spouses can negotiate around the norms that each spouse may assume in the private or public treatment groups.
Key finding 9: Having a spouse to take care of children and higher education are associated with selecting more training hours, indicating a lifting of constraints to participation in paid work.	1. Care responsibilities play a role in training attendance for both women and men. When the scenario indicates their spouse can take care of children, both men and women express a greater desire to attend training for more hours. 2. Education is positively linked to the number of training hours chosen by both women and men. Those with higher education tend to select more training hours than those with lower levels.

Key finding 10: There is misalignment in assumptions about training preferences between spouses, with men often underestimating their spouses' s preferred hours.	1. Most men (52%) reported fewer hours of preferred training attendance for their spouse than their spouse's own preference. Though only 27% of women report wanting to attend more hours than they would like their spouse to attend. 2. Both men and women tended to overestimate the number of preferred training hours for their spouse. For women, the gap between how many hours they would like their spouse to attend and the number of training hours they would like to attend themselves is driven by the amount of unpaid work they do. The more unpaid work they have, the bigger the gap between their preferred training hours and their spouse's (i.e. increasing by 0.167 hours), indicating that women with more unpaid work are expressing a greater desire to attend training, though the evidence base is still preliminary. For men, being in the negotiation treatment may contribute to this gap.

Piloting the two work packages suggest that women participate in the formal labour market less than men due to higher allocation of unpaid work, societal attitudes and less agency over decision to pay. The time use survey findings validate gender disparity in the way paid and unpaid work are distributed between women and men in the household. When we account for their work, it becomes clear that women, especially those with young children, work longer hours than men, leaving them with limited time to work for pay. The agency findings reveal that attitudinal barrier compound women's challenges. There is a broad agreement that men should provide for family and women are responsible for domestic work. Women also exhibit less agency when it comes to deciding to work for pay. Although more women are aware and readier to engage in paid work, male attitudes do not match women's willingness. However, the field experiment reveals that when spouses can negotiate with each other, the preferred hours of training to earn income increases for both women and men. And if care is not a constraint (spouse is around to look after children), the desired training hours go up too. This suggests, improving communication with spouses and relaxing social norms and childcare-related constraints could potentially increase women's participation in the economy.

The study findings suggest investing in education, campaigns, care ecosystem, and data could be effective measures for addressing gender disparity in unpaid work and enhancing economic opportunities available for women. The results show men and women who have more years of education have more equal views of paid and unpaid work and gender roles within a household. Therefore, education is a long-term investment that can shift gender norms. Encouraging sharing unpaid domestic work regardless of gender is an important step to reducing women's workload. Campaigns targeting attitudes about women

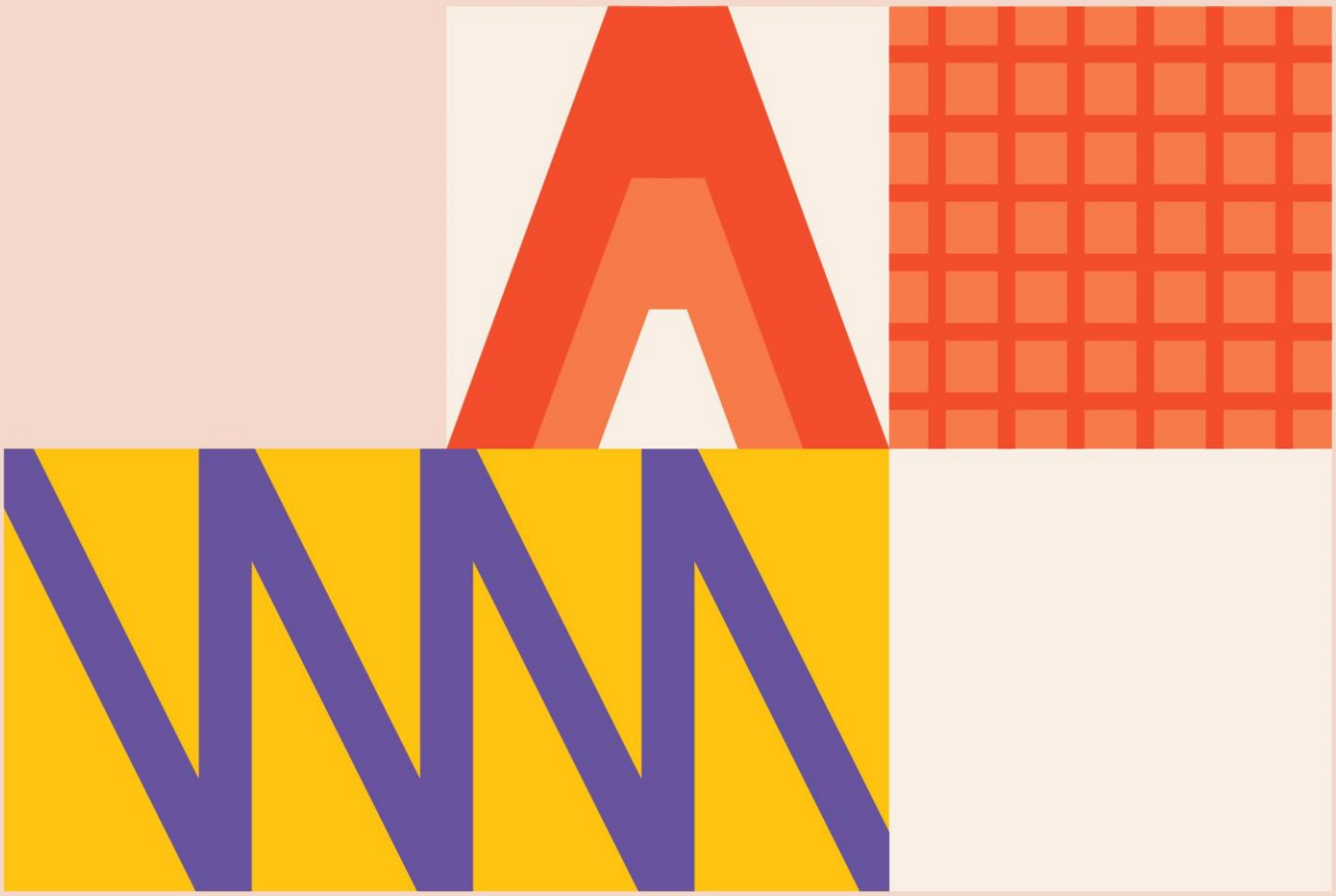
entering the paid labour force is important. Women, especially those engaged in high levels of unpaid work, have demonstrated an unmet demand for training that could lead to paid employment. However, many women will need support to overcome societal expectation to provide unpaid work first and exercise agency to pursue paid work. Moreover, without care infrastructure (such as childcare facilities, elderly care facilities and paid leave), caregivers, primarily women, are unable to fully participate in the labour market. Investing in high-quality and affordable care ecosystem can have multiple benefits for society, including redistributing women's workload, promoting children's development, supporting families, promoting job creation and boosting economy. Collective investments in the care ecosystem can address Indonesia's childcare gaps by unlocking the potential of the private sector, supporting the government in improving the quality and standardisation of childcare providers, and adjusting labour laws and workplace regulations. Finally, good quality time use data underpins effective policy and programs to lift female labour force participation, and boost income earning capacity, such as entrepreneurship and skills training. Therefore, it is imperative to collect comprehensive and accurate data on time use, including unpaid work, to inform evidence-based policies and programs that promote gender equality and women's economic empowerment.

The implementation of the time use instrument has provided valuable insights that can guide future time use data collection efforts in Indonesia. The time use survey has been piloted with the intention of assisting Statistics Indonesia in integrating a time use modular into Indonesia's Labour Force Survey (Sakernas), undertaken each six months. First, there is a need to account for both paid and unpaid work, as using only a narrow definition of work results in mismeasurement of total domestic work time and prolongs the undervaluation of women's contribution to their households, communities and the overall economy. Second, measuring the density and intensity of activities that people engage in is important and can be achieved by considering tasks that are done simultaneously. Third, distinguishing between supervisory and active care allows us to better estimation of care work time. From a methodological standpoint, the study suggests that the recovery section of the pilot time use module – a section which returns to supervisory care specifically after the diary day activities have been fully reported as a consecutive series of activities – may improve the measurement of supervisory care hours. The study has also offered valuable lessons on methodological issues such as selection of representative households, training enumerators and allowing sufficient time for multi-modal approach. However, further testing of the instruments in rural areas is necessary to ensure the readiness of the instruments for national level roll out.

Structure of the report. The report is divided into five sections. Section 1 discusses the background and context; Section 2 describes the instruments, sampling and analytical strategy used; Section 3 offers a comprehensive overview of the research results and findings from the pilot implementation; and Section 4 outlines the conclusions and programmatic recommendations based on the findings.

Section 1

Introduction, Background and Programmatic Focus



1. Introduction

1.1. Background

Despite considerable changes to the Indonesian economy, progress in reducing the gender gap in education and labour stalled in the two decades prior to the pandemic, with female labour force participation (FLFP) now at just 52%, compared to 83% for men. Women's participation in the labour market declines significantly following marriage and childbirth. Even when women do participate, gender wage gaps remain high, and women are overrepresented in informal and low-quality jobs (Cameron et al., 2018; Schaner & Das, 2016). They are also underrepresented in leadership and decision-making roles, in both public and private sectors (Prospera et al., 2022; IDX et al., 2022).

The onset of the Covid-19 pandemic exposed and exacerbated these gaps. On the surface, women's economic participation appears to have been boosted during the pandemic. From August 2019 to August 2020, FLFP increased by 1.2 percentage points, which is a notable rise and a deviation from the experiences of many other countries. However, a closer investigation reveals that this increase was made up of younger women who would otherwise have been in school and older women entering the labour force as added workers, to compensate for negative shocks to family income. Meanwhile, the labour participation of women of childbearing age declined, likely due to the increased care responsibilities they had to shoulder (World Bank, forthcoming; Sijapati Basnett et al., 2022).

Findings focused on the second year of the pandemic also point to the lingering setbacks that women have experienced relative to their male counterparts. By 2022, women who reported having worked in 2019 faced unemployment at a rate four times that of men (UNICEF et al., 2022).

Since the onset of the pandemic, Investing in Women and Prospera have collaborated to better understand its impact on women's economic participation in Indonesia, and to leverage research to inform our respective work programs.

First collaboration round between Prospera, Investing in Women and Lembaga Demografi. The first round of the study, undertaken by Lembaga Demografi Universitas Indonesia, focused on young adults in urban areas of Jakarta and Surabaya (see Setyonaluri et al., 2021). It sought to better understand the role of social norms in influencing FLFP, and whether the pandemic contributed to shifts in gender roles at home and work. The study employed a mixed-method approach, combining analyses of Indonesia's national socioeconomic and labour force surveys. The objective was to explore what shifts occurred during the pandemic, and to probe deeper into the inner workings of the gender norms influencing these trends. It used an intersectional approach, examining impacts across different groups of women.

A key finding from the quantitative analysis of Indonesia's labour force survey between 2009 and 2019 (the decade prior to the pandemic) was that family responsibilities limited women's economic participation, particularly among women of childbearing age. While there had been a modest rise in FLFP between 2009 and 2019 (from 55% to 56%), it continued to lag behind men's. The presence of young children negatively affected FLFP, and women who quit the labour force often cited pregnancy and domestic work as the main reasons.

Most respondents who participated in focus group discussions firmly believed in *kodrat*, or the view that women are naturally better carers of the family and children, while men are inherently better at being breadwinners. Women's work outside the house was seen as secondary and had to be flexible enough to allow them to perform their primary role. Men who stayed at home and looked after their homes were a source of 'embarrassment and burden', while women who pursued their careers were considered neglectful. The responses appeared to suggest that such norms did not necessarily prevent women from taking up paid work but rather limited their work options.

Three notable changes emerged during the pandemic – a slight increase in FLFP (from 51.9% to 53.1%), a slight decline in men's labour force participation (from 83.1% to 82.4%), and a significant rise in the percentage of men who reported doing domestic work (from 55% to 68.4%). Upon closer investigation, however, the economic slowdown caused by the pandemic apparently discouraged men from looking for work because of the limited opportunities in the labour market, while women took up low-quality jobs to compensate for the reduced income earned by men during the crisis. Because Indonesia does not yet collect time use data, it was unclear how much domestic work men do or did, the nature of the work, and whether this led to any significant shift in the division of domestic and paid work.

Most participants in focus group discussions regarded working women positively during the pandemic because they were perceived as keeping their families afloat. However, women were still expected to perform the majority of household tasks. Families who experienced a significant shift in roles – women went outside for work while men stayed at home – were subject to gossip and reproach. Thus, norms of *kodrat* remained intact even as families were readjusting their engagement on work and home fronts during the economic crisis.

Both the quantitative analysis and the follow-up focus group discussions revealed that the effect of the pandemic on women was not uniform. Educated women experienced fewer changes because they could outsource care work, resign from their jobs, start an online business, update their skills and/or pursue new opportunities in the event of furlough or reduction in income. Their privileged socioeconomic situation appeared to expand the range of choices available to them and their ability to negotiate for improved outcomes. In comparison, less educated women and single parents faced the greatest employment penalties and experienced more pressure to supplement household income, despite incurring a higher care burden during mass school closures.

This first phase of the study highlighted the need to pay attention to the gender norms that underpin women and men's preferences and labour market behaviour, but it also pointed to

many issues to explore further. Having to negotiate between work and home responsibilities is costly for women, and more information is needed about how this negotiation is affected by socioeconomic status. Negotiation implies choice – women are not merely victims of restrictive gender norms and, in fact, demonstrate micro agency, or the ability to define one's own goals and act on them (Kabeer, 1999) to some extent. Preferences and choices are, however, culturally, and socially embedded. There is a need to unpack the difference between what people view as acceptable behaviour (norms) and what they actually do (real behaviour, such as allocating time); explore the motivations behind women's decisions on how to participate in work and home spheres; whether women's preferences and labour market behaviour are consequences of their own decisions, or lack of awareness of their rights and/or inequities ('critical consciousness'), or merely symptoms of pressures and constraints imposed on them.

These issues are important to investigate and understand in order to design, fine-tune and evaluate policies and programs that can augment women's economic empowerment (for instance, by expanding women's understanding of how the imposition of norms affects their agency and urging them to negotiate for improved outcomes by voicing their preferences and influencing decision-making) without reinforcing inequalities (by increasing inequalities in the division of work, reducing women's leisure and overall wellbeing, increasing their time poverty, etc.). Such programs and initiatives include campaigns to shift gendered norms around women as primary carers and men as primary wage earners; private and public sector policies to encourage women to enter the labour force through improved information and/or skills boosting; investment in quality and affordable childcare; and/or other family-friendly policies.

Why do gender gaps in economic participation persist, and what can be done about them?
Available time use data from around the world suggests that women perform most unpaid care and domestic work (Addati et al., 2018), and this inequality is associated with reduced participation in the paid labour market (World Bank, 2018), less access to formal employment and decent work, and lower involvement in the wider public sphere (UN Women, 2021). The inverse relationship between time allocated for paid and unpaid work is likely to hold in Indonesia as well. Many women who stopped working between 2019 and 2022 in Indonesia cited care responsibilities as the main reason, while men were more likely to report pandemic-related reasons (e.g. furlough and shutdowns). Compared to fathers, 2.7 times more mothers said they supported their children to learn online, pointing to the additional unpaid care work mothers had to shoulder during mass school closures (UNICEF et al., 2022).

How society, employers and policymakers address unpaid domestic and care work has important implications for achieving gender equality, driving productivity and building an inclusive economy (Ferrant et al., 2014). Indeed, the UN's Sustainable Development Goal (SDG) 5.4 elevates the recognition and valuation of unpaid care and domestic work as an integral dimension of gender equality, and calls for investment in public services,

infrastructure and social protection policies, and the promotion of shared responsibility within households and families.

As the saying goes, however, we cannot improve what we do not measure. Reliable and timely data is needed in Indonesia to better understand how individuals use their time and what drives these choices; prioritise investments that can help reduce gaps; learn what works and what does not; and assess how Indonesia fares on gender equality relative to other countries.

Time use surveys help reveal details of a person's daily life that are not otherwise captured in standard household surveys. They collect data on the amount of time individuals spend over a specific period (typically 24 hours), capturing both paid and unpaid work, including caring for children and others. The results can reveal inequities in the distribution of work and bring to the fore unpaid work and the time constraints that act as barriers to engaging in activities that could expand economic opportunities (e.g. going to school, seeking employment, etc.). There are many examples of countries and development efforts leveraging time use data to inform programs and policies (Buvinic & King, 2018).

Statistics Indonesia (BPS) has trialled time use surveys three times in the past, the last time in 2005.[2] These efforts were fraught with challenges that are not unique to Indonesia. Time use surveys are more time-consuming and complicated to administer compared to standard national socioeconomic surveys. Past surveys tended to underestimate the time women with care responsibilities spend on being 'on call', thereby reducing available time or quality of engagement for other activities, including paid work and leisure (Folbre, 2018). These challenges are even greater in emerging countries, including Indonesia, where there are variable levels of education and literacy, many workers in informal and seasonal work, families with complex and fluid arrangements, and traditional gender roles that may bias responses. Moreover, these surveys are not yet designed to reveal the degree to which individuals can exercise agency (the ability to make choices) over allocation of time, and what factors contribute to it.

Fortunately, recent studies have found that these challenges can be attenuated through careful design. There are global efforts underway to refine instruments to collect information on unpaid care and domestic work and sharpen the guidelines provided to national statistics offices to produce internationally comparable data. Technological advancement has improved data collection and sped up analyses. For instance, the use of computer-aided personal interviewing (CAPI) allows survey practitioners to collect time use data more efficiently and with fewer errors (ADB, 2019). There are also emerging efforts to attach other data collection methods to standard time use surveys in order to better investigate the

[2] The first pilot was conducted in 1998, when the time use module was embedded in the Survey of 100 Villages (*Survei Seratus Desa*). The second pilot was done in 2004 in Greater Jakarta, and the third survey in 2005 in four provinces: West Sumatra, Bali, Central Java and North Sumatra.

relationship between time use and agency, such as the one seen in Sinharoy et al. (2021), who trialled a time-use agency scale instrument in Ghana.

What is key is an iterative dialogue between data producers (national statistics offices, study teams) and data users (government agencies, donor programs) to drive the inquiry, share the approach and findings, and time the data collection to align with policy and programmatic cycles (Buvinic & King, 2018).

How are we contributing? Second round of collaboration with additional partners. In November 2022, Prospera and Investing in Women, in collaboration with Lembaga Demografi, the International Labour Organization (ILO) and UN Women, kickstarted a time use and agency pilot study. The aim of the study is twofold. Its first objective is to assist BPS in developing methods, tools and guidelines to produce time use data more efficiently, and report how Indonesia is faring in the SDG 5.4.1 indicator (unpaid domestic and care work by sex, age and location). With this pilot, Indonesia is also participating in global efforts to refine the guidance provided to other emerging economies on the production of time use statistics.

The second objective is to elucidate women's agency over time use. We investigate whether women and men are aware of their time use, time use inequalities and personal aspirations, and assess their level of confidence in their ability to (re)allocate their time. Our data collection and analysis seek to understand whether people's time allocation is an expression of their voice in actual decisions, and whether the time allocated for activities (paid or unpaid) is in line with their needs and personal aspirations (Sinharoy et al., 2021). Finally, by offering respondents the opportunity to boost their personal or family income through a hypothetical training event, our effort seeks to reveal their time use preferences and measure how different levels of communication between spouses might affect them.

People's agency, and specifically their time use agency, is not easily observed using standard data collection methods (Donald et al., 2020). Reasonable measures, or even proxies for agency, do not exist in administrative or previous survey data. Our project is unique in that it combines detailed time use data with a focus on unpaid labour and supervisory care (see Box 1), and two measures of agency.

The partners trialled a new pilot modular light diary tool developed by ILO's Department of Statistics to assess how men and women allocate time. Prospera and Investing in Women also experimented with a novel methodology to measure time use agency that included a set of attitudinal questions and a lab-in-the-field experiment. Spouses from 226 households (a total of 452 respondents) living in Greater Jakarta and Greater Surabaya participated in both these data-collection initiatives. Furthermore, the partners refined the instruments and methodology throughout three distinct phases: proof-of-concept testing, cognitive testing, and piloting.

The overall effort is motivated by our mutual commitment to contribute to the production of reliable information on time use and agency to inform our respective programs and policy advice, while assisting Indonesia to measure its progress on gender equality. For Prospera

and Investing in Women, this endeavour is also underpinned by our interest in continuing our ongoing collaboration and deepening our knowledge of women's economic participation in Indonesia, including how gender norms impact women's work.

1.2. Project Objectives

Our global objective is to build an evidence base for policies and initiatives aimed at freeing up women's time to engage in meaningful paid work (e.g. through investments in childcare, campaigns to redistribute unpaid care work in the home, etc.). Our study is also an opportunity to continue our joint investment in foundational research by addressing follow-up questions raised during the first phase of our collaboration on women's agency, and to generate reliable data and statistics to help Indonesia meet its global commitments.

Our current collaboration, and thus the work described in this report, builds on its successful predecessor project with two interrelated objectives:

- First, to implement a time use survey that enables us to better measure women and men's time use allocation. Specifically, in addition to measuring forms of time use typically included in such surveys, we pay particular attention to the measurement of an activity that is often underreported, namely, supervisory care.
- Second, we aim to develop an instrument that can be used to better understand the factors influencing women's agency over time use.

The output of our efforts, comprehensively detailed in this report, is the development of (a) recommendations to support and inform the rollout of the Indonesian Government's national survey on time use in 2023 and beyond, and (b) a time use agency module that can be used to design, deliver and evaluate policy advice and programs.

Our project entails two work packages. The first work package piloted a time use survey, contributing to the development of international guidelines for collecting standardised time use statistics on women's unpaid work, in collaboration with ILO, UN Women's 'Women Count' program and BPS. In parallel, the second work package aims to understand whether and to what extent women and men exercise choice over their allocation of time, and how they do so. We used a lab-in-the-field experiment and an attitudinal survey about time use agency to complement the time use survey in Work Package 1. Time use agency is a multidimensional construct that aims to 'capture a person's critical consciousness of time use inequities, rights, and personal aspirations; confidence in their ability to (re)allocate their time; expression of voice about allocation of their time; and actual influence over decisions about allocation of their time across the full range of needs and choice-based activities, in line with their personal aspirations' (Sinharoy et al., 2021, p. 8).

Box 1. Definitions of work used in this report

Paid and unpaid work

This report defines work following the International Conference of Labour Statisticians' resolution concerning statistics on work, employment and labour underutilisation. Work is recognised in five forms:

1. own-use production work, or goods and services for own final use
2. employment, or production of goods/provision of services for pay or profit
3. unpaid trainee work, or any unpaid production of goods/provision of services to acquire workplace experience or skills in a trade or profession
4. volunteer work, or any unpaid, non-compulsory activity to produce goods or provide services for others
5. other forms of work, such as unpaid community or military service.

These five types of work are part of a broader classification of nine major categories of activity, as defined by ICATUS 2016:

Major division	Classification	Activity	Paid or unpaid
1	Productive activities (work)	Employment and related activities	Paid
2		Production of goods for own final use	Unpaid
3		Unpaid domestic services for household and family members	
4		Unpaid caregiving services for household and family members, including supervisory care	
5		Unpaid volunteer, trainee, and other unpaid work	
6	Personal activities	Learning	
7		Socialising and communication, community participation and religious practice	
8		Culture, leisure, mass media, and sports practices	
9		Self-care and maintenance	

Supervisory care

This report adopts the provisional definition of supervisory care proposed by the Sub-Committee on Supervisory Care convened by the UN Statistics Division's Expert Group on Innovative and Effective Ways to Collect Time-Use Statistics:

> *Unpaid supervisory care refers to the time a person is 'available and in close proximity' to provide active care for a child or a dependent adult should the need arise. Supervisory care may occur at any location when children or care recipients are also present and in close proximity to the care provider. That is, the respondent is near enough to the care recipient to provide immediate assistance, if necessary. There is no requirement for the care provider and care recipient to be in the same room nor for the care provider to be aware of what the care recipient is doing. (Annex I, Minimum Harmonized Instrument Background Paper for the 53rd Session of the UN Statistical Commission.)*

Supervisory care is a relevant dimension of caregiving for children below a specified age, and for older children and adults who require assistance with daily activities owing to illness, disability or age-related frailty.

'Supervisory' refers to time spent being 'on call', that is, ready to respond should the need arise. In all cases, the caregiver must remain available and in sufficiently close proximity to be able to assist the care recipient. Supervisory care excludes time when the caregiver is directly interacting with the person requiring care.

Depending on the analytical question to be addressed, supervisory care can include, for example, time when the care recipient is sleeping and time when the caregiver is engaged in parallel activities, including paid activities, provided that they remain available and in sufficient proximity to deliver immediate assistance if needed.

1.3. Policy and Programmatic Motivation

1.3.1. Work Package 1: Measuring Time Use

Work Package 1 pilots a time use survey to collect data on how many minutes and hours individuals in two-parent households spend on a wide array of activities. It pays particular attention to comprehensively identifying and measuring dimensions of unpaid domestic and care work within the household (see Box 1), including *supervisory care* – an often overlooked but important use of time, particularly by women.

According to the preliminary working definition by the Expert Group of the UN Statistics Division (UNSD), unpaid supervisory care refers to the time a person is available to provide active care for a child or a dependent adult should the need arise. Supervisory care may occur at any location when children or other care recipients are present and in close proximity to the care provider. That is, the respondent is near enough to the care recipient to provide immediate assistance if necessary.

This pilot aims to contribute to the development of international guidelines for collecting standardised time use statistics on women's unpaid work; test/refine a set of instruments that BPS can eventually include in its National Labour Force Survey (Sakernas) in 2023 or beyond so Indonesia can successfully report against SDG 5.4.1 (proportion of time spent on unpaid domestic and care work, by sex, age and location) and other relevant SDGs (such as SDG 8 on decent work); and support the development and monitoring of national policy objectives related to gender equality in the world of work and society more broadly.

In recent years, the measurement of unpaid domestic and care work has received increased attention in official statistics. There is growing interest in unpaid domestic and care work in national and international policy circles, largely resulting from an acknowledgement of its economic contribution. Similarly, there is growing concern that, globally, women and girls contribute over three-quarters of the total hours spent daily on unpaid domestic and care work, to the detriment of their labour force participation and retention, access to formal employment and decent work, and involvement in the wider public sphere (Ferrant et al., 2014).

Historically, independent time use surveys (TUSs), usually in diary format, have been the preferred survey approach to generate data on unpaid domestic and care work. While a small number of high-income countries have successfully integrated time use measurement into their national statistics system, the complexity and costs of conducting dedicated TUSs have deterred their widespread uptake in national statistics systems. In settings where dedicated TUSs are not feasible, interest has turned to alternative approaches to meet the growing demand for data on unpaid domestic and care work.

This study utilises a time use module developed by ILO with support from Data2X (a civil society organisation working to improve the production and use of gender data) and the UN Foundation, as part of a wider program to support the implementation of international standards on statistics relating to work, employment and labour underutilisation within countries' national labour force surveys (LFSs). As part of a pilot project, ILO developed and piloted a series of hybrid light diary and stylised modules, with testing undertaken in several countries (with piloting to refine the modules ongoing).

Our project used a version of the pilot light diary tool with the fullest coverage of simultaneous activities, given our focus on unpaid domestic and care activities and supervisory care (see Box 2), which are often performed simultaneously with other activities. The ILO pilot time use module implemented and adapted here has been tested as a classic rider or add-on integrated within a parent survey (in this case, a truncated version of a model LFS), and it is designed for a variety of modular attachment approaches.

Box 2. Unpaid domestic and care work: Why focus on supervisory care?

Under current international standards on labour statistics, 'work' comprises any activity performed by persons of any sex and age to produce goods or provide services for one's own use or for use by others. Work thus includes both paid work (i.e. employment for profit or pay) and unpaid work (i.e. volunteer work, unpaid trainee work, and activities people do to produce goods and provide services for their own use).

Unpaid domestic and care work refers to non-market, unpaid work carried out in households (Folbre, 2018; Singh, 2020). Unpaid domestic work involves performing routine household jobs such as cooking, cleaning, fetching food, collecting water, and so forth. Unpaid care work includes direct (or active) care involving hands-on or face-to-face personal engagement as well as 'supervisory care' or 'being on-call', often performed in conjunction with other forms of unpaid work.

Such distinctions are economically significant. Direct care is more resource-intensive, whereas supervisory care can often be combined with unpaid domestic work and non-productive activities. Supervisory care is also known to be commonly performed alongside informal (and sometimes formal) employment in home-based work or telework, though it is an impediment to formal employment situations where presence is required in the workplace. While supervisory care may be less intensive than direct care, it can be more time-consuming, as young children and others who are unable to care for themselves cannot be left alone for long periods of time.

There is a growing recognition that unpaid domestic and care work is vital for the maintenance and development of human capabilities, while it generates benefits for recipients and society as a whole. Despite this, such work is largely unrecognised, inadequately valued and not shared equally between women and men and at various levels of society. Women, especially mothers, carry a disproportionate burden of unpaid domestic and care work; they work longer hours than men do, and when they enter paid employment, they often cannot reduce their hours of unpaid work commensurately. This undermines the quantity and quality of FLFP and lowers economy-wide productivity (Alonso et al., 2019). Consequently, SDG 5 on achieving gender equality and empowering girls duly recognizes reducing the burden of unpaid work as an important indicator of progress.

Accurate and reliable data on time use can inform levels and changes in individuals' wellbeing, and it can inform public policies or programs to promote women's economic empowerment and wellbeing while tackling cultural norms, labour market features or lack of public services that undermine FLFP (Ferrant et al., 2014). Such data may, for instance, incentivise public and/or private provisioning of childcare services, thus reducing the constraints of supervisory care that otherwise keep mothers close to home while also allowing them to devote more time to active and developmental care of their children while at home.

Globally, however, this data is lacking; even where it does exist, much of its focus is on activities (useful for measuring direct or active care). Supervisory or on-call responsibilities are generally excluded, significantly underestimating the temporal constraints they pose. Supervisory care is notoriously difficult to capture, as active and supervisory work are often conducted in tandem, but evidence shows that small changes in the wording of survey questions can make a dramatic difference in researchers' ability to isolate it (Folbre, 2018).

1.3.2. Work Package 2: Measuring Agency

The second work package adds an agency module to better understand the factors influencing women's agency over their time use, or the ability to define time use goals and pursue them. Promising recent practices indicate that adding an agency dimension to standardised time use surveys can help unpack whether and how men and women are able to exercise choice over the allocation of their time (Eissler et al., 2021). Measuring agency in this study allows us to better respond to follow-up questions raised in the first phase of the IW–Prospera–UI collaborative study on gender norms.

Time use agency definition. Agency is an individual's ability to define goals and pursue them. Eissler et al. (2021) suggest that a person's agency is reflected in their time use choices, since these choices reveal trade-offs that people are willing to make to pursue their goals. Time use agency aims to 'capture a person's critical consciousness of time use inequities, rights, and personal aspirations; confidence in their ability to (re)allocate their time; expression of voice about allocation of their time; and actual influence over decisions about allocation of their time across the full range of needs and choice-based activities, in line with their personal aspirations' (Sinharoy et al., 2021, p. 8).

In sum, we conceptualised and measured four primary dimensions of time use agency as set out in Sinharoy et al. (2021, p. 15):

1. Critical consciousness – the understanding that men and women have unequal time allocations, and different preferences and time poverty outcomes.
2. Self-efficacy in time use – the feeling that one has the power to make decisions about how to spend one's time.
3. Instrumental time use:
    a. Voice – how often one talks with one's husband/wife about time use choices
    b. Decision-making – how one decides when to undertake certain activities and how much time to spend on them.

We implemented a set of attitudinal questions and a small lab-in-the-field experiment for all respondents of the time use survey.

Attitudinal questions. We asked respondents a set of attitudinal questions to measure their agency over time use across the four main dimensions described above. We grouped the responses to the attitudinal questions (on a scale of 1 to 5, strongly disagree to strongly agree) around the four different dimensions of agency: critical consciousness, self-efficacy, decision-making, and voice. We analysed responses to each individual category and combined the scores for the four types of agency to obtain an aggregate score. A higher score indicates a higher degree of agency.

Lab-in-the-field experiment. A lab-in-the-field experiment combines a standardised lab methodology with a field experiment conducted in a naturalistic setting. Our lab-in-the-field experiment tested whether the presence of a spouse or negotiation with a spouse changes the self-reported time allocation to a hypothetical training program to improve one's income earning potential. Each couple was randomly assigned to one of three treatment groups: private (each spouse responds independently and confidentially), public (each spouse

responds independently but their response is eventually revealed to their spouse), or negotiation (spouses respond in each other's presence and can discuss their responses). Each respondent was asked to respond to four different vignettes inquiring (1) how many hours they would attend a training or seminar if their spouse was at work all day; (2) how many hours they would attend if their spouse was at home; (3) how many hours a neighbour should attend; and (4) how many hours their spouse should attend.

We are interested in the number of hours people report, given that the choice to attend a training program is ultimately a choice to increase one's expected income at the cost of one's own immediate time use. The experiment compared responses across the different social scenarios to which each couple was assigned. By observing whether behaviour differed across experimental groups, we can determine whether levels of agency – particularly related to 'voice', or how much a spouse negotiates with the other about their time use choices, and 'decision-making', or the extent to which people feel they can independently decide the amount of time they spend on activities – affect a person's attendance at a professional training or seminar. For the experimental outcomes, we compared the average responses to each vignette by treatment group and examined whether people responded differently to the vignettes according to their treatment group assignment. We also examined the difference between each spouse's preferred hours of attendance at the training and their spouse's preference for them to highlight any divergence of expectations within the household, and the factors that drive it.

Taken together, this information helps us understand the characteristics that correlate with different levels of time use agency, as well as barriers to women's economic engagement in general. For example, the data gathered may be used to monitor and evaluate the effect of investments in the expansion of jobs and businesses on women's wellbeing; inform campaigns to transform care into a shared responsibility within and outside the home; and/or design better policies and programs to address gaps in child and family care.

1.3.3. Policy and Programmatic Focus Questions

Measuring agency alongside time use can help answer a number of questions about women's economic empowerment and labour decisions. Here, we pose several policy and programmatic questions and discuss how our current project may help address them.

1. **Policy/program question 1:** Why do women participate less in the formal labour market than men?

In Indonesia, it has long been the case that women participated in the formal labour market at a lower rate than their male counterparts. This disparity has negative consequences for the economy as a whole and for individual households' economic outcomes (Legarde & Ostry, 2018). Our previous work employed an intersectional approach to understand the shifting norms and labour decisions within the household; it led to many new questions about how women's agency influenced such decisions. In this study, we are specifically interested in

how unpaid work and supervisory care impact FLFP, including during a crisis like the Covid-19 pandemic.

Women's entrance into paid employment may increase resources available to them and their families, and likely increases macroeconomic growth. However, women who intensify their participation in the labour market may not experience a commensurate decline in their non-labour-market responsibilities, thereby reducing their individual wellbeing by leaving them with less time for personal care, sleep and leisure. The need to balance paid and unpaid work may also limit the jobs women are able to take on, reduce their overall productivity and stifle their opportunities to progress at work. Such patterns intensify inequalities among women, particularly between well-educated, high-earning women who can afford to outsource their domestic and care responsibilities, and poorly educated, low-earning women who may have less flexibility (Folbre, 2006).

To address the above policy question, we hypothesise that women must assume the great majority of unpaid care responsibilities and thus have less time to engage in paid work, even though they may be interested in obtaining paid work for their own economic independence and/or to support their families. Alternatively, their care work burden reduces their ability to take up quality jobs, even when such jobs are available.

We employ a time use survey with special attention paid to a comprehensive accounting of intra-household and intra-familial unpaid domestic and care work, including neglected dimensions, to investigate the following related sub-questions:

(i) How much time do women and men in a household spend on unpaid domestic and care work (both active and supervisory care)? How does this differ by household composition or type (e.g. presence of children and/or persons with disability, elderly, rural or urban)?
(ii) How does (i) impact the amount of time women and men can dedicate to work for pay? How much does the time allocated to paid work activities vary with gender?
(iii) How much agency do women and men exercise over the allocation of time dedicated to paid work?

2. **Policy/program question 2:** What policy levers and programmatic priorities can promote women's equitable participation in the economy?

To formulate effective and efficient policy approaches and social programs to improve FLFP, there is a need to better understand the nuanced cultural, social, logistical and economic reasons why women participate less in the labour force.

There are many examples of how countries and development partners are using TUSs to influence policies, as outlined in the table below. They range from overarching strategies to justify the provision of more child and elderly care, extend maternity leave and offer flexible

working arrangements, to evaluating the impact of family leave on time use and household income earning strategies.

Table 1.1. Summary of time use studies in selected countries

No.	Country	Method of data collection	Data users	Examples of how time-use data influences policies
1	Albania (2010–2011)	• Face-to-face interviews	• Ministry of Social Welfare and Youth • Social Insurance Institute	• National Strategy and Action Plan for Gender Equality: to measure the proportion of time spent on paid and unpaid work. • Children and care policy: to justify the provision of more childcare and elderly care, and to support the enforcement of up to four-month maternity and paternity leave for a parent of a child between birth and age six.
2	Cambodia (2004–2005, 2007, 2009–2014)	• Time diary sheet attached to national household survey	• Multiple stakeholders related to gender and care policies	• Children and care policy: to measure the time spent on care work. • Eliminating and preventing child labour: evidence on the total burden of (paid and unpaid) work on children and youth. • Gender policy: maternity leave for women employed in the formal sector; and the obligation of enterprises employing a minimum of 100 women to construct a nursing room and a childcare facility.
3	Finland (1979, 1987–1988, 1999–2000, 2009-2010)	• Computer-aided telephone • Face-to-face interviews • 'Leave behind' diaries	• Social Security Institution • National Consumer Research Centre • Family Federation • National Broadcasting Company • Ministry of Education and Culture • Ministry of the Environment	• Production of the household satellite account: to value own-use production and use of market services. • Child and family policy: to identify the impact of family leave policy on time use of families. • National program to increase the attractiveness of work life (Veto program): to understand whether people exercise work and family life balance.
4	Moldova (2011–2012)	• Face-to-face interviews • 'Leave behind' diaries	• Ministry of Health, Labour, and Social Protection • Multiple ministries involved in National Employment Strategy	• National Strategy on Gender Equality: to empower women with small children, establish childcare centres in the workplace, enhance childhood education, and eliminate gender pay gaps. • National Employment Strategy: to justify the need for innovative and flexible forms of employment and support women in business.

No.	Country	Method of data collection	Data users	Examples of how time-use data influences policies
5	Uruguay (2007, 2011, 2013)	• TU module attached to national household survey	• Ministry of Social Development • Centres for the Integral Attention of Early Childhood Development and Family (CAIF) • Multiple stakeholders in national care ecosystem	• Care policy: time use data helps the government prioritise early childhood programs. In Uruguay, the TU module also collected information on the social organisation of care.

Source: Data2x (2018)

We also investigate the following sub-questions:

- In the face of new economic opportunities (e.g. training to secure a new job, higher income), how do women and men negotiate time allocation?
- Does men and women's behaviour conform to societal expectations about what men and women should do in the face of such opportunities?
- Do women and men's decisions to take up new opportunities change depending on the presence of their spouse or the need to negotiate with them?

3. **Policy/program question 3:** How can Statistics Indonesia efficiently administer data collection on time use in the near future as part of Sakernas?

Within the context of this project, Indonesia is participating in global efforts to produce guidance on how to collect data in emerging and developing countries. Thus, Indonesia will not be passively accepting methods/guidance developed in a different national context.

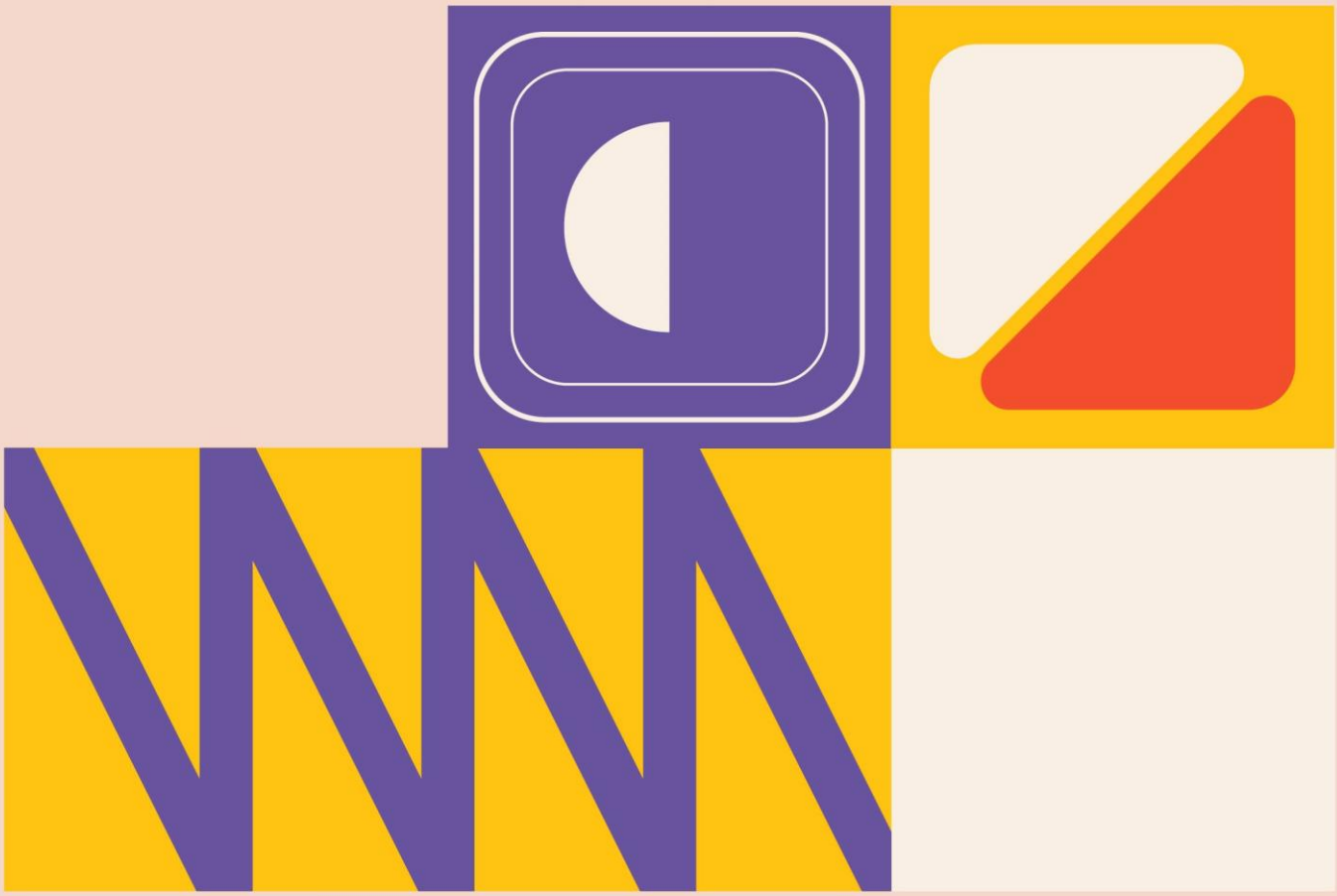
Other objectives of this pilot study are:

- Explore how different dimensions of intra-household and intra-familial unpaid care work, including supervisory dimensions of care (see Box 1), are perceived and articulated (including social norms around kin-based obligations and responsibilities classifiable as child and adult care).
- Inquire into how couples in two-parent households report periods of responsibility for care provision when both are present.
- Inform refinements to ILO's light time use diary's preliminary pre-coded activity listing developed for piloting, as well as the wording of other items.

Assess the ease with which informants recall and express the timing and duration of activities. For example, when and for whom is 'clock time' a suitable organising framework? What other temporal logic is at work? When and for whom?

Section 2

Summary Of Methods: Time Use and Agency Survey Instruments, Implementation and Analysis



2. Summary of Methods

This section provides an overview of the survey instruments used to measure both time use and agency, the implementation of these instruments, the sample used and the methods employed to analyse the resulting data.

The refinement of the survey instruments and implementation occurred in three distinct phases: proof-of-concept testing, cognitive testing, and piloting.

The first step consisted of a small set of survey tests acting as proof of concept. During this phase, the time use and agency instruments were tested on a small, non-random set of people (fewer than 10). Using feedback from this phase, the instruments were refined and cognitive testing (i.e. assessing the development and refinement of instruments) began in August. After the cognitive testing phase,[3] the instruments and implementation strategies were further adjusted. We simplified language, changed the order of the survey modules and refined the experimental vignettes to make them more consistent and concise.

After revisions from the cognitive testing round, we implemented the pilot survey on an urban sample of 452 individuals (226 male-female couples or dyads). The data generated in this round includes rich data on time use at the individual level, with special attention paid to unpaid work and various types of care. We also collected data on agency for both men and women related to both their time use preferences and their attitudes about time use and agency.

In this section we provide an overview of the strategies used to analyse the resulting data, which include visualisations, descriptive means and means comparisons, and regression analysis.

2.1. Survey Design and Implementation: Time Use Instrument

While interest in the potential for time use surveys to inform labour force participation studies has increased in recent years, few countries have fully integrated them into their national household survey programmes, as they can be challenging and costly to implement. Indonesia does not conduct a standard time use survey on a regular basis. The most recent time use survey was piloted in 2005.

The time use module tested in this project is aligned with the latest international standards and guidance on the measurement of unpaid domestic and care work, as set out in the 19th International Conference of Labour Statisticians resolution on statistics of work, employment and labour under-utilisation (ICLS, 2013). This resolution, adopted in October 2013, was a major step change in the measurement of productive activities for national statistics.

[3] A full report on the cognitive testing phase can be found in the *Report on Time Use Agency Cognitive Testing*, an online appendix to this report.

The resolution also provided a reference definition of the term 'work' (a first in international labour statistics), which was defined as 'any activity performed by persons of any sex and age to produce goods or to provide services for use by others or for own use' (ICLS, 2013, p. 2). This definition encompasses all production of goods or services for use by oneself or others. It includes both paid and unpaid productive activities and applies regardless of the (in)formality or (il)legality of the sector and status of the work, or the economic unit in or for which it is performed. It excludes any activities that do not involve production of goods or provision of services (e.g. begging and stealing), relate to 'self-care' (e.g. personal grooming and hygiene), or meet the 'third-person criterion' (that is, activities that cannot be performed by another person on one's own behalf, such as sleeping, learning, recreation, etc.). This concept of work is aligned with the general production boundary as defined in the 2008 System of National Accounts (SNA).[4]

Work, as defined by the 19th ICLS, is categorised into five different forms based on (1) the beneficiary of the work (oneself and one's family, or others), and (2) whether there is an expectation of pay or profit, as follows:

- **Own-use production work:** Work done to produce goods or services for the benefit of oneself or one's household or family without pay or profit. This includes:
    - Own-use provision of services, such as (i) unpaid domestic work (e.g. food and meal preparation, shopping for family members and cleaning of one's dwelling), and (ii) unpaid care work including active caregiving (e.g. feeding, cleaning and providing physical care) and supervisory (or passive) care.
    - Own-use production of goods, such as subsistence farming, where the goods are consumed mainly by the family.

- **Employment:** Work done in exchange for pay or profit, including self-employment, paid employment and work in a family or household business.

- **Unpaid trainee work:** Work done without pay to gain workplace experience or skills in a trade or profession.

- **Volunteer work:** Unpaid, non-compulsory work for the benefit of others (i.e. the community, formal/informal organisations and people who are not members of the volunteer's household or family).

- **Other work activities:** Any activity to produce goods or services for others without pay not already covered above, for example, compulsory unpaid work ordered by a court.

[4] 2008 SNA is an international statistical standard for national accounts.

The adoption of the new reference definition of work greatly expanded the scope of work situations within the remit of labour statistics. One important result of the changes was the need to integrate data on *own-use provision of services* (the term used to refer to unpaid domestic and care work) in official national labour statistics. Therefore, in recent years the measurement of unpaid domestic and care work has received greater attention in official statistics. There is a growing interest in unpaid domestic and care work in national and international policy circles, largely resulting from the acknowledgement of its significant economic contribution.

Similarly, there is growing recognition of the problematic status quo whereby, globally, women and girls contribute over three-quarters of the total hours spent daily on unpaid domestic and care work, to the detriment of their labour force participation, access to formal employment and decent work, and involvement in the wider public sphere.

The ICLS standards permit improved analysis of gender-based inequalities in labour force participation, employment characteristics, the division of paid and unpaid labour, and total work time, as well as assessments of the relationships and trade-offs between participation in, and access to, paid and unpaid work.

2.1.1. Time Use Instruments: Available Options and Instruments Piloted

TUSs record how individuals spend their time for a specified period, usually a 24-hour day or a seven-day week. The resulting data provides insights into the routines of daily life, illuminating patterns, interdependencies, inequalities and trade-offs in the time people allocate to paid work, unpaid work and non-work activities. Since time use is highly gendered, TUSs are an important source of gender statistics.

Two main survey approaches are available for time use measurement, termed ‘diary’ and ‘stylised’.[5] A variety of different formats are available within each broad approach. In recent years, hybrid diary instruments, which combine aspects of each, have been the subject of growing interest.

Diary approaches to time use measurement are characterised by the chronological reporting of time use over the 24 hours of a day. The respondent records (if self-administered) or reports (if interviewer-administered) how they spend (if contemporaneous) or spent (if retrospective) their time from a designated moment – conventionally 4 am or midnight – or from the moment they wake/woke up. The exercise is sometimes repeated for multiple days.

Within this broad approach, several diary formats can be used. Formats vary according to whether episode timings are open or pre-defined. In the former, the respondent records (or reports) the start and end times of each activity. In the latter, the 24 hours of the day are

[5]Approaches used outside of survey contexts, such as immersive observation, experience sampling methods (ESMs) and deployment of wearable technologies are omitted from discussion, as they are beyond the scope of this research.

divided into (usually) equal[6] intervals of between 5 and 60 minutes, ready to be populated with respondents' activities.

Formats also vary according to whether activities are open-coded or predefined. In the former, activities are recorded verbatim, in the respondents' own words, and coded at the data entry stage. In the latter, the respondent (if self-administered) or interviewer (if interviewer-administered) selects the code that most closely corresponds to each activity from a predefined list. The convention is to refer to diaries with pre-designated activity codes as 'light' or 'lite' diaries and to open-code diaries as 'full' diaries. Further variations, applicable to both light and full diaries, include the presence of fields to record multitasking, that is, activities undertaken simultaneously or overlapping, and/or contextual information such as location, presence of others, beneficiary, remuneration, linkage with employment, and more.

In a **stylised** time use format, respondents report participation in, and estimate the total amount of time dedicated to, an activity or activity class over a specified reference period, usually a seven-day week or a 24-hour day. While diary formats record the timing, sequence and duration of activities, stylised formats provide only participation and total duration. Stylised questions may be framed in terms of a specific reference period, such as, 'Yesterday (or last week), how much time did you spend doing (activity X)?' Alternatively, questions may be phrased more generally in terms of usual or typical practices, such as, 'How many hours a day (or week) do you usually spend doing (activity X)?'[7]

Hybrid diary instruments include direct question items, characteristic of the stylised approach, alongside a core diary. Such combined approaches have developed in response to a range of different measurement, operational and statistical considerations. One such consideration is the inadequacy of the diary reference period for measuring the prevalence of activities that occur on a less than daily or weekly basis. Time use research on volunteer work, performance of civic duties, in-work training, attendance at events or exhibits, and illicit or illegal behaviours has employed direct, stylised questions and a longer reference period to supplement a core diary to good effect (see Juster et al., 2003).

Stylised questions can also serve an important recovery function, either as targeted probes or longer, checklist-based probes. In both cases, the goal is to address under-identification of activities omitted from spontaneous diary reports. The direct probes are sequenced subsequent to initial diary completion, with newly recalled/reported activities flagged as recovered items in the original diary. Targeted probes are increasingly included in hybrid diaries as a means to recover supervisory or passive care responsibilities.

[6]In some cases, longer intervals are assigned to night-time hours, during which a majority of respondents are presumed to be sleeping, in order to condense the diary exercise.

[7]It is generally agreed that 'typical' period questions place greater cognitive demands on respondents than specific period questions (where the specific period is sufficiently short and recent), though there are dissenting views. Prior research also indicates that some respondents may report a 'typical' period despite being asked about a specific period. See Chang and Krosnick (2003) for a review of the evidence on typical vs specific time periods.

Checklist-based probes perform a similar recovery function and may even substitute for dedicated 'simultaneous activities' diary fields in situations where the comprehensive collation of simultaneous activities is assessed to be excessively burdensome and/or error prone. Where the analytical interest is in recovering one or more particular simultaneous activity classes – and not in the performance and characteristics of multitasking/simultaneity/activity segues and sequencing – the combination of a diary format to record main activities and direct questions to identify secondary activities of interest may offer a viable compromise. As is the case generally, the use of direct prompts requires careful testing, as there is a risk that they may bias responses, skewing respondents' recall and/or communicating social desirability of certain activities.

The choice between a diary or stylised time use instrument and decisions about the content and organisation of the selected form are guided by the measurement objectives as well as available resources.

In our study, we asked people about their time use allocations directly (i.e. proxy response was not permitted, in line with ILO pilot protocols). Previous experience suggests that generating reliable time use data is challenging because of response errors originating in decay of recall and/or social desirability (Bell et al., 2019). There is often a trade-off between accuracy and complexity or resource intensiveness. Full time use surveys – the historical gold standard – are very costly to administer, especially in emerging and developing countries where lower numeracy prevents respondents from accurately filling out complicated surveys. People's time use can also vary significantly and depend on when a survey is administered.

Given these challenges, we implemented and tested a pilot light time use diary designed to be less burdensome for BPS to administer as part of Sakernas. This light time use diary also contains improved questions to measure supervisory care.

2.1.2. Time Use Instruments: Pilot Implementation

The ILO's pilot time use module has been designed for face-to-face interviewer administration using a CAPI tool developed in CSPro software, meaning that the information provided by respondents is entered by interviewers via an electronic tablet. The use of a CAPI tool, together with the pre-coding of activities and the use of fixed time episodes, results in a much faster turnaround time from data collection to analysis and dissemination. It also has the potential to minimise data entry errors thanks to several features that reduce erroneous or missing data fields (including restrictions on data entry format, automated updates and warning flags activated by incongruous entries).

The time use module is embedded within a wider pilot survey composed of four sections:

1. Section 1 collects background information about the selected household. Only one member of each household is interviewed for this section. This person provides basic information about the dwelling and the characteristics of each inhabitant.
2. Section 2 collects more detailed background information on labour force participation and employment characteristics for up to two eligible household members (one couple dyad, probabilistically selected). It is a truncated version of the model LFS.[8] Each eligible household member must complete this section for themselves.
3. Section 3 is the time use module. It records what the selected individuals did in the day before the interview. Each selected household member must complete this section for themselves.
4. Section 4 is the agency module. It asks respondents a series of attitudinal questions and measures responses to vignettes after respondents are randomly assigned to an experimental group.

The embedding of the time use module within a shortened survey reproducing standard LFS items is intended to contextualise the questions contained in the time use modules for respondents and support meaningful analysis of time use module performance. Table 2.1 summarises the key features of the pilot time use module utilised in the present study.

Table 2.1. Key features of the pilot time use module

Feature	Summary
Mode	Face-to-face interviewer-administered
Administration	CAPI, CSPro
Format	Hybrid light diary
Reference period	One 24-hour day from 04:00 to 03:59
Activity timing	96 fixed episodes, each 15 minutes long (multiple rosters record simultaneous activities separately in 15-minute episodes)
Activity coding	41 pre-coded activities + 'other, specify' (aligned to ICATUS-16 coding scheme)
Contextual items	Five *conditionally activated* contextual items are targeted for pilot testing (location, co-presence, beneficiary, job linkage, income generation).
Treatment of simultaneity / multitasking	Multiple activities can be recorded as occurring simultaneously (i.e. during the same 15-minute episode).
Treatment of supervisory / passive care time	A dedicated recovery series is activated after completion of the diary day. Separate items target supervision/passive care of children and adults.

[8]Model LFS resources are available here.

The time use module records everything the respondent reported that they did on their diary day. The diary day is divided into 96 consecutive and non-overlapping fixed episodes of 15 minutes each, from 4 am on the day before the day of the survey administration until 4 am on the day of the survey administration.

The initiation of the diary day at 4 am is a standard convention in time use research (an alternative convention is to begin at midnight). Generally – but not always – respondents are asleep at 4 am, so by beginning at this time the diary catches the start of the waking day. The time use module begins by asking respondents to recall what they were doing at 4 am. Once the activity is recorded (selected from a dropdown menu), the respondent is asked until when they did this activity. The start time of each activity is automatically updated in the question wording based on the ending time entered for the prior activity. The ending time of the activity is recorded via a dropdown menu that is updated automatically to exclude times prior to each activity start.

The pilot time use module is designed to comprehensively record respondents' time use for one 24-hour day. The module is programmed with a dropdown menu of 41 pre-coded activities, along with an option for 'other, specify' (code 42) featuring an open field to record activities that do not correspond to any of the 41 specified codes. Codes are aligned to ICATUS-16 at the two- or three-digit level, and all nine of the ICATUS major divisions are included (ICATUS has 165 groups classified into 9 major divisions and 56 subdivisions). The level of disaggregation for the pilot activity codes varies by activity domain, with a greater number of codes dedicated to priority domains and/or domains known to be prone to undercounting (such as unpaid domestic and care work), and fewer codes assigned to domains less prone to measurement error (such as paid work) and/or of lower substantive priority for the measurement objectives, where less disaggregation is required at the output stage.

The pilot diary instrument selected for the study in Indonesia permitted multiple simultaneous activities to be recorded via a dedicated questionnaire item, 'Were you doing anything else while you were [spontaneously reported activity]?' This item was included to permit multitasking to be recorded. For instance, the respondent may report that *while* doing paid work or unpaid housework (e.g. cleaning, cooking, laundry) or care work (e.g. feeding a young child), or while travelling/commuting, they were listening to the radio or watching TV. Or they may report that they were talking with colleagues while having lunch, or grazing livestock while weeding an adjacent field or harvesting fruit in a nearby orchard, or cooking an evening meal while cleaning up the kitchen and supervising a child to make sure they were doing homework. Interviewer training highlights that care should be taken to ensure that only genuinely simultaneous activities are recorded for the same episode(s) of time, with interruptions to an activity recorded sequentially.

To be able to correctly assign respondents' time use to higher-level activity domains (such as unpaid domestic work, unpaid care work, employment [i.e. work for pay or profit], volunteer work, study, personal care, etc.), it is often vital to have access to contextual

information. Contextual items may also aid respondent recall, particularly location and co-presence (Tulving, 1983). The pilot time use module included five *conditional* context items (the final instrument is expected to have fewer contextual variables, based on the results of piloting). Contextual items are activated only when a coded activity is eligible for classification to multiple higher-level domains.

The five conditional context items included in the pilot instrument are:

1. location[9]
2. co-presence (with whom)
3. beneficiary (for whom)
4. job linkage
5. income generation.

For some activities, no contextual items are activated; for others, two or three may be required to correctly classify activities to higher-level domains. The following are additional features of the pilot time use module utilised in the study:

- A passive care recovery series prompts respondents to recall time when they were minding, watching over or supervising dependents, but not directly interacting with them. It asks separately about times when the respondent had supervisory/passive care responsibilities for children who are household/family members (anyone aged under 18) and dependent adult household/family members (aged 18 and over) who require assistance from others to undertake daily activities due to illness, injury, frailty or disability, whether temporarily or long-term. The recovery sequence is activated only once the diary has been filled out in full. A separate roster permits the timing, sequence and duration of supervisory/passive care episodes to be recorded in fixed 15-minute episodes.
- A short quality audit series asks respondents to confirm – and revise, if desired – the timings of key moments in the diary day (wakeup time, bedtime, mealtimes, etc.).[10]
- A short 'typical day' series asks respondents to report whether the diary day reported in the survey is unusual in any way. It permits any apparent anomalies in the resulting data to be assessed.

[9]The *location* item includes a quality control feature to limit underreporting of travel time (a known tendency in recall diaries). If the interviewer codes a change in location in the absence of an intervening travel episode (for instance, a respondent reports being at home readying a child for school [location: own home] and next reports employment [location: office]), a warning will activate, requiring the interviewer to probe for intervening travel episodes. Coding allows for transit to be recorded where mobility is characteristic of the activity. For instance, if the interviewer codes the respondent being at home readying a child for school [location: own home] and next reports the respondent accompanying the child to school [location: in transit] before attending employment [location: office], no warning will appear.

[10]An item embedded at the beginning of the individual labour force module invites respondents to spontaneously report the time, with a related item for the interviewer to code the means by which the respondent estimated the time. This provides a way to assess respondents' comfort with clock time and performs a quality assessment role for the pilot survey.

The following are the seven steps to record activities in the light time use diary, though in practice the interview adopts a conversational style, with the interviewer probing for additional information as needed:

1. We start by asking the respondent to report what they were doing at 4:00 and record the first activity reported, termed 'Activity 1' for clarity's sake (without imposing a hierarchy), as 04.00–05.30: Sleeping, followed by 05.30–06.00: Eating breakfast. The respondent reports their time use in their own words, and the interviewer selects the appropriate activity label from a drop-down listing of pre-coded activities.
2. We ask respondents whether they did anything else while engaged in the first spontaneously reported non-sleeping activity (in our example, eating breakfast). For instance, a respondent might report feeding their children at the same time.
3. For activities that exceed a single 15-minute episode, we ask the respondent to indicate the duration of the subsequently reported activity/ies, allowing for simultaneous activity/ies to vary in length. For instance, the respondent may report eating breakfast from 05.30 to 06.00 and eating their own breakfast while feeding their children from 05.45 to 06.00. The pilot module places no restrictions on the number of simultaneous activities that can be recorded for a single episode, as this is a topic of interest for the broader ILO pilot project for which the pilot instruments were prepared. In practice, genuinely simultaneous activities tend to be limited by practical considerations.
4. Once the activity/ies are recorded as above, conditional contextual items are activated in the CAPI tool, directing the respondent to probe for further details necessary to support the classification of activities. Some context items are active for all coded activities (location), while others are only activated when relevant to the eventual higher-level coding of activities (with whom, for whom, job linkage, income generation).
5. The interviewer then asks about the respondents' time use beginning at the subsequent interval, in this case 06.00. Steps 1–4 are repeated until the diary day is complete.
6. The interviewer then moves to the recovery section and asks whether the respondent had any supervisory care responsibilities during the diary day. If yes, the respondent must indicate the time(s) when they did supervisory care. A separate roster records the supervisory care time in the same 96 15-minute slots.
7. A final validation sequence assesses the stability of respondent recall on select items.

The pilot time use measurement tools have been developed by ILO to permit the evaluation of alternative measurement approaches introduced to the light format instruments to ease the response burden and resource intensity of traditional full diary approaches while maintaining data quality. Multi-country piloting has focused on several key design features and/or evidence gaps in modular time use measurement, including:

- concordance of estimates produced by a range of different modular time use formats

- alternative measurement strategies to reduce known underreporting of unpaid domestic and care work (via contextual items, items for simultaneity, and dedicated probing items)
- alternative measurement strategies to produce estimates of total (paid and unpaid) work time
- adequacy/redundancy of the pre-coded listing and contextual items in the data input and analysis phases
- implementation of a fixed 15-minute episode format for diary reporting, including the fungibility of clock time in a range of settings
- data quality impacts of modifications to sample designs and field operations related to the requirements introduced by a time use module
- workability of alternative schemes to minimise non-response, with days of the week as a sampling dimension.

2.2. Agency Instruments

Women's labour force participation patterns are, in part, formed as a result of care constraints faced by households. However, there are likely other factors that play a role in people's time use decisions. Time use agency, consisting of critical consciousness, self-efficacy and instrumental time use (including both voice and decision-making), reflect how people choose to spend their time and how this time use is negotiated within the household. While we explore care constraints and unpaid labour constraints to labour force participation, we also examine the role of time use agency in observed patterns of time use to provide a more complete picture of what barriers to labour force participation may exist. Understanding how people's time allocations may change or adjust (potentially strategically) when information is private, public or negotiated helps us understand how agency may affect observed time use allocations.

People's agency, and specifically their time use agency, is not easily observed using standard data collection methods. Further, we know that private information like time use agency is not well accounted for in models of the household. It is only recently that researchers and practitioners have attempted to carefully measure time use agency. Reasonable measures of agency, or even proxies for it, do not exist in administrative or previous survey data. This project is unique in that it combines detailed time use data with a focus on unpaid labour and supervisory care, and measures of agency. In this section, we describe the methods and instruments we use to measure different forms of time use agency. We also provide a justification for these methods and some preliminary hypotheses.

Development practitioners employ a variety of methods to measure so-called 'unobservable' variables. We draw on some of these techniques in our study. Behavioural experiments (such as dictator games, preference games and ultimatum games) are commonly used to detect

preferences for things that are difficult to observe, such as agency. Role-playing games and agent-based modelling are examples of behavioural experiments used to simulate complex socio-ecological systems in an understandable and feasible way in order to observe participants' behaviour and understand their decisions, as well as provide a platform for the discussion of the implications of their actions. Role-playing games have been used to facilitate informed decision-making, scenario planning and joint management of a common resource. These games can also encourage participants to reflect on their own circumstances, raise awareness about their place in complex systems, and come to understand others' perspectives (Wesselow & Stoll-Kleemann, 2018). For these reasons, economists and social scientists use role-playing games in creative ways to measure agency and other aspects of women's economic empowerment.

While qualitative interviews record self-reported preferences, behavioural experiments have the added benefit of exposing participants to different scenarios, encouraging them to question and evaluate their own understanding, knowledge and experience of gender relations. At the same time, the experimenters are able, to some extent, to control the parameters of the environment, thus limiting the potential inconsistency of participants' behaviour (Smith, 1989). If designed properly, behavioural experiments can produce robust information without necessarily adding to the cost.

Many behavioural experiments are rightly criticised for operating in a social vacuum. We attempt to shield our design from this criticism by interlinking the experimental approach with findings from the time use survey as well as answers from the direct agency question module (i.e. the attitudinal questions). Another rationale for simultaneously using attitudinal questions and an experimental approach is that while we think that the lab-in-the-field experiment may be a more theoretically justifiable and consistent way to measure agency, it has not been tested in the context of measuring time use agency specifically, and therefore presents some risk. We therefore combine the experiment with the more field-tested method of asking people to signal their agreement or disagreement, on a scale of 1 to 5, with a set of attitudinal questions.

We use attitudinal questions similar to those used by Sinharoy et al. (2021) to gauge people's perceptions and attitudes about the three main types of time use agency. Asking both men and women, for example, whether women are expected to work longer hours or be the primary caregivers helps us understand how people's attitudes might affect their time use allocation.

Our experimental approach allows us to learn the most about instrumental time use. Understanding how differing social scenarios affect the answers given by men and women helps us understand the impact of 'voice', or how much a spouse negotiates with the other spouse about their time use choices, and of decision-making, or the extent to which people feel they can independently decide the amount of time they spend on activities – in this case, a professional activity like a job training session. The experimental portion aims to test whether voice and decision-making, two important components of instrumental time use,

change a person's time use decisions. If we hold childcare constant and vary the information and social interactions that spouses have around time use decisions, we can start to gain an understanding of whether voice and decision-making affect time use decisions, and, potentially, how much.

Our primary efforts to measure critical consciousness and self-efficacy are captured in the attitudinal section of the agency module. In the pilot survey, the attitudinal questions were asked directly after the time use questionnaire, and the lab-in-the-field experiment was conducted last (but within the same day). The results from the pilot implementation of these instruments are presented in Section 4.

2.2.1. Attitudinal Questions

As detailed in the following tables, we grouped attitudinal questions in four categories: critical consciousness (who should work and how much women should work), self-efficacy in time use, instrumental time use, and voice. Answers to the questions were measured on a Likert scale of 1 to 5.

Table 2.2. Attitudinal questions measuring self-efficacy

To what extent do you agree or disagree with the following statements?					
	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
a. You have the ability to change your daily schedule.	1	2	3	4	5
b. You can ask a household member to do some of your household duties.	1	2	3	4	5
c. You can ask a household member to help you take care of a child or other family member.	1	2	3	4	5
d. You can change the amount of time you spend on paid work.	1	2	3	4	5

Table 2.3. Attitudinal questions measuring decision-making/instrumental agency

To what extent do you decide when and how much time you spend on the following activities?					
	Someone else decides for me all the time	Someone else decides for me most of the time	Discussion between me and someone else	I decide most of the time	I always decide for myself
a. Your daily schedule	1	2	3	4	5
b. Your agricultural activities (RURAL ONLY)	1	2	3	4	5
c. Your job or other paid work	1	2	3	4	5

d. Household duties, such as cooking, cleaning, washing clothes, or collecting water or cooking fuel	1	2	3	4	5
e. Caring for household members, such as children or elderly family members	1	2	3	4	5
f. Shopping, such as going to the community market	1	2	3	4	5
g. Attending a social gathering in the community, such as a wedding or other celebration, or a community meeting	1	2	3	4	5
h. Leisure activities, such as listening to music or chatting with friends; resting; sleeping other than resting	1	2	3	4	5

Table 2.4. Attitudinal questions measuring critical consciousness

To what extent do you agree or disagree with the following statements?					
	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
a. Men are responsible for providing for the family and earning an income; women are responsible for taking care of children and doing domestic work.	1	2	3	4	5
b. Women can work, even as the main earners.	1	2	3	4	5
c. Women can work, but only as secondary earners.	1	2	3	4	5
d. In this community, women are expected to work longer hours than men.	1	2	3	4	5
e. Compared to a woman, a man can change his daily schedule more easily.	1	2	3	4	5
f. Because of their responsibilities, women generally sleep less than men.	1	2	3	4	5
g. Women’s responsibilities take more time than men’s responsibilities.	1	2	3	4	5

Table 2.5. Attitudinal questions measuring voice

To what extent do you agree or disagree with the following statements?					
	Strongly disagree	Disagree	Neither disagree nor agree	Agree	Strongly agree
a. A spouse can change their daily schedule without consulting the other spouse.	1	2	3	4	5
b. A spouse can ask the other spouse to do more household duties so they can do something they want to do.	1	2	3	4	5

c. I can ask a household member to help me take care of a child or other family member.	1	2	3	4	5
d. I have total control to change the amount of time I spend on paid work.	1	2	3	4	5

2.2.2. Lab-in-the-field Experimental Design

In this section, we describe the strategy, assumptions and justifications of the experimental design. We further outline potential hypotheses and outcomes related to each of the treatment groups.

Our experimental approach, which employed vignettes or stories, allowed us to hold constant childcare, the type of opportunity presented to the respondent, and tasks that each respondent will forgo to attend the seminar/workshop. In each vignette, we offered the respondent an opportunity to attend a seminar that could improve the income earned by the household. The opportunity was based on a real training program that people are familiar with, Kartu Prakerja,[11] managed by the Indonesian Government. As such, a choice to attend the training program was a choice to increase one's expected income at the cost of one's immediate time use.

In each vignette, we explicitly stated, 'You cannot bring your children to the meeting, but assume you can find a responsible caretaker.' The intention of this statement was to eliminate childcare as a constraint to participation, thereby focusing on people's willingness to attend.

We aimed to create two types of variation during the experiment. The first source of variation was the content of the four vignettes. In the first vignette, each person was asked how many hours they would attend a seminar if their spouse was at work all day. The second vignette asked how many hours they would attend if their spouse was at home. The third vignette asked how many hours a neighbour should attend (to gauge social norms), and the last vignette asked how many hours the respondent thought their spouse should attend. Understanding people's expectations about their own participation helps us observe their agency, especially when compared with their spouse being at home or not. Examining response patterns to the third vignette (how many hours a neighbour should attend) helps us understand social norms, and lastly, asking about the spouse's attendance helps us understand intra-household expectations. Each respondent was exposed to this same variation across vignettes.

The second source of variation was the social and communication situation to which we randomly assigned each couple. Each treatment group varied in whether and how

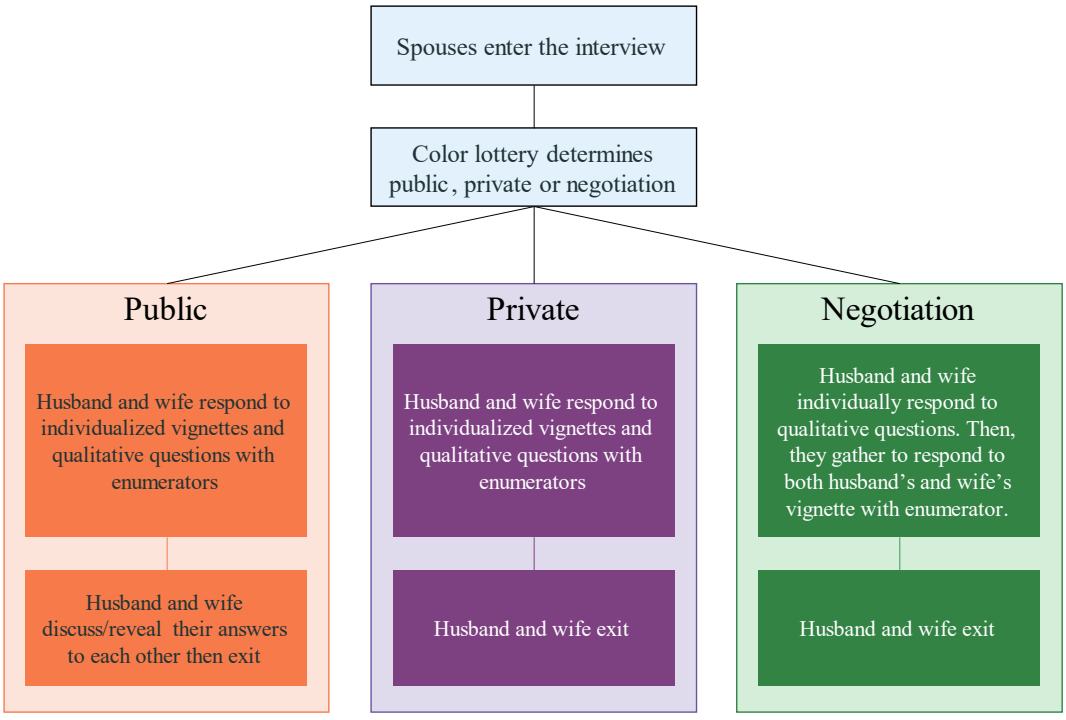
[11] Kartu Prakerja is a work competency and entrepreneurship development program aimed at jobseekers, workers/labourers affected by termination of employment, and/or workers/labourers who need competency improvement, including micro and small business actors (see https://www.prakerja.go.id/tanya-jawab/tentang-kartu-prakerja).

respondents' responses were shared between the individuals making up each couple and the mode of communication by which this information was shared. By observing whether behaviour differs across experimental groups, we can examine whether levels of agency – particularly related to 'voice' (how much a spouse negotiates with the other about their time use choices) and decision-making (the extent to which people feel they can independently decide the amount of time they spend on activities) – affect a person's willingness to attend a professional training or seminar.

The experimental element was that couples, upon entering the interview, were randomly assigned to one of three treatment groups that determined the social construction under which they would respond to each vignette. The social construction of the experiment asked people to respond to the vignettes either (1) independently and confidentially, or (2) independently, but with their answer later revealed to their spouse, or (3) after direct negotiation with their spouse. Each spouse then reported their preferred hours of attendance to the enumerator. The random assignment ensured that couples would not select the most comfortable social construct, thus invalidating the results.

While we are interested in the number of hours reported by the participants, the logic of the experiment was to *compare* the hours reported given the different social scenarios to which each couple was assigned. We can imagine that the negotiation group is similar to real life, where couples negotiate time use among themselves. We are interested in whether the answers, particularly for women, changed depending on whether they (a) responded to enumerators' inquiries without their husband, but knew he would see the answers, or (b) responded to enumerators' inquiries knowing that their answers would be observed only by the researchers. For clarity, a diagram of the experimental design is presented in Figure 2.1.

Figure 2.1. Experimental design



Incorporating lessons from cognitive testing resulted in the survey instrument and vignettes being scripted as follows:

Vignette 1. Imagine you are offered the opportunity to attend a seminar or a rotating savings (e.g. *arisan*) meeting that could increase your income (e.g. a job training workshop similar to Program Kartu Prakerja in some cases, or a loan) on a normal weekday. This would mean not doing some of your daily tasks or doing them later. You cannot bring your children to the meeting, but assume you can find a responsible caretaker. The workshop will run for eight hours, but you can choose to attend only parts of the workshop, and it would still be useful.

EXP.V1.Q1.a. Please choose how many hours you would attend, remembering that you will need to give up other tasks to attend, if **your spouse was gone all day because they're at work**.

0 hours	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours	7 hours	8 hours

EXP.V1.Q1.b. Please choose how many hours you would attend, remembering that you will need to give up other tasks to attend, if **your spouse was around because they're not working that day.**

0 hours	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours	7 hours	8 hours

Vignette 2.
FOR MEN: Imagine that one of your male neighbours has the opportunity to attend a similar event of great interest to them. They have many daily chores to accomplish. Their wife is away for the day. They cannot bring children to the meeting but find a responsible caretaker. How many hours do you think they should attend?

FOR WOMEN: Imagine that one of your female neighbours has the opportunity to attend a similar event of great interest to them. They have many daily chores to accomplish. They cannot bring children to the meeting but find a responsible caretaker. Their husband is away for the day. How many hours do you think they should attend?

0 hours	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours	7 hours	8 hours

Vignette 3. Your spouse is invited to a workshop of great interest to them during a typical workday. They could attend the workshop but may not accomplish all their daily chores. You'll need to mind the children while they're away. How many hours do you think they should attend?

0 hours	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours	7 hours	8 hours

Attitudinal questions were left largely unchanged from their formulation in the cognitive survey and are presented in full in Appendix 1.

The goal here is to understand whether aspects of time use agency related to voice may present additional barriers to labour force participation, especially women's (proxied by attendance at an income-improving seminar), even though we observe both spouses' answers. How each treatment group tested specific aspects of voice is presented below.

1. **Private treatment group:** In this treatment, spouses were asked to respond to the vignettes separately, and they were assured that their answers would not be shared with their spouse. Because spouses did not talk to each other and had the choice to not reveal their answers to each other, this treatment represents no communication between spouses about attendance to the seminar. It presents the lowest social pressure to adjust their answers according to what they think their spouse will want.
2. **Public treatment group:** In this treatment, spouses were asked to respond to the vignettes separately, but they were told that their answers would be revealed to their spouse later. The logic of this treatment is that while there is no communication, the possibility of social pressure exists because each respondent anticipates that their spouse will learn their answers.
3. **Negotiation treatment group:** In this treatment, spouses sat together and discussed their responses to the vignettes in real time. This scenario reflects the highest level of communication and adherence to the household's social norms.

Our first hypothesis is that in people who demonstrate high levels of time use agency, particularly related to voice, we expect to observe no significant differences between treatment groups. Note that because each person is assigned to only one experimental condition, we do not observe the same individuals under different experimental scenarios. Therefore, our results only tell us how different demographic groups respond in aggregate to these conditions.

The next hypothesis is that the negotiation treatment group will experience the highest level of communication (and is likely the closest to real-life conditions). Spousal negotiation could result in (a) a higher number of hours attended, particularly if spouses encourage each other to attend the seminar if the value of the potential income from the seminar is greater than the opportunity cost of not attending the seminar, or (b) a lower number of hours attended if home time is valued more than the seminar.

We hypothesise that voice in time use agency may vary by demographic characteristics like gender, age and family structure. If this is not true, again we expect to see no difference across demographic characteristics. This approach allows us to measure agency between treatment groups. The attitudinal questions help us identify specific individual agency attitudes, thus allowing us to observe agency differences in aggregate and at individual level.

This type of experiment can also help us understand other agency-based factors that determine participation (or lack of participation) in the labour market. Policies that simply focus on providing care so that people can enter the labour market are a necessary first step,

but there are other, less observable barriers to labour market participation that we need to attend to as well.

2.3. Sample and Research Sites

Our sample for the pilot phase included 452 individuals from Greater Jakarta and Greater Surabaya, namely, 226 men and 226 women. We intentionally sampled couples in order to be able to implement the time use survey as well as the attitudinal questions and the agency experiment with the same respondents.

2.3.1. Challenges in Implementing Time Use Surveys

Time use surveys have additional design requirements over and above those imposed by standard household sample surveys. These additional requirements have implications for the timing and distribution of survey field operations because, in addition to generating a probabilistic sample of persons, the sample design must often generate a probability sample of days of the week. Failure to randomise days of the week may bias estimates of population-level time use. In practice, a probability sample of days of the week is achieved by the randomised pre-assignment of each sample unit to one or more designated diary days.[12] In the case of retrospective 'yesterday' diaries, the random assignment of respondents to designated diary days directly conditions the survey participation day (i.e. the day immediately following the diary day). Respondents assigned to report on Monday's time use must be surveyed on Tuesday, those assigned to report on Tuesday's time use must be surveyed on Wednesday, and so on.

While it is relatively straightforward to extend a household survey sample design to obtain a probability sample of days of the week (often supported by adjusted sample weights), the designation of a specific diary day presents challenges for survey operations. Upholding the design increases the time and effort required to obtain a complete response, since a proportion of the sampled individuals will be unavailable, unable or unwilling to participate in the survey on their assigned day.

Sometimes in household sample surveys, proxy reporting (whereby a household member provides information on behalf of other eligible household members) is permitted in order to reduce the number of contact attempts required to obtain a response. The use of proxy reporting is discouraged in existing international guidelines on time use measurement, as the risks of information loss and inaccuracy are substantial (see UN Department of Economic and Social Affairs, 2005).

Taken together, the twin requirements of pre-assignment of reporting days and direct reporting present challenges when it comes to obtaining a sufficiently high response rate,

[12]The design utilised in the present study involves the pre-designation of a single diary day. Schemes with multiple designated diary days per respondent (e.g. one weekday and one weekend day) may be preferred when self-administered survey modes (e.g. mail-out, leave behind, online or app-based diaries) or mixed-mode (e.g. interviewer-administered for the first diary day and self-administered for the second diary day) permit prospective reporting.

which, in turn, presents risks for data quality due to nonresponse bias.[13] In a modular design, this may risk undermining the response rate for both the parent survey and the time use module.

Various postponement and/or substitution strategies have been proposed to reduce the challenges imposed by the designated diary day feature. Some strategies retain the probabilistic sampling of days of the week in modified form, while others relax or even dispense with it completely. The different strategies seek to find a balance between retaining a probabilistic mechanism for diary day assignment (at least initially) and minimising the extra burden, duration and cost of field operations imposed by a strict designated day scheme while maximising the response rate. Different strategies involve trade-offs in exposure to selection bias, measurement error, and the complexity and costs of field operations (Alonso et al., 2019). Our strategies to deal with these issues are presented in the next section.

2.3.2. Time Use and Agency Sampling

The study selected two urban areas as research sites: Greater Jakarta and Greater Surabaya. These areas were selected because the current study also serves as a follow-up to Lembaga Demografi's study on social norms and women's economic participation (Setyonaluri et al., 2021). Our pilot study focuses on urban areas because urban women have a more distinctive, M-shaped[14] FLFP compared to women in rural areas. Focusing on cities, particularly large metropolitan areas, can provide an understanding of the time allocation and intra-household negotiation that affect women's decision to work.

Greater Jakarta has a population of more than 10.56 million (BPS, 2021) and is one of Indonesia's most densely populated areas. As a growth centre, DKI Jakarta is attractive for people living nearby who come to the city to find work and access service facilities. As a result, the people of Jakarta have more diverse and pluralistic characteristics. Similarly, Greater Surabaya, the second largest metropolitan area in Indonesia, is home to 9.6 million people (BPS, 2021). Surabaya is a main destination for migrants from East Java and eastern parts of Indonesia who seek employment and education. These characteristics make the two urban centres appropriate sites for a TUS that will sample a variety of respondents and data. The sample area and sub-regions are shown in Table 2.6.

In each research site, we randomly selected three cities, and within each city we randomly selected two *kecamatan* or sub-districts. In each *kecamatan*, two *kelurahan* or villages were selected to represent middle-upper-class and middle-low-class areas. The selection of *kelurahan* was based on average household expenditure, distance to the city's main road, distribution of amenities (e.g., minimarkets, public transport, government offices), and distribution of household composition at *kelurahan* level, estimated using the 2021 National

[13]Earlier research found that response probability for time use surveys varies systematically with individual demographics, impacting time use estimates (Abraham et al., 2006; Abraham et al., 2009; Ingen et al., 2009; Fricker & Tourangeau, 2010).

[14]M-shaped FLFP suggests that women's participation tends to increase substantially at around 20–24 years of age. It dips slightly at 25–29 during their reproductive age, increasing again when children enter school, and levels off thereafter.

Socioeconomic Survey (SUSENAS). In each *kelurahan*, one *rukun tetangga* (RT) or neighbourhood was selected as the smallest enumeration area. To reduce the cost of travel for the enumerators, the distance between RTs became the main consideration during fieldwork.

Table 2.6. Sample regions and subregions for pilot activities

Jakarta		
City	**Subdistrict**	**Village/Kelurahan**
Jakarta Selatan	Tebet	Menteng Dalam
		Tebet Barat
	Jagakarsa	Tanjung Barat
		Lenteng Agung
Jakarta Timur	Pulo Gadung	Rawamangun
		Pulo Gadung
	Jatinegara	Cipinang Muara
		Kampung Melayu
Jakarta Utara	Penjaringan	Pluit
		Kapuk Muara
	Koja	Tugu Utara
		Tugu Selatan

Greater Surabaya (*Gerbang Kertasusila)*		
City	**Subdistrict**	**Village/Kelurahan**
Surabaya	Wonokromo	Darmo
		Wonokromo
	Bubutan	Gundik
		Tembok Dukuh
Sidoarjo	Waru	Pepelagi
		Tambak Rejo
	Prambon	Prambon
		Bulang
Mojokerto	Magersari	Gedongan
		Magersari
	Prajurit Kulon	Prajurit Kulon
		Blooto

In the pilot study, households were assigned a diary day (or interview day) randomly chosen from the seven days of the week. Interviews were conducted on every day of the week, with respondents reporting on their activities for the prior (pre-allocated) day.

As mentioned earlier, the designation of a specific diary day presents challenges to obtain a complete response, as some respondents may be unavailable or unable/unwilling to participate on their assigned day. In order to maintain an adequate response rate and minimise the duration and cost of field operations, our survey team put in place a household substitution scheme. The steps taken to select and substitute households are described below.

In each *kelurahan*, a field coordinator was responsible for sampling the households in each RT. Field coordinators, along with the enumerators, obtained the overall population list of the households living in the RT from local authorities one or two days before the interview was to be conducted. Whenever the local authorities did not have a complete list of

households in the RT, the field team performed door-to-door listing of households within the enumeration areas.

From the list, 6–10 households were selected probabilistically as 'main respondents', and the rest were selected (also probabilistically) as substitutions in case the main respondents refused or were unavailable for the interview. Since the pilot survey adopted the balance sample scheme, 10 households were interviewed each day during the first week of fieldwork, while 6–7 households were interviewed every day in the second week.

During the first visit (usually in the morning), enumerators asked respondents to agree to be involved in the study. If the respondents agreed and both husband and wife were ready to be interviewed, then enumerators proceeded with the interview immediately. When a respondent was not available for an interview during the day, the enumerator rescheduled it for later on the same day or after hours. If the respondent was not available throughout the day or refused to participate in the survey, the enumerator substituted the household with one from the reserve list.

As a result of the substitution strategy, we managed to obtain a balanced sample of diary days. In total, there were 32 or 33 households (64 or 66 respondents) interviewed for each day of the week. Both members of the selected couple-dyad (where relevant) were interviewed on the same day and reported their activities for the prior day. The distribution of respondents for each day and enumeration area in the work plan is illustrated in Table 2.7. In total, the pilot survey collected data from 452 respondents (226 couples) in Greater Jakarta and Greater Surabaya.

Table 2.7. Daily distribution of respondents during fieldwork period

	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu
Enumeration area	1	2	3	4	5	6	7	8	9	9/10	10	10/11	11	12
Week 1	40	40	40	40	40	40	40							
Week 2								26	26	24	24	24	24	24

The agency experiment was conducted on all 226 couples (or households). We randomly pre-assigned couples to a treatment group – private, public or negotiation – as explained in Section 2.3. As a result, 150 participants (75 households) fell into each of the private and public categories, and 152 participants (76 households) were in the negotiation category. Contrary to the cognitive testing implementation, the experiment was conducted at the end of the survey, that is, after household and time use modules and attitudinal agency questions had been administered. The reason for implementing the experiment last was to ease

logistics/movement, since for those in the negotiation group, participants must gather in a room to discuss answers, while the other part of the survey can be completed by each respondent alone.

Each interview lasted 35–75 minutes on average. The time use module lasted 15–30 minutes, and it was longer for households with many members. The time use module took 10–25 minutes and was typically longer for mothers with young children due to the varied activities conducted in a day. The agency module, both attitudinal and experimental, took 10–20 minutes.

Interviews were longer, on average, for respondents who multitasked while being interviewed. On the other hand, interviews were shorter for respondents who were in a hurry to get to work or open their business, and those who were interviewed at night. There was no significant difference in mean interview duration across the two study sites.

Sixteen enumerators were employed in this study. The profiles of selected enumerators are presented in Table 2.8. We paid attention to gender in the recruitment to ensure that respondents not only felt comfortable in answering the questions but also understood the questions. First, all hired enumerators were locals comfortable communicating in the local language. Second, we assigned male enumerators to interview male respondents and female enumerators for female respondents. Third, we conducted role-playing simulations during the training and conducted strict monitoring of enumerators during the interview process to ensure that questions were understood as intended by both enumerators and respondents.

Table 2.8. Enumerator profile

Variable	N
Panel A. Demographic frequencies	
Total enumerators	16
Male	7
Female	9
Attained diploma 1/2/3	1
Attained diploma 4/bachelor's	15
Panel B. Age and experience (years)	**Mean**
Average age	30.063
Average enumeration experience	5.625

2.4. Data Analysis Strategy

Our analysis utilised three separate types of data collected with a single survey instrument:

- time use (modified LFS module)
- attitudes about agency
- experimental agency outcomes.

2.4.1. Data Visualisation

The first phase of the analysis provides a description of the attitudinal and quantitative variables collected. To implement the descriptive analysis we provide means, standard deviations and other relevant comparisons, as well as visualisations where appropriate.

Time use variables. The first approach to describing time use is to compute the average number of minutes people spend on each generalisable activity. This data can be visualised in bar graphs.

One of the main objectives of this data collection effort was to better understand multitasking and supervisory care duties. We present the average number of hours women spent multitasking. We also present a description and visualisation of the kinds of activities done while multitasking.

We provide detailed descriptions of variables measuring supervisory care, including what type of labour was being performed while supervising children or other family members.

Agency. The survey instrument contains two different measures of agency. The first consists of a series of attitudinal questions (strongly disagree to strongly agree) and the second is an experiment. The questions are grouped around four different subcategories of agency: critical consciousness, self-efficacy, decision-making, and voice (instrumental). Within each of these categories we present a descriptive summary of the response to each individual question using a stacked bar graph. The graph shows the percentage of each response to each question.

In addition to the individual analysis by question, we also aggregate scores within the four agency types to obtain an aggregate score, with a higher score typically indicating a higher degree of agency in a certain category. We first explain the variation in these scores by demographic characteristics and then use them to analyse differences in time allocations among respondents.

Experimental outcomes. The objective of the experimental design was to test whether different modes of communication (represented by four vignettes) change the self-reported time allocation to a training program. Given that there were three different treatment groups, we compare average responses to each vignette across treatments. We also present a heterogeneity analysis by gender and family composition, particularly examining how households with young children responded differently to the vignettes depending on their treatment assignment. We present the means and standard deviations, and statistically test whether the means differ between each paired grouping using a t-test. The t-statistic and resulting p-value are reported.

2.4.2. Multivariate Analysis

We employ a multivariate regression to examine how different levels of agency affect time use outcomes. First, we use a summary of the Likert scales to explain variations in time use and hours reported, including domestic work, paid work, supervisory care and multitasking. This multivariate analysis allows us to control for rural/urban status, age and other demographic characteristics. The regression framework is as follows:

$$TimeUse_i = \beta_0 + \sum_{z=1}^{5} CategoryScore\ \alpha_z\ + \beta_2 Education_i\ + \beta_3 Age_i + \beta_4 Disability_i\ +\ \beta_5\ HHSize_i + \epsilon_i \quad (1)$$

$CategoryScore$ is the Likert measure (z = critical consciousness, self-efficacy, decision-making, and voice) aggregated within each of the four agency categories. We also use the experimental data to examine whether each treatment arm has a significant impact on the number of hours that people report allocating to different vignette activities.

In the first regression we are interested in estimates for α_z. These parameter estimates will help us understand the direction and magnitude of different measures of agency over time use decisions within the household, while controlling for the respondent's education (β_2), age (β_3), disability status (β_4), and number of household members (β_5). Further, including supervisory care and multitasking will uncover the relationship between agency measures and these time uses.

We also examine the factors that are associated with labour force participation for both men and women. Using a dichotomous variable that equals 1 if a person participated in any paid work, and 0 otherwise, we build a model to determine if and what kinds of paid work are correlated with entrance into the labour force. We employ the following regression model:

$$LFP_i = \beta_0 + \beta_1 Supervisory_i + \beta_2 UnpaidCare_i + \beta_3 UnpaidDomestic_i + OtherTimeUse\gamma + X\alpha + \epsilon_i \quad (2)$$

where LFP is a dichotomous variable indicating if person i participated in paid work, $Supervisory_i$, $UnpaidCare_i$, and $UnpaidDomestic_i$ are the number of hours person i spent on supervisory care, unpaid care, and unpaid domestic work apart from care respectively. $OtherTimeUse$ is a vector is a vector of other time uses including leisure, community and religious activities, volunteer work, and travel. The vector X is a set of control variables including household size, age of the individual, education level, and if the household has children under 5 years old. We use both a probit model and a linear probability model to estimate equation 2.

We examine the responses to the vignettes in the following form:

$$VignetteTime_i = \beta_0 + \beta_1 Public_i\ + \beta_2 Negotiation_i\ + Demographics_i\alpha + \epsilon_i \quad (3)$$

In equation 3, we are most interested in testing whether $\beta_1 or\ \beta_2$ are different than zero. Our null hypothesis is that the form of communication between spouses will not affect the length of time for which each spouse reports wanting to attend the training themselves, or how much they would like the other spouse to attend (vignette 3). In this case, we would expect both $\beta_1 and\ \beta_2$ to be equal to zero for both spouses.

One concern here is that our analysis may be underpowered due to a low sample size. However, even with a small sample we may get suggestive evidence of an experimental treatment effect.

In addition to this analysis, we examine the difference between the number of hours for which a husband/wife reports wanting to attend the training (vignette 1) and the number of hours

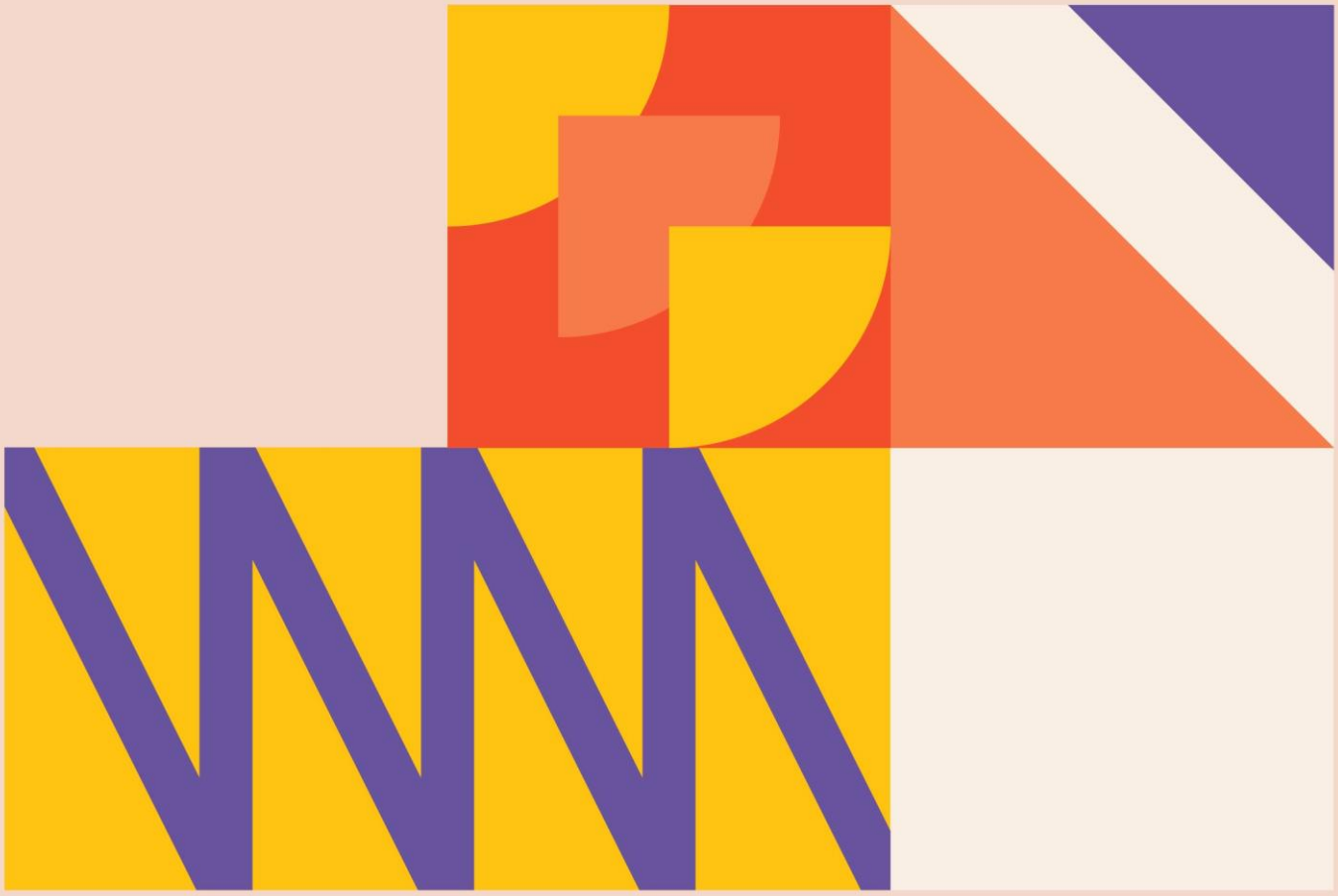
that their spouse wants *them* to attend (vignette 3). We estimate what factors explain this difference for both men and women. This relationship is examined in the following regression:

$$V1(spouse1) - V3(spouse2) = \beta_0 + \beta_1 Public_i + \beta_2 Negotiation_i + \beta_3 TimeUse_{s1} + Demographics_{s1}\alpha + \epsilon_i \quad (4)$$

In equation 4, the demographic and time use variables are related to spouse 1 in the difference definition. Again, if we believe that the form of communication does not matter, we expect $\beta_1 and\ \beta_2$ to be equal to zero for both spouses.

Section 3

Research Results



3. Pilot Study Research Results

This section describes the findings from the pilot activities in which we deployed the refined survey instruments on a sample of 452 individuals, as discussed in Section 2. We include data visualisations about time use patterns by demographic group, descriptive statistics about time use patterns, a discussion of simultaneous activities with a focus on supervisory care, an analysis of the experimental activities to detect changes in preferences depending on spousal involvement, and an analysis of the attitudinal questions. We also provide a multivariate analysis of experimental results explained across demographic characteristics, and of time use choices explained by attitudinal questions.

3.1. Time-Use Survey: Descriptive Results and Analysis

The demographic and social characteristics of the sample of respondents are presented in Table 3.1. More than 70% of respondents in the pilot study are aged 40 and older (rows 1.d to 1.h), while only 5.5% are aged 29 or younger. The majority of respondents are graduates of upper-secondary education (40.49%), with male respondents in Jakarta having a higher percentage of upper secondary education compared to female respondents (row 2.d). In Surabaya, the educational level of male and female respondents is fairly balanced. Regarding employment, female respondents have a lower share of workers (either for pay, for profit or in a family business) compared to their male counterparts (rows 3a to 3c). However, among those who are employed, the share of female respondents owning a business is higher compared to male respondents in both Jakarta and Surabaya (rows 4a to 4b).

Table 3.1. Selected respondent characteristics

		Jakarta		Surabaya		Total (%)
		Male (%)	Female (%)	Male (%)	Female (%)	
1	***Age group***					
1.a	under 29	3.5	11.5	2.7	4.4	5.5
1.b	30–34	8.9	13.3	3.5	6.2	8.0
1.c	35–39	15.0	13.3	10.6	9.7	12.2
1.d	40–44	16.8	15.0	8.9	18.6	14.8
1.e	45–49	17.7	21.2	17.7	19.5	19.0
1.f	50–54	17.7	15.0	20.4	13.3	16.6
1.g	55–59	10.6	8.0	10.6	12.4	10.4
1.h	60+	9.7	2.7	25.7	15.9	13.5
	N	113	113	113	113	452
2	***Education***					
2.a	Less than primary	10.6	14.2	3.5	3.5	8.0
2.b	Primary	18.6	29.2	15.9	19.5	20.8
2.c	Lower secondary	23.9	24.8	15.0	14.2	19.5
2.d	Upper secondary	38.1	24.8	51.3	47.8	40.5
2.e	Post-secondary/tertiary	8.8	7.1	14.2	15.0	11.3
	N	113	113	113	113	452
3	***Employment status***					
3.a	Works for pay	76.1	46.0	62.8	21.2	51.6
3.b	Works for profit	7.1	12.4	12.4	38.9	17.7
3.c	Works in family business	6.2	6.2	8.9	8.0	7.3
3.d	Not employed	10.6	35.4	15.9	31.9	23.5
	N	113	113	113	113	452
4	***Employment relationship in main job***					
4.a	Employee	62.75	39.73	43.75	24.68	44.25
4.b	Owns the business	34.31	46.58	42.71	62.34	45.4
4.c	Helping family/household business	2.94	13.7	12.5	12.99	10.06
4.d	Helping family work for someone else	0	0	1.04	0	0.29
	N	102	73	96	77	348

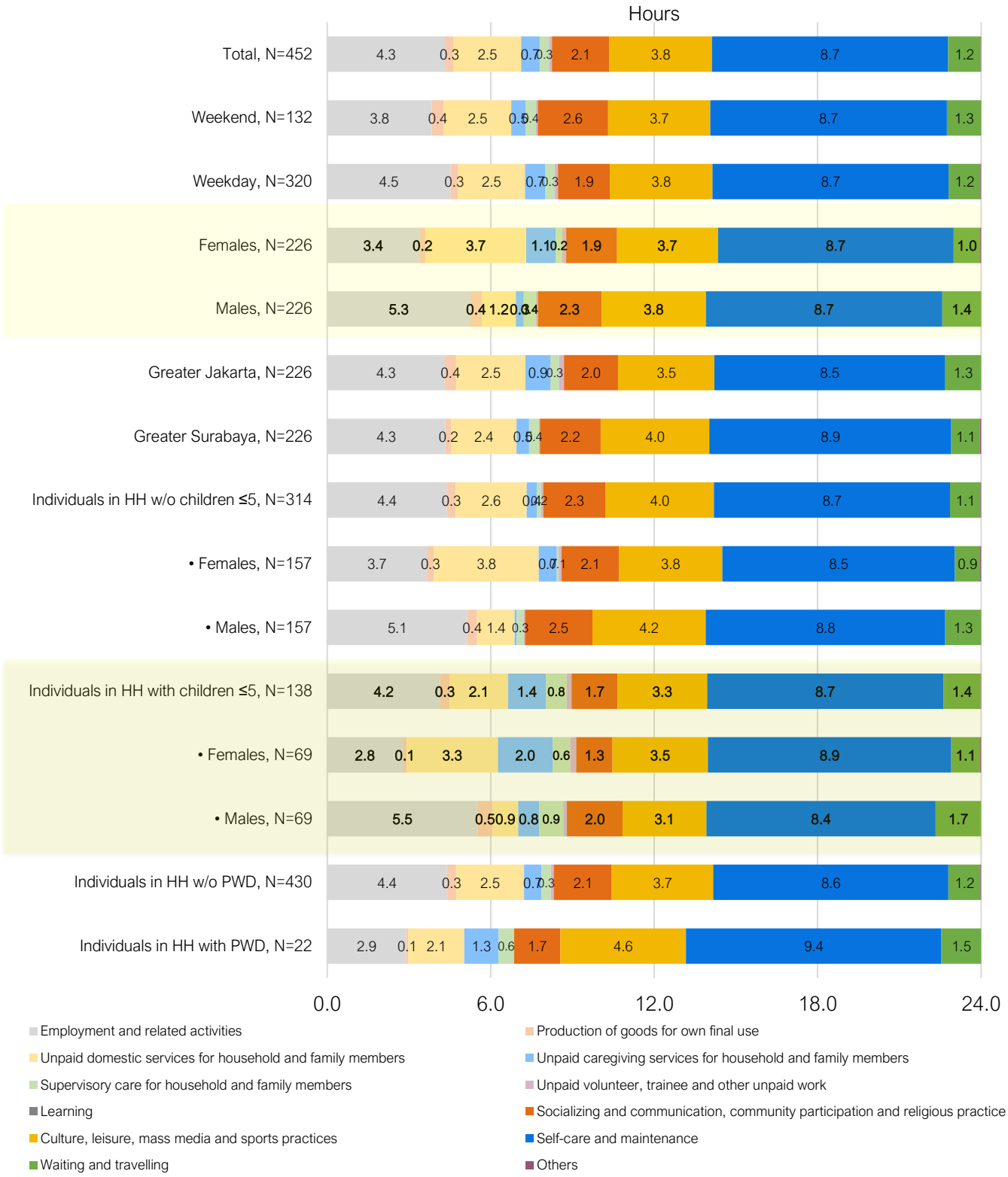
The following graphs illustrate how people allocate their time throughout the day. We investigate both primary (the first activity spontaneously reported) as well as secondary (simultaneous activities reported in addition to the first specified activity) activities in order to understand the role of multitasking by household members throughout the day, and to better capture care obligations that are often reported as secondary rather than primary activities.

The averages presented relate to 'social time' indicators, that is, they are calculated inclusive of participants and non-participants in the activity domain. The alternative volume measure, 'participant time', includes only time for those who engaged in each activity.

3.1.1. Time spent on primary activity.

Figure 3.1 illustrates the average number of hours that respondents spend on each primary activity, broken down by demographic group. There are several notable patterns.

Figure 3.1. Average number of hours excluding simultaneous activities by type of activity and population group



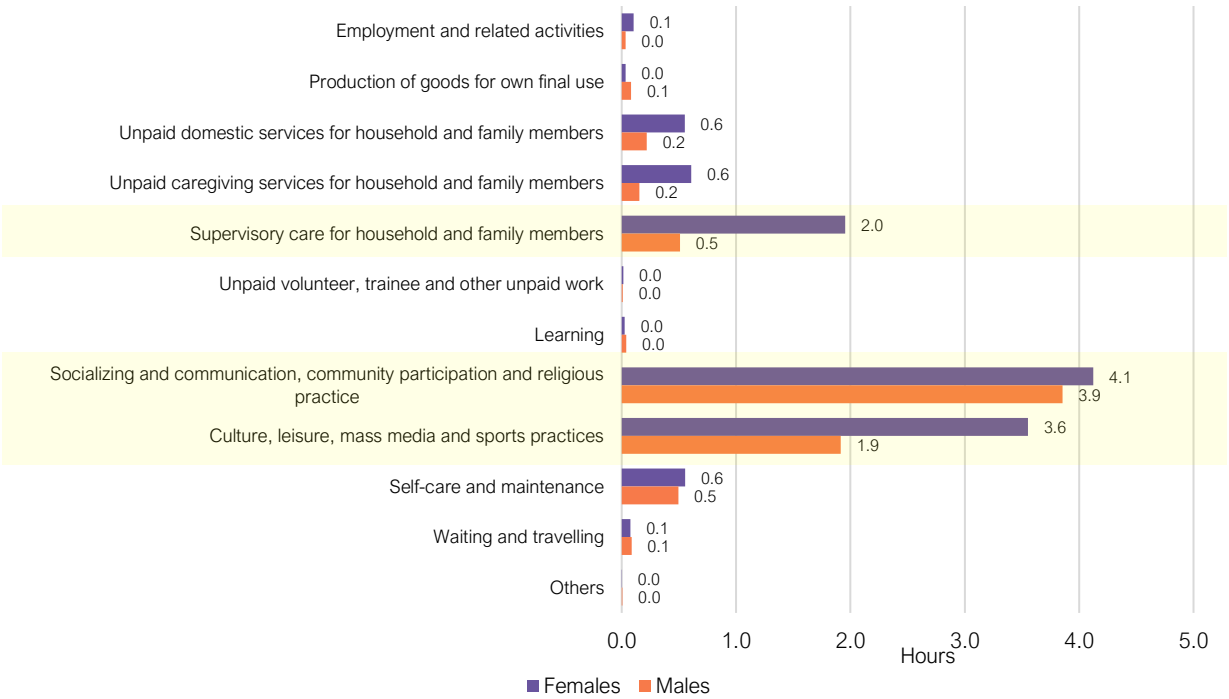
On average, respondents spend the most time on 'self-care and maintenance', which includes sleep. There does not appear to be any difference by gender, with both women and men (irrespective of the type of household they live in) spending between 8.7 and 8.9 hours on this activity. Compared to their male counterparts, mothers of children below the age of 5 spend slightly less time on socialising but slightly more on self-care and maintenance.

In comparison, the amount of reported time spent on paid and unpaid care and domestic work varies significantly by gender and age of the child. First, men spend an average of 5.3 hours per day on paid work, whereas women spend an average of 3.4 hours per day. This difference is larger in households with children under 5, where men work 5.5 hours per day for pay and women just 2.8 hours.

Conversely, we see women carrying more of the unpaid domestic work burden, with an average of 3.7 hours per day for women and 1.2 for men. Again, this difference is larger in families with children under 5, at 3.3 hours per day for women and 0.9 for men.

We also see evidence that women spend more time on unpaid active and supervisory care work (1.3 hours) compared to men (0.7 hours) when considering only their primary activity (this is higher when simultaneous activities are included – see below). This is consistent in subsamples of households with and without children below 5 years of age.

Figure 3.2. Average number of hours of simultaneous activities by type of activity



Importantly, Figure 3.1 does not include simultaneous activities, which often comprise a significant portion of unpaid work. Figure 3.2 shows the average number of hours (which is between 0 and 4.1) of simultaneous activities broken out by gender.

We see that ‘socialising and communication, community participation and religious practice’ is the most common simultaneous activity, with women reporting 4.1 daily hours and men reporting 3.6 hours. ‘Culture, leisure, mass media and sports’ is also a common grouping of simultaneous activities. Also, relevant to our programmatic objective of understanding unpaid work, women spend an average of 2 hours per day on supervisory care, whereas men spend only 0.5 hours per day on the same, when reported as occurring simultaneously with other activities.

3.1.2. Multitasking and Supervisory Care

Figure 3.3 shows the average number of hours spent on each type of activity based on the aggregation of primary and simultaneous activities. On average, respondents spent the longest time (9.2 hours) on self-care and maintenance, which includes sleeping, followed by culture and leisure (6.5 hours), and socialising and communication (6.1 hours).

Men spent more time on paid employment activities compared to women (5.3 vs 3.5 hours). On the other hand, the time allocated to unpaid domestic work by women (4.2 hours) was almost three times as much as for men (1.5 hours). Women engage in unpaid active caregiving services for 1.7 hours per day on average, while men only spend 0.4 hours on the same.

Supervisory care jumps significantly, particularly for women with young children, once we consider simultaneous activities. Women (2.2 hours) spent more than twice as long as men (1 hour) on supervisory care. Mothers in households with child(ren) aged 5 or below spent 5.2 hours on supervisory care, compared with only 2.1 hours for their male counterparts.

When summing time spent on all types of work, women worked longer hours (11.8 hours) than men (8.2 hours). This suggests that women typically shoulder the largest portion of domestic work, and the gender gap in the amount of time allocated for domestic work appears to be pervasive. For women with children aged 5 or below, however, unpaid care (supervisory and active) is added on top of their heavy domestic load, such that women in this group spend 12.3 hours per day doing unpaid domestic and care work, which is almost three times longer than their male counterparts, who do only 4.3 hours, and twice as long as women without children (6.3 hours).

Figure 3.3. Average number of hours including simultaneous activities by type of activity and population group

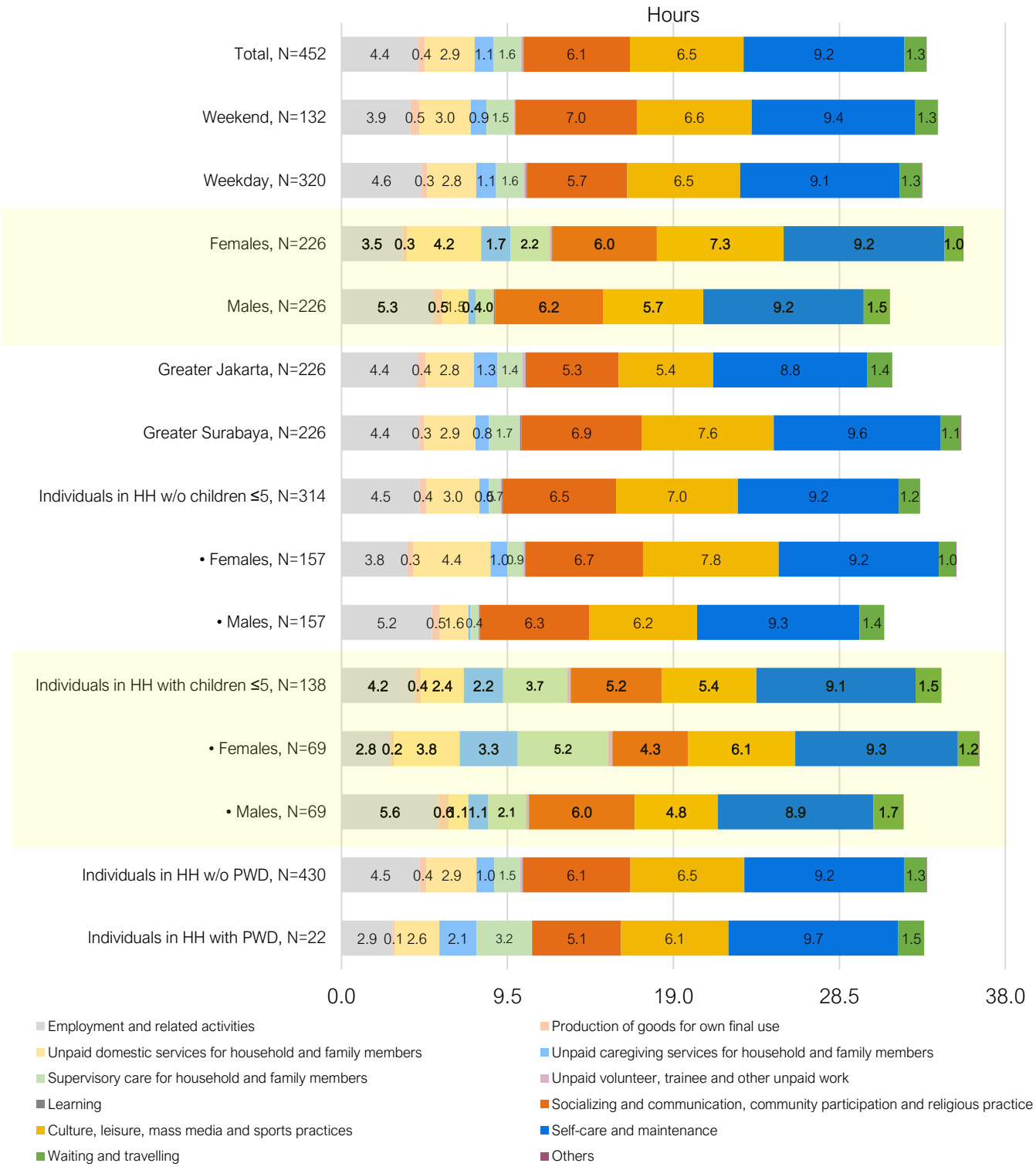


Table 3.2. Mean hours by gender and type of activity

		Mean (men)	Mean (women)	t-test			
				Diff.	Std. err.	t-stat	p-value
1.	**Panel A. All Sample (N = 452)**						
1.a	Employment and related activities	5.292	3.49	-1.802	.388	-4.65	0
1.b	Unpaid domestic and care work	2.865	8.126	5.261	.438	12	0
1.c	Total work (paid work, unpaid domestic work, unpaid care work, and volunteer/trainee work)	8.238	11.767	3.529	.508	6.95	0
1.d	Socialising and communication, community participation and religious practice	6.192	5.991	-.2	.431	-.45	.643
1.e	Culture, leisure, mass media and sports practices	5.741	7.258	1.516	.41	3.7	0
1.f	Self-care and maintenance	9.177	9.209	.032	.231	.15	.89
2.	**Panel B. Subsample: Individuals with child(ren) ≤ 5 in the household (N = 138)**						
2.a	Employment and related activities	5.562	2.819	2.743	.652	4.2	0
2.b	Unpaid domestic and care work	4.42	12.286	-7.866	.829	-9.5	0
2.c	Total work (paid work, unpaid domestic work, unpaid care work, and volunteer/trainee work)	10.14	15.341	5.203	.93	5.6	0
2.d	Socialising and communication, community participation and religious practice	4.753	6.131	-1.377	.614	-2.25	.026
2.e	Culture, leisure, mass media and sports practices	8.906	9.305	-.399	.4	-1	.32
2.f	Self-care and maintenance	2.148	5.181	-3.033	.586	-5.2	0

Table 3.2 statistically tests the differences in means between men and women across different time use categories. Results are largely consistent with expectations, with men engaged significantly more in paid work (row 1a), while women report a higher average number of hours of unpaid domestic and care work (row 1b), total work (row 1c), and culture and leisure activities (row 1e). For households with small children, the paid and unpaid work patterns remain, but women and men are at parity in terms of culture and leisure activities, and women are engaged in more self-care and maintenance.

Figure 3.4 illustrates the percentage of each group that performs supervisory care, as reported in the recovery sequence of the pilot light diary. Individuals with young children in the household have the highest share of those providing supervisory care compared to other demographic groups, and this is true for both male and female subsamples.

Figure 3.5 presents the number of supervisory care hours performed by each demographic group. The graph confirms that women with young children[15] report the highest number of supervisory care hours (5.2), while males with young children spent less than half as many hours (2.1) as their female counterparts. Given our results above, we see that women with young children are the most likely to engage in the highest number of hours of supervisory care.

In the time use section, respondents reported providing supervisory care for 1.6 hours on average. In the recovery section, the time spent on supervisory care was 2.2 hours (combining supervisory care for children and adults). From a methodological standpoint, this suggests that the recovery section of the pilot time use module – a section which returns to supervisory care specifically after the diary day activities have been fully reported as a consecutive series of activities – may improve the measurement of supervisory care.

[15] Women with young children (highlighted in blue in Figure 3.3) is a subset of carers of young children (labelled carer of children ≤ 5 in Figure 3.3).

Figure 3.4. Percentage of respondents providing supervisory care by population group

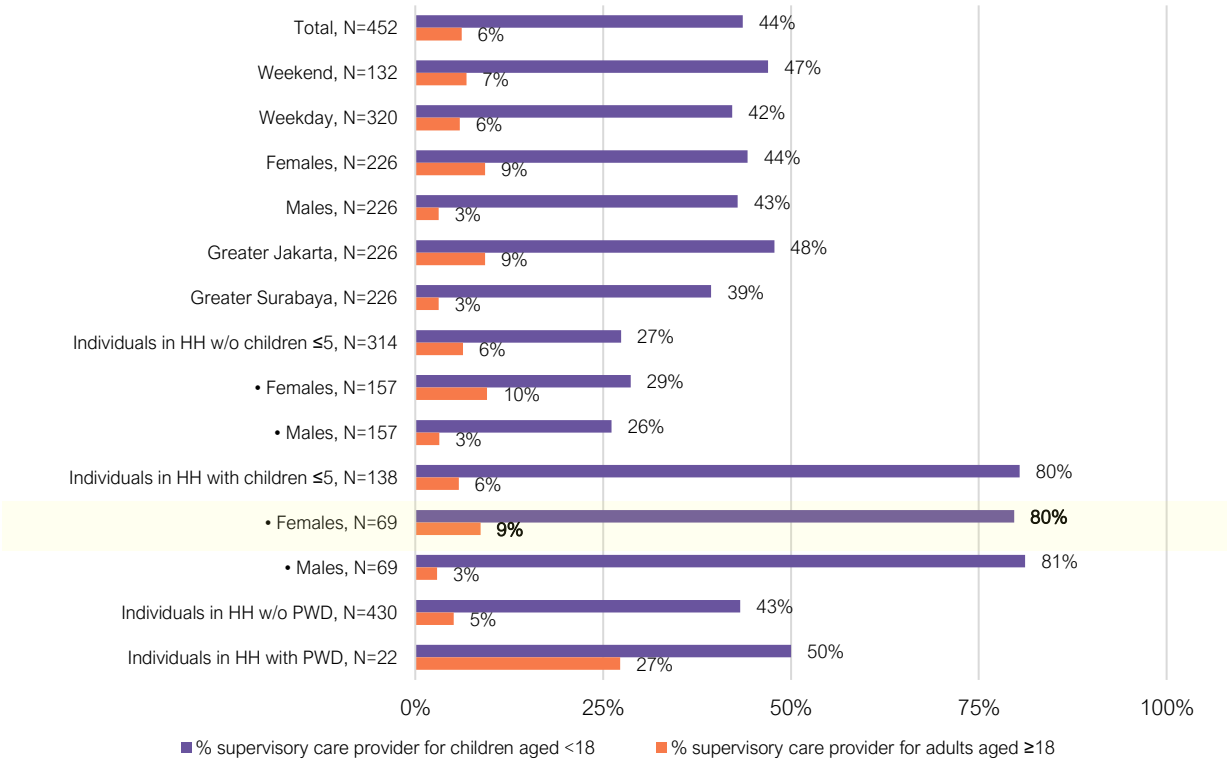


Figure 3.5. Average number of hours of supervisory care by population group

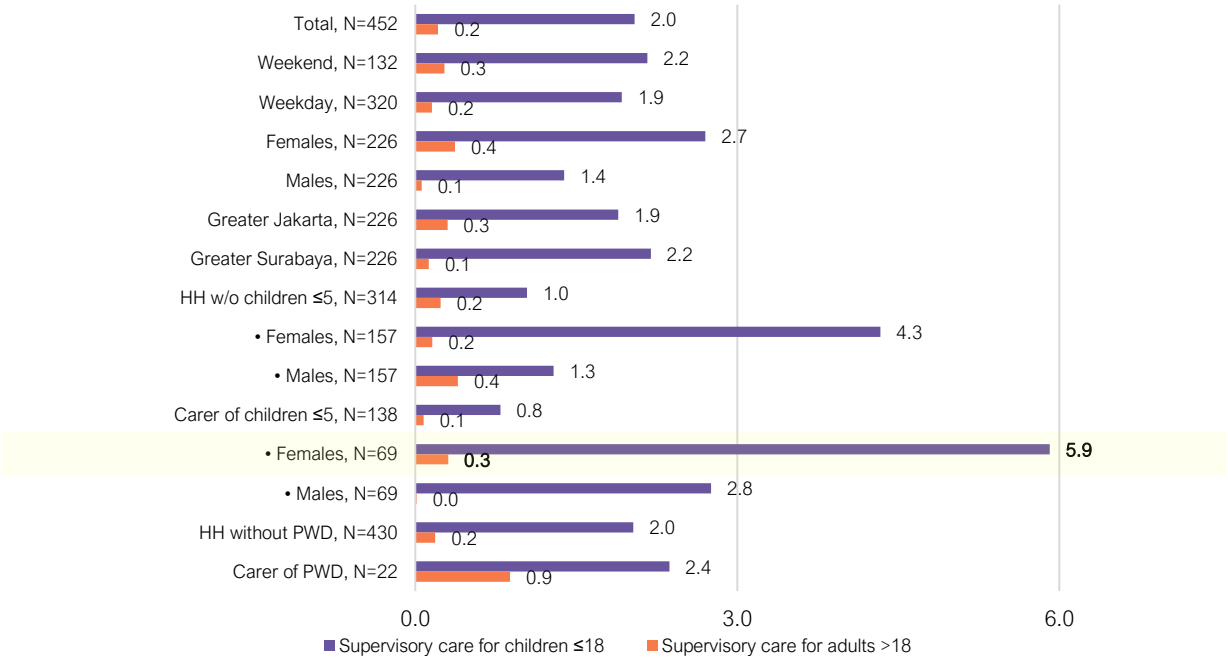


Table 3.3 is a frequency table showing the top activities conducted simultaneously with supervisory care. For those with young children, ‘resting, relaxing’, which includes doing

nothing, is the most common response. Although other forms of self-care, such as eating or drinking and watching TV, are also top responses, we see a considerable share of the population performing unpaid domestic work (i.e. cooking and cleaning indoors) while supervising their children. Similar results are also found for activities done simultaneously with supervisory care for adults aged over 18, in which self-care activities and unpaid domestic work are the top reported activities. These results may be indicative of the challenges of combining supervisory work with paid work.

Table 3.3. Top activities reported while providing supervisory care

	Count	% of sample
Panel A. Activities while doing supervisory care of children aged ≤18 (N = 197)		
39. Resting, relaxing (includes 'doing nothing')	61	30.96%
04. Cooking, baking, preparing or serving meals, snacks, beverages, and cleaning dishes / utensils / OR preserving, storing, arranging food stocks	53	26.90%
03. Eating or drinking	44	22.34%
09. Cleaning and tidying indoors, and disposal of household rubbish	36	18.27%
37. Watching TV shows, movies, online entertainment, listening to music / radio, other mass media	35	17.77%
14. Providing physical care for children (e.g. feeding, cleaning, medical care)	82	41.62%
Panel B. Activities while doing supervisory care for adults aged >18 (N=28)		
04. Cooking, baking, preparing or serving meals, snacks, beverages, and cleaning dishes / utensils / OR preserving, storing, arranging food stocks	9	32.14%
09. Cleaning and tidying indoors, and disposal of household rubbish	8	28.57%
37. Watching TV shows, movies, online entertainment, listening to music / radio, other mass media	6	21.43%
11. Laundry (washing, drying, ironing, folding, etc.) and repair / maintenance of clothes/textiles/shoes	5	17.86%
39. Resting, relaxing (includes 'doing nothing')	4	14.29%
14. Providing physical care for children (e.g. feeding, cleaning, medical care)	3	10.71%

We also find that some respondents reported engaging in active and supervisory care at the same time. It is possible for someone to actively care for someone else at the same time as they are passively caring for others, and/or that the fixed episode length of 15 minutes is insufficiently sensitive to capture genuine task switching, such as passive care interrupted by shorter active caregiving interactions (something noted in the wider literature as characterising supervisory care). An alternative and perhaps worrying explanation is that there is an issue of comprehension around the difference between active and supervisory care. At the level of statistical reporting, however, this issue is not of great concern, since estimates produced in line with international guidance allow that simultaneous forms of caregiving are counted only once.

Where the concern is to untangle the precise moment-by-moment nature of the caregiving relationship so as to fully parse out the nature and extent of task switching and simultaneity at a finer level than permitted by the 15-minute episode, the tendency to report simultaneous active care and passive care poses more intricate methodological challenges. For our

purposes, the focus on supervisory care is motivated by its known tendency to be underreported as a dimension of unpaid care work, which, as we have seen above, disproportionately falls to women. Based on further analysis indicating the existence of a comprehension gap, refinements to scripted prompts designed to aid respondents' comprehension of supervisory dimensions of care may resolve the issue.

3.1.3 Relationship between paid and unpaid work

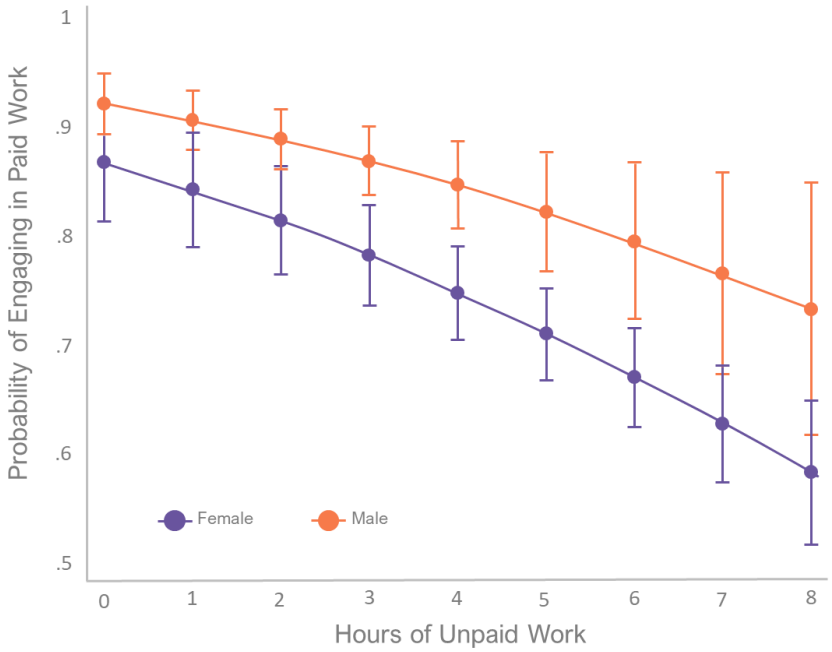
Further analysis[16] suggests there is a negative relationship between amount of unpaid work and the probability of employment, which applies for both men and women. This is depicted in Figure 3.6, mapping the probability of labour force participation for both men and women at different levels of unpaid work (0-8).[17]

On average, an increase in 1 hour of unpaid work decreases the probability of employment by about 3% for men and 4% for women. This implies the negative impact of unpaid work on employment is almost similar for both men and women, even as the baseline level of unpaid work differs greatly between men and women. Women typically have a higher baseline of unpaid work relative to men, with an average of 4.5 hours compared to 1.5 hours for men.

At low levels of unpaid work, the probability of engaging in paid work is similar for both genders (the "whiskers" in Figure 3.6 overlap), but as the amount of unpaid work increases, the probability of engaging in paid work drops more quickly for women than men, becoming significantly lower at about 3 hours of unpaid work (the "whiskers" do not overlap in Figure 3.6). We see that men and women do not differ in their probability of employment when engaging in 0, 1 or 2 hours of unpaid work. However, at 4 hours of unpaid work, the probability of engaging in unpaid work is roughly 78% for the average women and 84% for men.

[16] Using both a probit regression model and a linear probability model (Refer to 2.4.2).
[17] These results are based on a probit model explaining the probability of paid labour force participation (0/1) explained by number of hours of unpaid work. The model includes individual and household level controls.

Figure 3.6. Probability of engagement in paid work predicted by hours allocated towards unpaid work



It is important to note that these results demonstrate a correlation rather than a causation, and reverse causation may be at play as well, meaning individuals who are out of the paid labour market may have more time available to engage in unpaid work. Nevertheless, the overall study finds a consistent negative relationship between unpaid work and employment for both women and men across different hours of unpaid work, thereby suggesting that the direction of causality is likely from unpaid work to employment, rather than the other way around.

Disaggregating the analyses by type of unpaid work helps to shed light on the specific unpaid responsibilities most likely to impact women's engagement in paid work, with supervisory care has the largest negative effect on women's probability of engaging in paid work at 8.0% with each additional hour (row c, column 1). Each additional hour of supervisory care is also associated with men's lower probability of paid work, but they are less responsive at 4.4% reduction (row c, column 2). Unpaid domestic duties, also have a negative effect on the probability of engagement in paid work for women, decreasing the probability by 2.1% for each additional hour (row a, column 1). Meanwhile, unpaid domestic duties have no effect on men.

Our findings also suggest community and religious activities, leisure and travel and volunteering and other unpaid work have a more significant impact on women's probability of engaging in paid work than men's, as reflected in the Table 3.4 below. An additional hour spent on community and religious activities decreases women's probability of working by a substantial 7.6%, while for men, the decreases is only 4.8%. Similarly, an extra hour of total leisure and travel and waiting time decreases women's engagement in paid work by 6%, compared to a 2% decrease for men. In contrast, an additional hour of volunteering and other unpaid work increases women's probability of engaging in paid work by 5.6%, which is not

the case for men. This finding suggests that volunteering activities may help women gain the experience and skills needed to participate in paid work.

Table 3.4. Regression analysis explaining probability of engagement in paid work

	Variable	Female's Probability of Engaging in Paid Work (1)	Male's Probability of Engaging in Paid Work (2)
a	Unpaid domestic work	-0.021**	-0.017
		(0.010)	(0.013)
b	Unpaid active care	-0.007	-0.057
		(0.018)	(0.037)
c	Unpaid supervisory care	-0.080**	-0.044**
		(0.032)	(0.022)
d	Leisure total	-0.060***	-0.055***
		(0.009)	(0.008)
e	Community and Religious Activities	-0.076***	-0.048***
		(0.014)	(0.015)
f	Volunteering and other Unpaid	0.056**	0.018
		(0.027)	(0.030)
g	Travel and Waiting	-0.021**	0.008
		(0.011)	(0.013)
h	Number of household members	-0.001	0.012
		(0.021)	(0.012)
i	Age (when missing birth date)	0.000	-0.003
		(0.003)	(0.003)
j	Secondary Education	-0.075	-0.010
		(0.058)	(0.052)
k	Tertiary Education	0.139	0.036
		(0.103)	(0.078)
l	Household has children < 5yo	-0.004	-0.020
		(0.090)	(0.051)
m	Observations	226	226
n	R-squared	0.299	0.296

3.2. Attitudinal Agency Questions: Descriptive Results and Analysis

This section reports results from the attitudinal questions administered after the time use section of the questionnaire. Respondents were asked a series of grouped questions in order

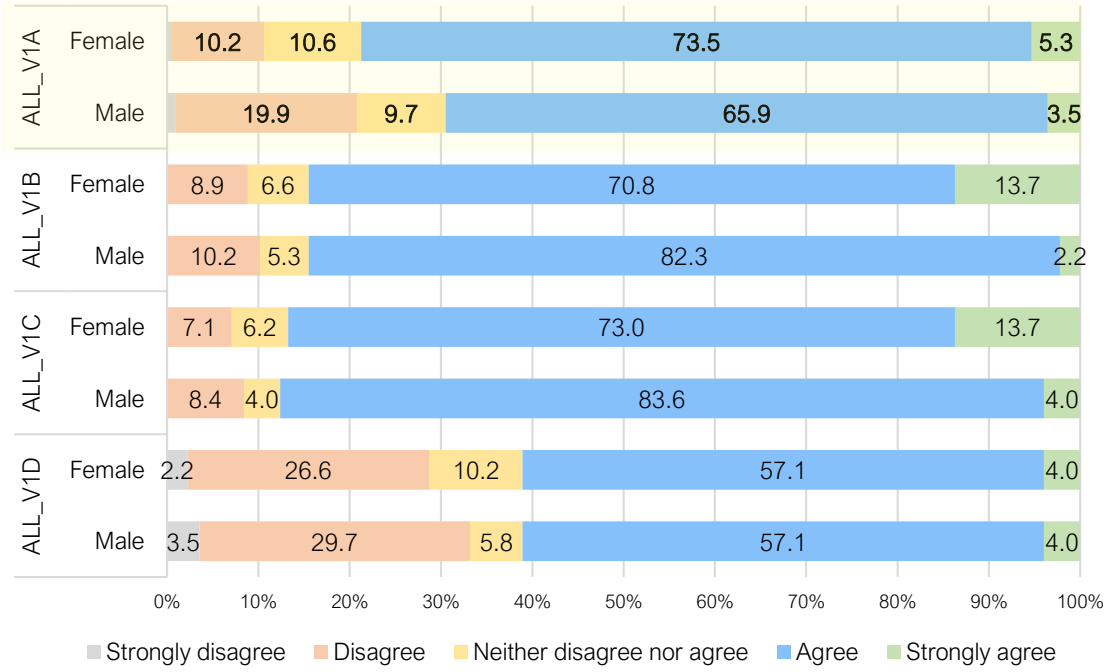
to gauge their level of different forms of agency: self-efficacy, decision-making, critical consciousness, and voice.

This section is laid out in four separate subsections. Section 3.2.1 provides a disaggregated summary of the Likert scale questions, reporting the percentage of respondents' answers for each individual question across the five-point scale. Section 3.2.2 aggregates responses to the attitudinal questions within agency categories and describes the aggregated scores. Section 3.2.3 provides a multivariate analysis to explain the variation in agency scores by educational level, age, household disability status and household size. Lastly, Section 3.2.4 explores the time use patterns explained by the attitudinal question scores.

3.2.1. Individual Attitudinal Question Frequency Visualisation

This section provides a visualisation of attitudinal question responses, grouped according to four different forms of agency. Figure 3.7 shows the responses to four questions measuring self-efficacy. The questions corresponding to each category are shown below the figure.

Figure 3.7. Self-efficacy: Tabulation of responses to four attitudinal questions

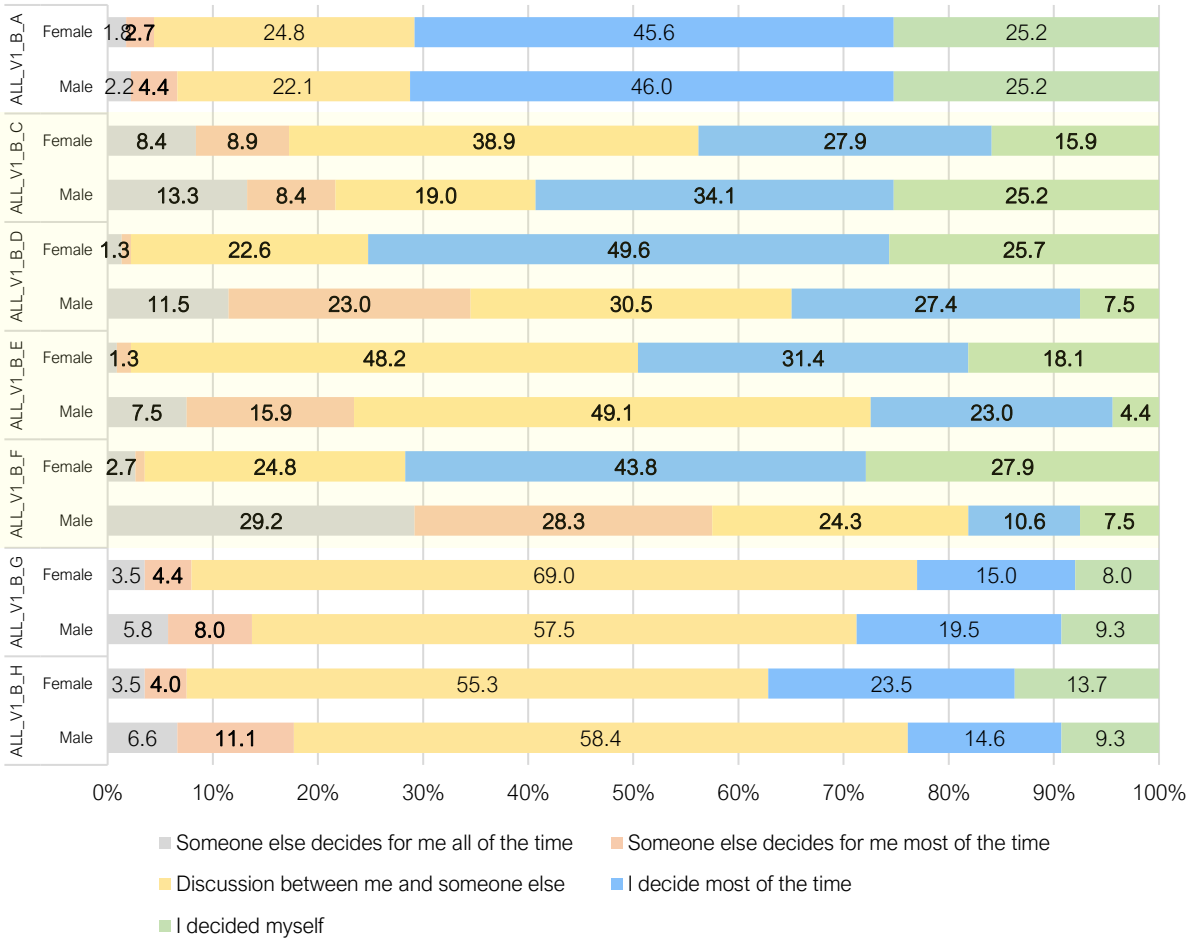


Question code	To what extent do you agree or disagree with the following statements?
ALL_V1A	a. You have the ability to change your daily schedule.
ALL_V1B	b. You can ask a household member to do some of your household duties.
ALL_V1C	c. You can ask a household member to help you take care of a child or other family member.
ALL_V1D	d. You can change the amount of time you spend on paid work.

Figure 3.7 shows that while a vast majority of both women and men agree with statements about their ability to exercise agency over their time use, men are more likely to agree or strongly agree that they can ask for help (ALL_V1C). The one exception is paid work, where men and women feel similar levels of constraint. The share of women who agree that they can change their daily schedule is higher than men's. However, men have a higher share of those agreeing that they can ask for help in performing household duties and childcare.

There are two interpretations for this finding. While women have agency over their own time use, they are less likely to reach out for help from others for domestic and caregiving activities, which are assumed to be women's responsibilities and best done by them, as indicated in the Social Norms Attitude and Practice (SNAP) Survey 2020 (YouGov & Investing in Women, 2020). Men are more able to ask for help with household duties and caregiving because they are not the ones primarily responsible for these tasks.

Figure 3.8. Decision-making: Tabulation of responses to seven attitudinal questions



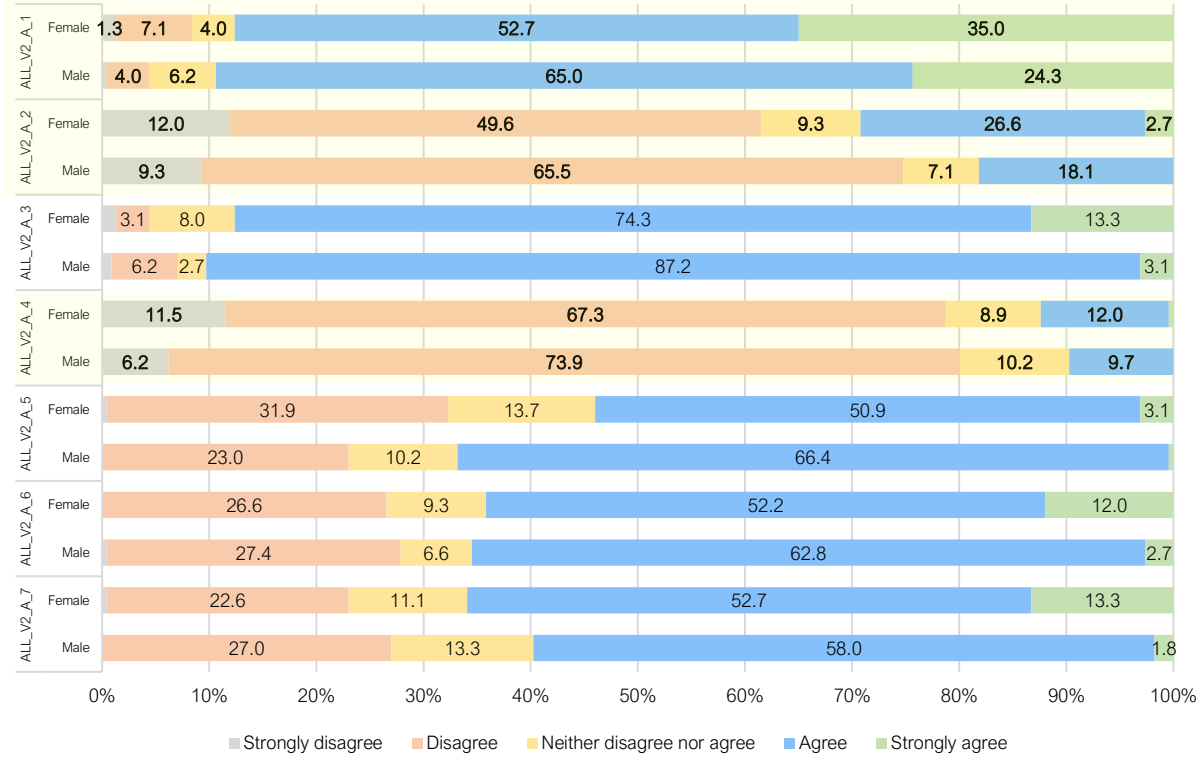
Question code	To what extent do you decide when and how much time you spend on the following activities?
ALL_V1B.A	Your daily schedule
ALL_V1B.C	Your job or other paid work

ALL_V1B.D	Household duties, such as cooking, cleaning, washing clothes, or collecting water or cooking fuel
ALL_V1B.E	Caring for household members, such as children or elderly family members
ALL_V1B.F	Shopping, such as going to the community market
ALL_V1B.G	Attending a social gathering within the community, such as a wedding or other celebration, or a community meeting
ALL_V1B.H	Leisure activities, such as listening to music or chatting with friends; resting; sleeping other than resting

Figure 3.8 shows who makes decisions – men or women – related to time use for different tasks. Most men and women seem to have similar levels of agency over their daily schedule, since most of them decide their schedule by themselves. However, the survey shows that there is a gendered pattern in decision-making around work for pay and doing domestic work/caregiving. Women seem to have more agency than men in deciding the amount of time spent on domestic work/caregiving, but they are more likely to depend on someone else for deciding the amount of time spent on paid work.

This is consistent with gendered norms on division of labour, where domestic work and caregiving are assumed to be women's responsibilities. Having internalised this norm, women have more control over their schedule within this realm compared to men. Meanwhile, the survey shows that women have lesser agency in deciding to work for pay, which provides confirmation of the norm that working for pay is a 'secondary' activity for women, to be undertaken only when it is deemed necessary (Setyonaluri et al., 2021). Interestingly, we see that women report feeling more control over their leisure time, although this is likely because they feel they can allocate their time within the broader structure of unpaid domestic work. They may also perceive some supervisory care time as leisure time (e.g. relaxing with their children).

Figure 3.9. Critical consciousness: Tabulation of responses to seven attitudinal questions



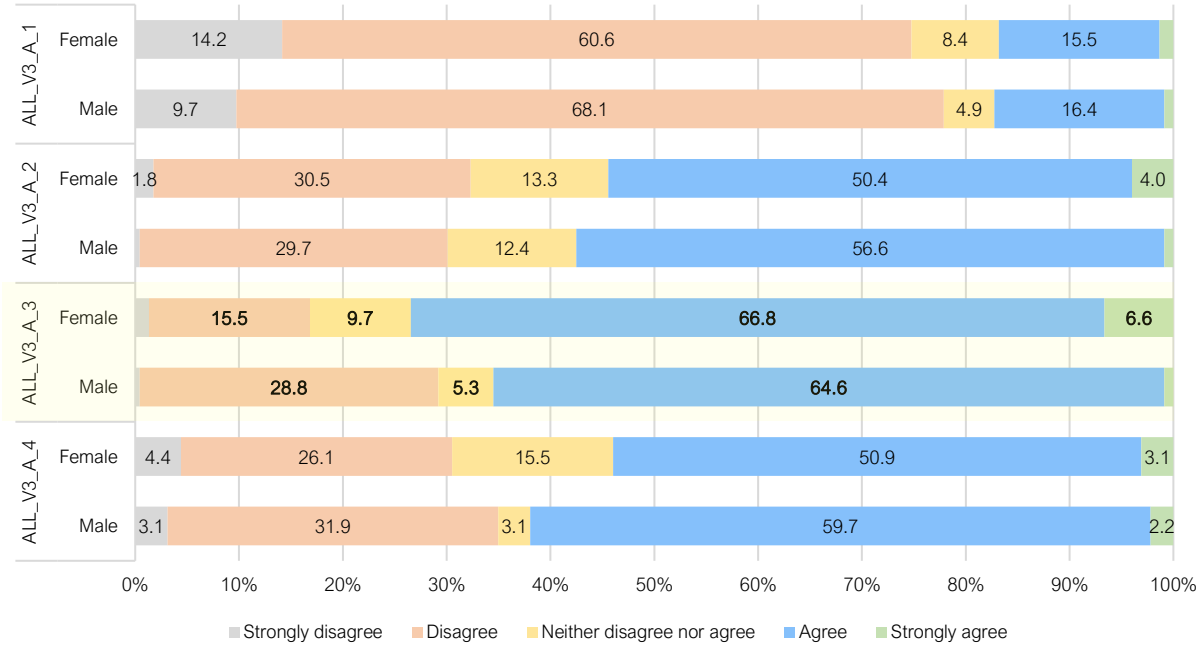
Question code	Table ALL.V2.A. Attitudinal questions (critical consciousness)
ALL_V2_A_1	Men are responsible for providing for the family and/or earning an income; women are responsible for taking care of children and doing domestic work.
ALL_V2_A_2	Women can work, even as the main earners.
ALL_V2_A_3	Women can work, but only as a secondary earners.
ALL_V2_A_4	In this community, women are expected to work longer hours than men.
ALL_V2_A_5	Compared to a woman, a man can change his daily schedule more easily.
ALL_V2_A_6	Because of their responsibilities, women generally sleep less than men.
ALL_V2_A_7	Women's responsibilities take more time than men's responsibilities.

Regarding expectations of women's work, there are three findings to note when it comes to saying who the main earner in the family should be. First, there is a broad agreement that men should provide for the family and women are responsible for domestic work. However, secondly, we see some divergence by gender in disagreement: about 4% of men disagree, whereas 8% of women disagree or strongly disagree. Approximately 75% of males disagree or strongly disagree that women can be the main earners, while 62% of women disagree or

strongly disagree.[18] Third, 28% of women versus 18% of men agree or strongly agree that women can work even as the main earners. These findings highlight the entrenched norm that the breadwinner of the family should be male.

A large share of respondents expressed that they disagree/strongly disagree that in their community, women are expected to work longer hours, but in contrast most respondents agreed that women's responsibilities take more time and women sleep less, indicating that most men and women are aware that women spend disproportionately more time on work (whether it is paid or unpaid). It is possible that respondents interpreted 'work' in question ALL_V2_A_4 as paid work, and 'responsibilities' in ALL_V2_A_7 to include unpaid work. Men and women also diverge when considering whether a man can more easily change his daily schedule – men are more likely to agree with this statement than women.

Figure 3.10. Voice: Tabulation of responses to four attitudinal questions



Question code	Table ALL.V3.A Attitudinal questions (voice)
ALL_V3_A_1	A spouse can change their daily schedule without consulting the other spouse.
ALL_V3_A_2	A spouse can ask the other spouse to do more household duties so that they can do something they want to do.
ALL_V3_A_3	I can ask a household member to help me take care of a child or other family member.
ALL_V3_A_4	I have total control to change the amount of time I spend on paid work.

[18] This finding suggests that 'sticky' norms around who should be the primary earner are prevalent in our sample. While this observation presents another possible barrier to FLFP, it is beyond the scope of this study to fully investigate it.

Notably in Figure 3.10, we see that women are more likely to agree that they can ask for help when it comes to care work (ALL_V3_A_3), whereas we see consistency between male and female responses to the other three questions.

3.2.2. Aggregated Agency Score: Basic Statistics by Demographic Group

This section reports the aggregated scores representing the five different categories of agency. We first illustrate the distribution of scores visually with plots broken down by gender.

Figure 3.11. Distribution of total self-efficacy scores

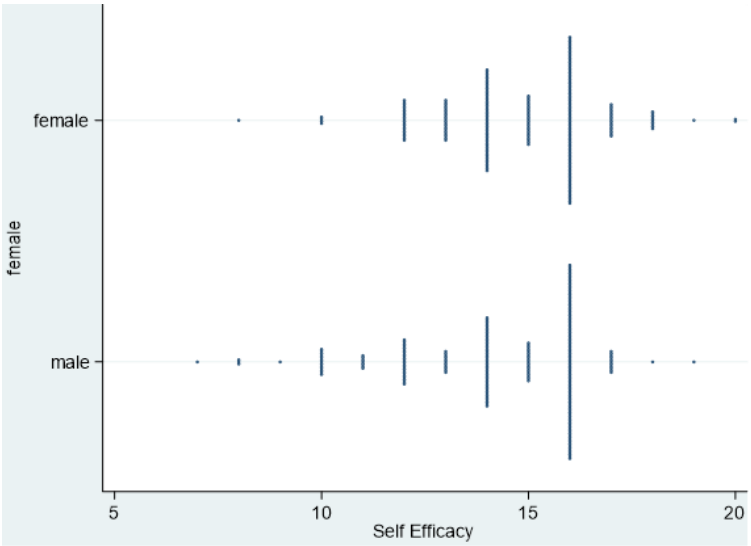


Figure 3.11 shows the distribution of the aggregated Likert scores for questions related to self-efficacy. Notably, for men we see a longer and thicker tail below the mean, indicating that while the means are similar for men and women, there is a cluster of men with relatively high self-efficacy scores and some with very low scores.

Figure 3.12. Distribution of total decision-making scores

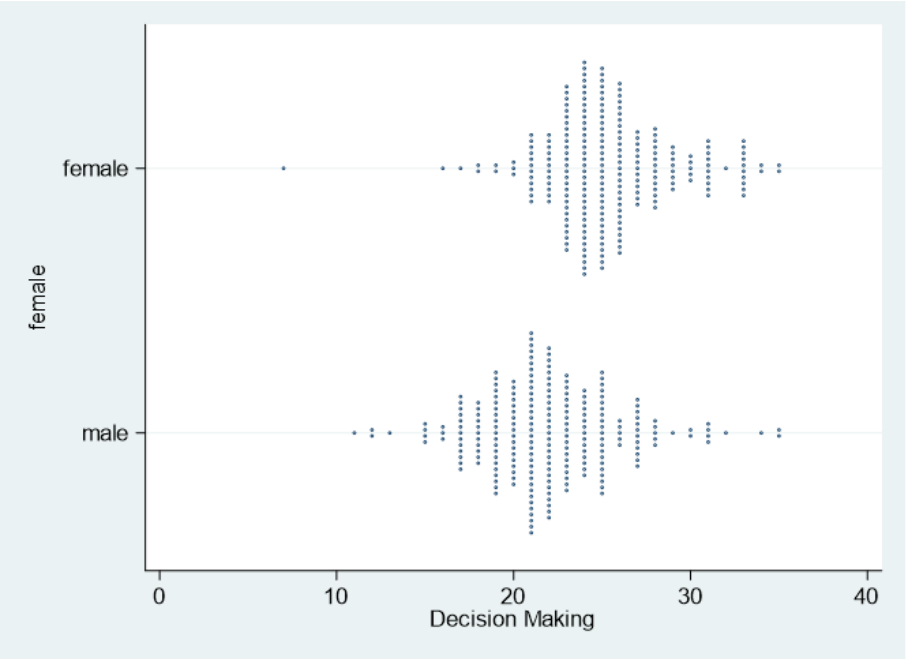
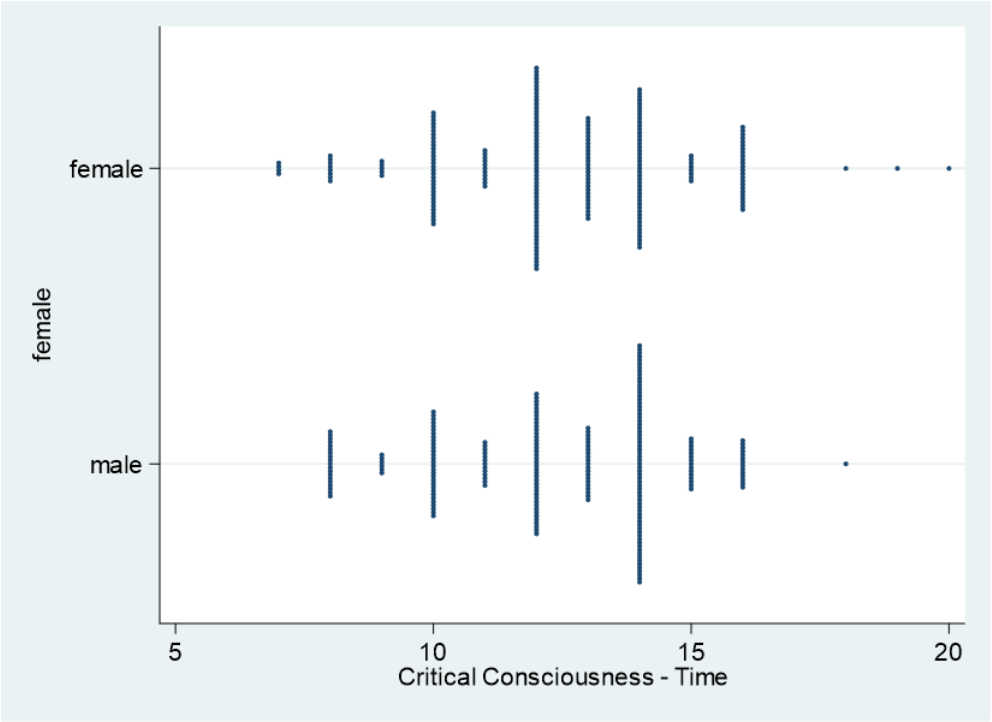


Figure 3.12 reports the distribution of total attitudinal scores for the ability to make decisions. The higher the score, the more the respondent feels like they can decide for themselves about how to allocate their time across daily activities. Interestingly, men report lower decision-making agency than women overall. This is likely because more men are engaged in work for pay, where they have an externally set schedule with little flexibility. Women, if engaged in home production, likely have more agency over the allocation of activities throughout the day.

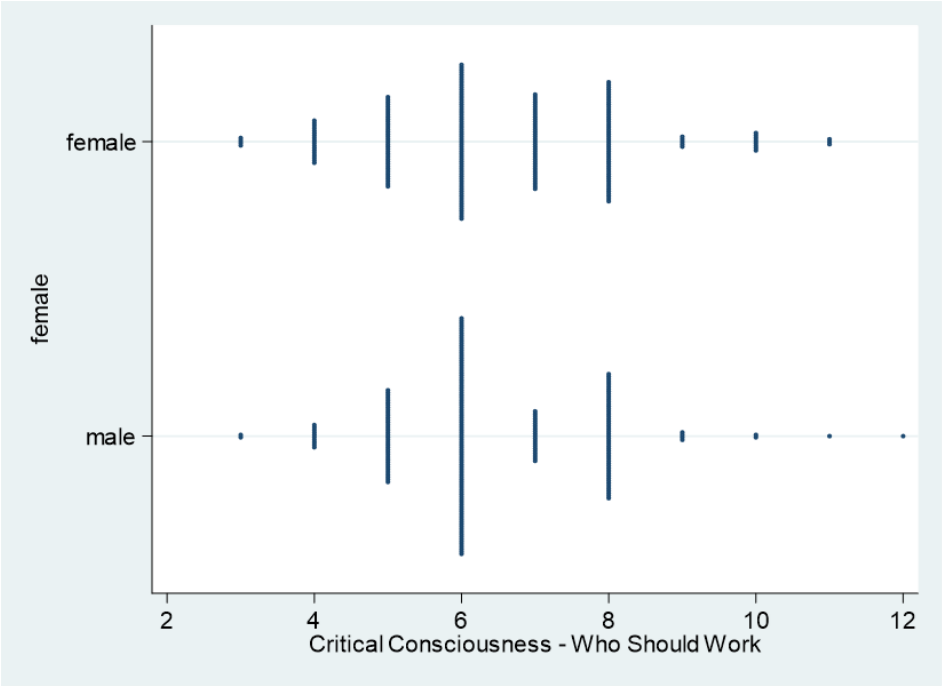
Figure 3.13. Distribution of critical consciousness about who should work



Note: This measure is calculated by aggregating scores for questions ALL_V2_A_1, ALL_V2_A_2 and ALL_V2_A_3. In questions ALL_V2_A_1 and ALL_V2_A_3, a higher score implies a higher level of agency, since the original wording of the questions was such that a lower score meant higher agency.

When it comes to analysing critical consciousness, we find it useful to split critical consciousness into two sub-categories: critical consciousness about who should assume the role of working (an aggregation of scores for ALL_V2_A_1 – 3) and critical consciousness around time use (an aggregation of scores for ALL_V2_A_4 – 7). When asked a series of questions about their attitudes regarding who in the household should work (Figure 3.13) as the main or secondary earner, there is broad agreement among men and women.

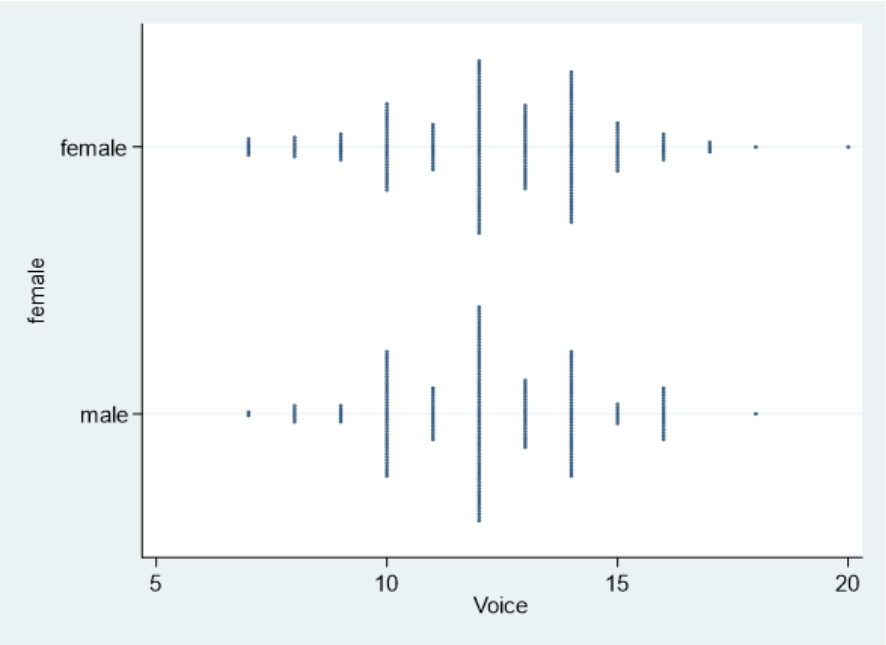
Figure 3.14. Distribution of critical consciousness about how much women work



Note: This measure is calculated by aggregating scores for questions ALL_V2_A_4, ALL_V2_A_5, ALL_V2_A_6 and ALL_V2_A_7.

In Figure 3.14 we report the distribution of scores asking men and women whether women are expected to work more hours in the day. We see that the distribution is similar, though this aggregation masks differences across age of children and other important demographic characteristics, which we examine in the next section.

Figure 3.15 Distribution of voice attitudinal scores



Note: This measure is calculated by aggregating scores for questions ALL_V3_A_1, ALL_V3_A_2, ALL_V3_A_3 and ALL_V3_A_4.

In Figure 3.15, we again see similar means in the aggregated scores but different shapes of the distribution. Women's attitudes about how much women work seem to be more varied than those of their male counterparts.

3.2.3. Multivariate Analysis of Factors Explaining Attitudinal Scores

This section presents a multivariate ordinary least squares regression analysis examining the effect of age, gender and education[19] on attitudes about different forms of agency.

In Table 3.5 we see that being a woman with primary school education results in a higher score for decision-making when compared to men with primary education (row c).[20] This is likely the result of a stronger adherence to gendered work allocations in households with lower levels of education. Further, we see that favourable male attitude toward women working increase as men's level of education increases (row b). It is likely that there is some degree of assortative matching in the marriage market such that men and women of similar educational levels tend to marry each other. Therefore, the higher the level of education, the more likely that both spouses are engaged in formal labour. Interestingly, men with higher education also report a higher critical consciousness about women working more and sleeping less (row b, column 3). In sum, Table 3.5 suggests that education plays an important mitigating role for both men and women's attitudes.

Table 3.5. Regression analysis explaining variation in attitudinal scores

		(1)	(2)	(3)	(4)	(5)
	Variable	**Self-efficacy**	**Decision-making**	**Critical consciousness time**	**Critical consciousness role**	**Voice**
a	Male x secondary (0/1)	0.22	-1.06	-1.044***	0.047	-0.259
		(0.36)	(0.74)	(0.36)	(0.21)	(0.33)
b	Male x tertiary (0/1)	-0.008	-0.309	-1.333**	0.872**	0.069
		(0.62)	(0.95)	(0.56)	(0.39)	(0.56)
c	Female x primary (0/1)	0.551	3.176***	-0.257	0.181	-0.261
		(0.41)	(0.85)	(0.43)	(0.26)	(0.41)
d	Female x secondary (0/1)	-0.201	0.141	-0.06	-0.039	0.184
		(0.47)	(0.97)	(0.50)	(0.32)	(0.47)

[19] Ideally, we should also control for income or economic status. However, measuring income is complex, and we do not have an adequate proxy for economic status in our data.
[20] Note that the questions on decision-making are predominantly related to household chores.

e	Female x tertiary (0/1)	0.307	-0.552	0.784	-0.189	0.185
		(0.77)	(1.19)	(0.85)	(0.57)	(0.85)
f	Age	-0.002	-0.012	-0.001	-0.015*	-0.011
		(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
g	Household has kids < 10 years old (0/1)	0.089	-0.376	0.133	-0.314**	-0.214
		(0.209)	(0.416)	(0.230)	(0.153)	(0.222)
h	Disability household (0/1)	0.089	-1.287	-0.339	-0.28	0.345
		(0.506)	(1.280)	(0.433)	(0.368)	(0.496)
i	No. of household members	-0.008	-0.028	-0.164***	0.05	-0.046
		(0.058)	(0.123)	(0.063)	(0.042)	(0.059)
j	Works for pay (0/1)	0.127	-0.02	-0.015	-0.013	0.488**
		(0.224)	(0.407)	(0.253)	(0.159)	(0.223)
k	Constant	14.148***	23.665***	13.797***	7.028***	12.507***
		(0.631)	(1.268)	(0.678)	(0.463)	(0.631)
l	Observations	452	452	452	452	452
m	R-squared	0.02	0.185	0.063	0.038	0.019

Notes: ***, ** and * indicate statistical significance at 1%, 5% and 10% levels, respectively. The first five explanatory variables are interaction variables to determine the separate effects of education for men and women. The omitted category is male x primary, so all categorical parameter estimates are relative to this category.

3.3. Analytical Results of Agency Lab-in-the-field Experiment

3.3.1. Agency Experiment Descriptive Analysis

This section describes the data collected via the lab-in-the field experiment, as explained in the methods section. Table 3.6 shows the demographic breakdown of each treatment group (public, private, negotiation). Given that people were randomly allocated to treatment groups, we would expect to see few statistical and systematic differences between them. In general, the randomisation into different treatment assignments worked to balance the demographic characteristics, with a few notable exceptions.

Table 3.6. Demographic balance across experimental treatment assignment

						t-test

			Public (1)		Private (2)		Negotiation (3)	Difference		
								Public – Private (4)	Public – Negotiation (5)	Private – Negotiation (6)
		N	Mean/ SE	N	Mean/ SE	N	Mean/ SE			
a	Number of HH members	151	2.377	149	2.403	152	2.52	-0.025	-0.142	-0.117
			[0.132]		[0.146]		[0.151]			
b	Age	151	45.675	149	48.45	152	46.671	-2.774**	-0.996	1.779
			[0.913]		[0.852]		[0.845]			
c	Education	151	1.795	149	1.758	152	1.849	0.036	-0.054	-0.09
			[0.052]		[0.044]		[0.046]			
d	Works for pay	151	1.43	149	1.517	152	1.507	-0.086	-0.076	0.01
			[0.040]		[0.041]		[0.041]			
e	Household has kids < 10 years old	151	0.457	149	0.409	152	0.579	0.048	-0.122**	-0.170***
			[0.041]		[0.040]		[0.040]			

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The values displayed for t-tests are the differences in the means across the groups.

We see that people assigned to the public group were significantly younger (by 2.7 years on average) than those in the private group, but not significantly younger than those in the negotiation group (row b). Furthermore, people in the negotiation group were significantly more likely to have children under 10 years of age (row e).[21] We control for these characteristics in the regression analysis to reduce concerns about these variables having a confounding effect.

The reader should keep in mind that the sample is relatively small, so the standard errors reported tend to be higher than they would be with a higher number of respondents.

Table 3.7 shows the averages of hours reported for each of the treatment groups in response to each vignette.

Table 3.7. Mean hours reported by all respondents by treatment group

	Public			Private			Negotiation		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Vignette 1A	151	3.007	2.977	149	3.107	2.895	152	3.447	3.149
Vignette 1B	151	3.947	3.079	149	4.235	3.133	152	4.717	3.100
Vignette 2	151	3.139	2.683	149	3.423	2.719	152	3.493	2.706
Vignette 3	151	4.57	3.069	149	4.396	3.053	152	4.441	2.987

[21] We define families with young children in this section to be those with children under 10 years old rather than 5 because our sample size of families with children under 5 was too small for reliable regression results.

Note: Each vignette presents the opportunity for someone to attend a training with different conditions associated.

Our hypothesis is that if agency does not matter, we would expect similar outcomes across all treatments. Further, we assume that the private treatment group, where spouses are promised that their responses will be kept confidential by the research team, represents the lowest level of communication between spouses, and therefore the lowest communication cost.[22] The public treatment group puts a single condition on the experiment by stipulating that each respondent's answers will be revealed to their spouse. Therefore, respondents know that their spouse will see their answers, but they do not have to interact with the spouse to determine their responses. Negotiation, where spouses discuss their answers in real time, is the most intensive form of communication and reduces asymmetric information between them. Indeed, negotiation is likely the closest treatment condition to the one that spouses actually operate in. As summarised in Table 3.8, there is preliminary evidence that people change their answers based on the treatment assignment.

Table 3.8. Mean hours reported by women by treatment group

								t-test Difference		
			Public (1)		**Private (2)**		**Negotiation (3)**	**Public – Private (4)**	**Public – Negotiation (5)**	**Private – Negotiation (6)**
		N	Mean/ SE	N	Mean/ SE	N	Mean/ SE			
a	Vignette 1A	76	3	73	2.712	77	3.156	0.288	-0.156	-0.444
			[0.327]		[0.303]		[0.352]			
b	Vignette 1B	76	3.934	73	3.973	77	4.532	-0.038	-0.598	-0.56
			[0.345]		[0.353]		[0.345]			
c	Vignette 2	76	3.368	73	3.329	77	3.429	0.04	-0.06	-0.1
			[0.308]		[0.326]		[0.325]			
d	Vignette 3	76	5.289	73	5.178	77	4.766	0.111	0.523	0.412
			[0.333]		[0.339]		[0.350]			

Note: Each vignette presents the opportunity for someone to attend a training with different conditions associated.

As shown in Table 3.8, women report the highest number of hours in the negotiation treatment for all but vignette 3, and roughly the same number of hours in the public and private treatments. Interestingly, in vignette 3, in which couples are asked how much time their spouse should attend the seminar, women report lower hours in the negotiated treatment than in the public or private treatments, indicating that after negotiation with their husband, they lower the hours they believe the husband should attend (row d). Using a simple t-test,

[22] Communication costs could result from (1) difficulties in bringing up the discussion, (2) possible monetary cost associated with the time allocated to negotiation, or (3) differing opinions. Future costs from negotiation outcomes may include (1) less time allocated to domestic work by a female spouse due to training, and (2) conflicts arising from the failure to reach an agreement.

which provides a pairwise test between treatment groups, we see that none of the differences are statistically significant using this sample.

Table 3.9. Mean hours reported by men by treatment group

								t-test Difference		
			Public (1)		Private (2)		Negotiation (3)	Public – Private (4)	Public – Negotiation (5)	Private – Negotiation (6)
		N	Mean/ SE	N	Mean/ SE	N	Mean/ SE			
a	Vignette 1A	75	3.013	76	3.487	75	3.747	-0.474	-0.733	-0.26
			[0.360]		[0.359]		[0.370]			
b	Vignette 1B	75	3.96	76	4.487	75	4.907	-0.527	-0.947*	-0.42
			[0.366]		[0.372]		[0.367]			
c	Vignette 2	75	2.907	76	3.513	75	3.56	-0.606	-0.653	-0.047
			[0.309]		[0.306]		[0.297]			
d	Vignette 3	75	3.84	76	3.645	75	4.107	0.195	-0.267	-0.462
			[0.355]		[0.348]		[0.332]			

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The values displayed for t-tests are the differences in the means across the groups.

We see in Table 3.9 that men in the negotiation treatment group consistently report the highest average hours in each of the four vignettes (column 3). This is an interesting result, as it suggests that men either feel compelled to report more hours or are being negotiated upward by their wives. Further, men consistently report higher hours in the private treatment group (column 2), with the public group being the lowest (column 1). Again, however, the raw mean differences are statistically insignificant overall.

3.3.2. Agency Experiment Regression Analysis

In the above analysis, we presented a comparison of means across experimental groups and by demographic characteristics. Unfortunately, we can only do this in pairwise fashion, and it is likely that the relationships underlying these results are more complex than presented above.[23]

In this section, we present a multivariate regression analysis in order to allow us to control for important factors that may influence the variation in hours reported across the different treatment groups. Specifically, we examine whether the treatment assignment that varies the social situation between spouses influences the hours reported. Using this method, we can control for gender, education, household size, and whether the household has young children.

[23] Ideally, we would combine this statistical analysis with comprehensive qualitative data. At this point, the qualitative data we have is limited, but we are planning to collect relevant qualitative information for future work.

Table 3.10. Regression results: Hours reported by experimental treatment group – Women

		(1)	(2)	(3)	(4)
	Variable	**Vignette 1a**	**Vignette 1b**	**Vignette 2**	**Vignette 3**
a	Public assignment	0.245	-0.059	0.065	-0.122
		(0.457)	(0.523)	(0.482)	(0.486)
b	Negotiation assignment	0.337	0.542	0.004	-0.429
		(0.472)	(0.521)	(0.474)	(0.497)
c	Works for pay	0.721*	0.473	-0.206	0.713*
		(0.413)	(0.456)	(0.428)	(0.422)
d	Age	0.028	0.028	0.03	0.028
		(0.021)	(0.020)	(0.018)	(0.019)
e	Secondary education (0/1)	0.318	0.615	1.094**	0.242
		(0.451)	(0.493)	(0.449)	(0.469)
f	Tertiary education (0/1)	1.539**	1.2	0.871	1.161
		(0.746)	(0.770)	(0.807)	(0.712)
g	Household size	0.127	0.206	-0.005	0.310**
		(0.147)	(0.151)	(0.141)	(0.135)
h	Unpaid work	0.024	-0.034	0.016	0.005
		(0.056)	(0.061)	(0.056)	(0.058)
i	Self-efficacy row total	0.075	0.263**	-0.006	0.306***
		(0.111)	(0.116)	(0.116)	(0.115)
j	Decision-making row total	-0.004	-0.003	0.008	0.017
		(0.048)	(0.057)	(0.047)	(0.055)
k	Critical consciousness – Who should work row total	0.068	-0.008	0.200*	0.155
		(0.119)	(0.133)	(0.109)	(0.123)

l	Critical consciousness – Time row total	0.079	0.047	0.035	0.279***
		(0.087)	(0.091)	(0.086)	(0.087)
m	Voice row total	0.073	-0.118	0.068	-0.174*
		(0.095)	(0.094)	(0.081)	(0.099)
n	Constant	-3.999	-2.055	-1.148	-5.951**
		(2.774)	(2.936)	(2.738)	(2.767)
o	Observations	226	226	226	226
p	R-squared	0.067	0.07	0.048	0.132

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by female respondents for each of the four vignettes. The omitted experimental base category is the private assignment. All regressions include heteroskedasticity-robust standard errors in parentheses. Each vignette presents the opportunity for someone to attend a training with different conditions associated.

For clarity, we present the multivariate analysis of the experiment by male and female respondents separately. The trade-off is that we reduce our already small sample size and as such our standard errors tend to be large. Table 3.10 shows the effect of different experimental assignments for women and Table 3.11 does the same for men. In addition to the experimental assignment (our variable of primary interest), we control for demographic, work and attitudinal variables to further isolate the effect of communication between spouses on the outcomes. Both regressions compare the public and negotiation treatments to the omitted category, namely, private treatment.

Women have higher reported attendance hours in vignette 1 for both public and negotiation treatments, compared to the private treatment (rows a and b, column 1 and 2). While neither is statistically significant, the point estimate is relatively large, representing about 30 of the mean number of hours reported. This pattern suggests that women report their lowest hours of attendance in the private treatment, when they are assured that their husbands will not learn their answers. While we do not have enough information to explain this fully, we explore several possibilities below.

Further, women report more hours that their husbands should attend the seminar (vignette 3) in the private treatment (rows a and b, column 4). Interestingly, the negotiated number of hours they would attend (vignette 1) and how many they think their husbands should attend (vignette 3) move toward a more equal sharing of hours between the spouses when spouses can negotiate, that is, in the highest level of communication.

Table 3.11. Regression results: Hours reported by experimental treatment group – Men

		(1)	(2)	(3)	(4)
	Variable	**Vignette 1a**	**Vignette 1b**	**Vignette 2**	**Vignette 3**
a	Public assignment	-0.472	-0.472	-0.544	0.028
		(0.511)	(0.532)	(0.446)	(0.510)
b	Negotiation assignment	0.169	0.443	0.065	0.273
		(0.529)	(0.516)	(0.430)	(0.493)
c	Worked for pay	0.486	0.513	0.858*	0.893*
		(0.543)	(0.517)	(0.448)	(0.495)
d	Age	-0.007	-0.008	-0.040**	-0.027
		(0.021)	(0.020)	(0.016)	(0.020)
e	Secondary education (0/1)	0.592	0.856*	0.818**	-0.05
		(0.476)	(0.518)	(0.395)	(0.472)
f	Tertiary education (0/1)	1.609*	2.898***	0.428	2.067***
		(0.828)	(0.734)	(0.650)	(0.755)
g	Household size	0.323**	0.155	0.118	0.252*
		(0.159)	(0.154)	(0.135)	(0.146)
h	Unpaid work	0.005	-0.037	-0.006	-0.02
		(0.106)	(0.106)	(0.091)	(0.087)
i	Self-efficacy row total	0.031	0.152	0.092	-0.076
		(0.112)	(0.118)	(0.086)	(0.097)
j	Decision-making row total	-0.037	-0.053	0.019	-0.084
		(0.058)	(0.060)	(0.052)	(0.052)
k	Critical consciousness – Who should work row total	0.092	0.027	-0.005	0.018
		(0.164)	(0.168)	(0.135)	(0.153)

l	Critical consciousness – Time row total	-0.008	0.043	-0.081	-0.04
		(0.094)	(0.096)	(0.080)	(0.087)
m	Voice	-0.073	-0.202*	-0.085	-0.013
		(0.113)	(0.110)	(0.092)	(0.105)
n	Constant	2.208	3.523	3.628	6.327**
		(2.843)	(2.825)	(2.411)	(2.493)
o	Observations	226	226	226	226
p	R-squared	0.066	0.103	0.085	0.092

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by male respondents for each of the four vignettes. The omitted experimental base category is the private assignment. All regressions include heteroskedasticity-robust standard errors in parentheses. Each vignette presents the opportunity for someone to attend a training with different conditions associated.

Table 3.11 reports similar results as those presented in Table 3.10, but for men only. Again, the omitted category is private treatment, where men are assured that their answers will not be shared with their wives. Interestingly, those assigned to the public category, where spouses do not communicate but their answers are revealed, decrease the number of hours reported that they would attend the seminar (row a, columns 1–3). This is true except for vignette 3, where men are asked how many hours their wives should attend. In the public treatment, they increase their reported hours (row a, column 4). It seems also that in the negotiation treatment, men are 'bargained up' in terms of the hours they report attending (row b). We also see that, overall, both men and women are thinking about care constraints, since the response to vignette 1b (row b, column 2), where the spouse would be around to care for the children, is higher than in vignette 1a in the negotiation treatment (row b, column 1). Interestingly, women decrease the number of hours they report attending the seminar if their spouse is around.

We also see that education is an important explanatory variable in each vignette. This is not surprising, given that the training is a form of education and higher-educated people are likely to select into this training more readily. Particularly tertiary education seems to have a significant and positive effect for both women and men's own attendance to the seminar, although this effect goes away for women with a tertiary degree when their spouse is around.

Overall, we see that men seem to be adjusting their hours down when they know their wives will learn their answers, while women are adjusting their hours up. We further explore this pattern by (1) examining whether these findings differ in households with or without young children, and (2) exploring factors that explain the *difference* between a respondent's reported

hours of attendance (vignette 1) and their spouse's preference for how long they should attend (vignette 3).

Table 3.12. Regression results explaining hours reported by experimental group – Women in households with children under 10

		(1)	(2)	(3)	(4)
	Variable	**Vignette 1a**	**Vignette 1b**	**Vignette 2**	**Vignette 3**
a	Experimental assignment , public	-0.219	-0.842	-0.266	-1.536**
		(0.700)	(0.902)	(0.911)	(0.749)
b	Experimental assignment negotiation	0.76	0.475	0.453	-0.453
		(0.799)	(0.925)	(0.803)	(0.594)
c	Worked for pay	1.394**	0.703	0.405	0.784
		(0.575)	(0.642)	(0.643)	(0.595)
d	Age (when missing birth date)	0.017	0.050*	0.014	-0.025
		(0.039)	(0.030)	(0.028)	(0.031)
e	Secondary education (0/1)	-0.011	0.78	1.101*	0.308
		(0.677)	(0.741)	(0.597)	(0.732)
f	Tertiary education (0/1)	1.663	1.964*	1.564	1.908*
		(1.156)	(1.094)	(1.100)	(1.124)
g	Household size	0	-0.112	-0.149	0.217
		(0.244)	(0.274)	(0.247)	(0.208)
h	Unpaid work	-0.033	0.042	0.106	0.025
		(0.096)	(0.103)	(0.090)	(0.118)
i	Self-efficacy row total	0.142	0.412*	0.161	0.142
		(0.164)	(0.226)	(0.152)	(0.196)
j	Decision-making row total	0.002	0.025	0.137*	0.191
		(0.093)	(0.083)	(0.071)	(0.128)

k	Critical consciousness – Who should work row total	0.303*	0.099	0.302**	0.07
		(0.159)	(0.204)	(0.151)	(0.171)
l	Critical consciousness – Time row total	0.196	0.057	0.122	0.280*
		(0.148)	(0.152)	(0.116)	(0.145)
m	Voice	0.055	-0.125	0.179*	-0.087
		(0.138)	(0.171)	(0.099)	(0.157)
n	Constant	-7.420**	-5.867	-10.515***	-5.872
		(3.770)	(4.347)	(3.463)	(4.144)
o	Observations	108	108	108	108
p	R-squared	0.167	0.142	0.164	0.188

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by male respondents for each of the four vignettes. The omitted experimental base category is the private assignment. All regressions include heteroskedasticity-robust standard errors in parentheses. Each vignette presents the opportunity for someone to attend a training with different conditions associated.

In Tables 3.12 and 3.13 we present the same results as in Tables 3.10 and 3.11, again separated by gender, but this time we limit our analysis to households with young children to see whether care constraints change people's responses. Unlike in the full sample, we see little evidence that care constraints significantly alter responses, but this is likely due to the small sample size (columns 1 and 2). Again, we can look at the signs of the coefficients for suggestive evidence. The only statistically significant result is that women reduce the number of hours they report their husbands should attend in the public treatment group (row a, column 4). This may indicate a wife's hesitation to report a high number of hours for her husband to attend, knowing he will learn her answer despite having no communication with him. Women in this treatment with young children shift their answers down based on the knowledge that their husbands will know their answers. Interestingly, when spouses are permitted to negotiate, women decrease the number of hours they report their husbands should attend, but by less (row b, column 4). This is not statistically different from the private treatment.

Interestingly, we see in both tables that, again, when spouses negotiate the number of hours each should attend, women are bargained up compared to the private treatment group, while men are bargained down. This may suggest that when given the ability to negotiate, spouses can negotiate around the norms that each spouse may assume in the private or public treatments. This may also indicate that asymmetric information, present in the private and public treatments, is at play in creating these differences.

Table 3.13. Regression results explaining hours reported by experimental group – Males in households with children under 10

		(1)	(2)	(3)	(4)
	Variable	**Vignette 1a**	**Vignette 1b**	**Vignette 2**	**Vignette 3**
a	Experimental assignment: public	-0.715	-1.265	-0.026	0.549
		(0.974)	(0.850)	(0.563)	(0.710)
b	Experimental assignment: negotiation	-0.42	0.754	0.457	0.967
		(0.804)	(0.833)	(0.454)	(0.617)
c	Works for pay	0.935	0.231	1.630***	0.766
		(0.871)	(0.896)	(0.549)	(0.691)
d	Age (when missing birth date)	0	0.016	-0.044	-0.038
		(0.041)	(0.037)	(0.029)	(0.028)
e	Secondary education (0/1)	0.03	0.344	1.1	-0.329
		(0.735)	(0.733)	(0.709)	(0.722)
f	Tertiary education (0/1)	1.036	2.142**	0.599	1.363
		(1.337)	(1.050)	(1.098)	(1.188)
g	Household size	0.34	-0.126	0.063	0.34
		(0.300)	(0.255)	(0.224)	(0.222)
h	Unpaid work	-0.071	-0.013	-0.037	0.001
		(0.144)	(0.142)	(0.120)	(0.102)
i	Self-efficacy row total	-0.051	0.114	0.051	-0.058
		(0.180)	(0.184)	(0.121)	(0.155)
j	Decision-making row total	-0.051	-0.046	0.032	-0.137*
		(0.101)	(0.089)	(0.078)	(0.070)
k	Critical consciousness – Who should work row total	0.327	-0.027	0.08	0.299
		(0.254)	(0.246)	(0.232)	(0.228)

l	Critical consciousness – Time row total	-0.022	-0.046	-0.126	-0.131
		(0.164)	(0.153)	(0.113)	(0.140)
m	Voice	0.008	0.038	-0.018	0.246*
		(0.185)	(0.151)	(0.145)	(0.135)
n	Constant	1.388	3.627	2.133	3.358
		(4.693)	(4.305)	(2.960)	(2.899)
o	Observations	110	110	110	110
p	R-squared	0.091	0.12	0.136	0.184

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by male respondents for each of the four vignettes. The omitted experimental base category is the private assignment. All regressions include heteroskedasticity-robust standard errors in parentheses. Each vignette presents the opportunity for someone to attend a training with different conditions associated

3.3.3. Analysis of the Difference Between Husband-and-Wife Time Use Preferences in the Experiment

Delving further into our experimental data, we can examine the difference between how many hours a person would like to attend the training and how many hours their spouse would like them to attend. Examining this difference allows us to understand another level of intra-household time use preferences. We also examine the factors that may explain this difference.

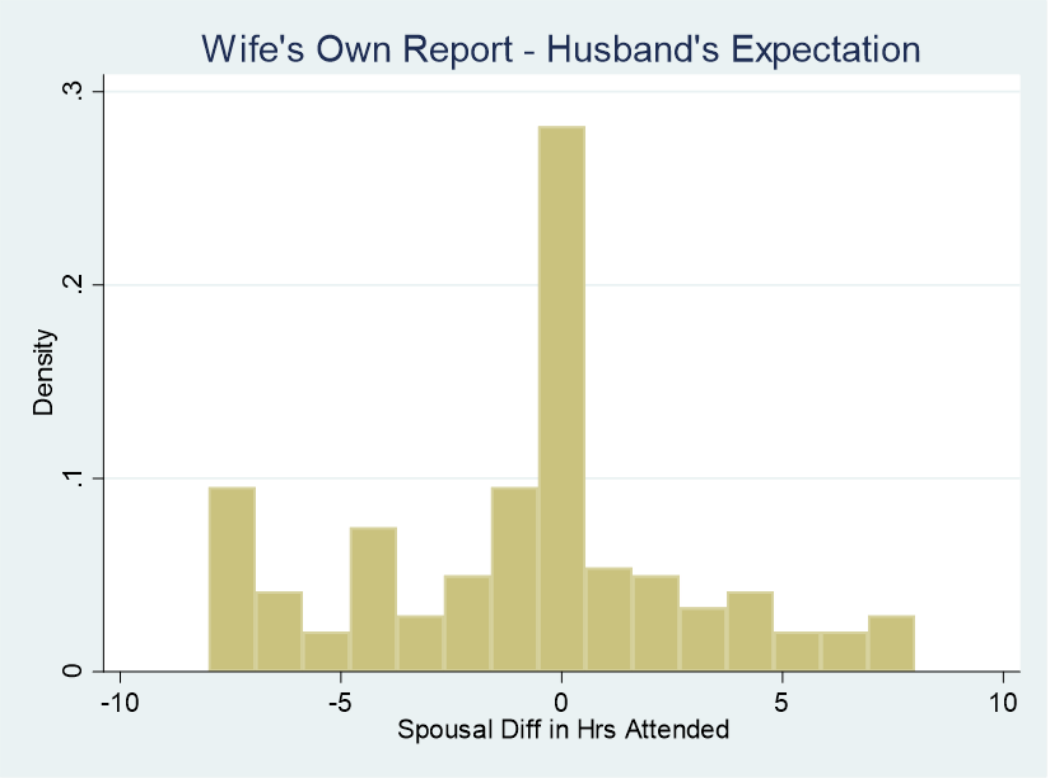
Spousal differences are defined as follows:

$$\text{Female difference} = (\text{Woman's response to vignette 1}) - (\text{Man's response to vignette 3}) \quad (3.1)$$

$$\text{Male difference} = (\text{Man's response to vignette 1}) - (\text{Woman's response to vignette 3}) \quad (3.2)$$

The response to vignette 1 is the respondent's stated preference for the number of hours they themselves would like to attend a training seminar. Vignette 3 asks how many hours the respondent thinks their spouse should attend. When comparing these responses, if the female difference (equation 3.1) is positive, then the wife would like to attend more hours than her husband would like her to attend. If the female difference is negative, then her husband would like her to attend more hours than she prefers.

Figure 3.16. Spousal differences in hours of training attendance (wife's minus husband's expectations)



In Figure 3.16 we see that the difference between a wife's desired number of hours and her husband's expectation about the duration of her attendance spans from –8 to +8. About 30% of couples agree on the number of hours that a wife should attend, which is represented by the highest bar at the zero mark. However, it is also the case that 43% of women report fewer hours than their husbands wish for them, and 27% report more hours of attendance than their husbands wish for them. Overall, the mean of this distribution is just below zero, at –0.87.

Figure 3.17. Spousal differences in hours of training attendance (husband's minus wife's expectations)

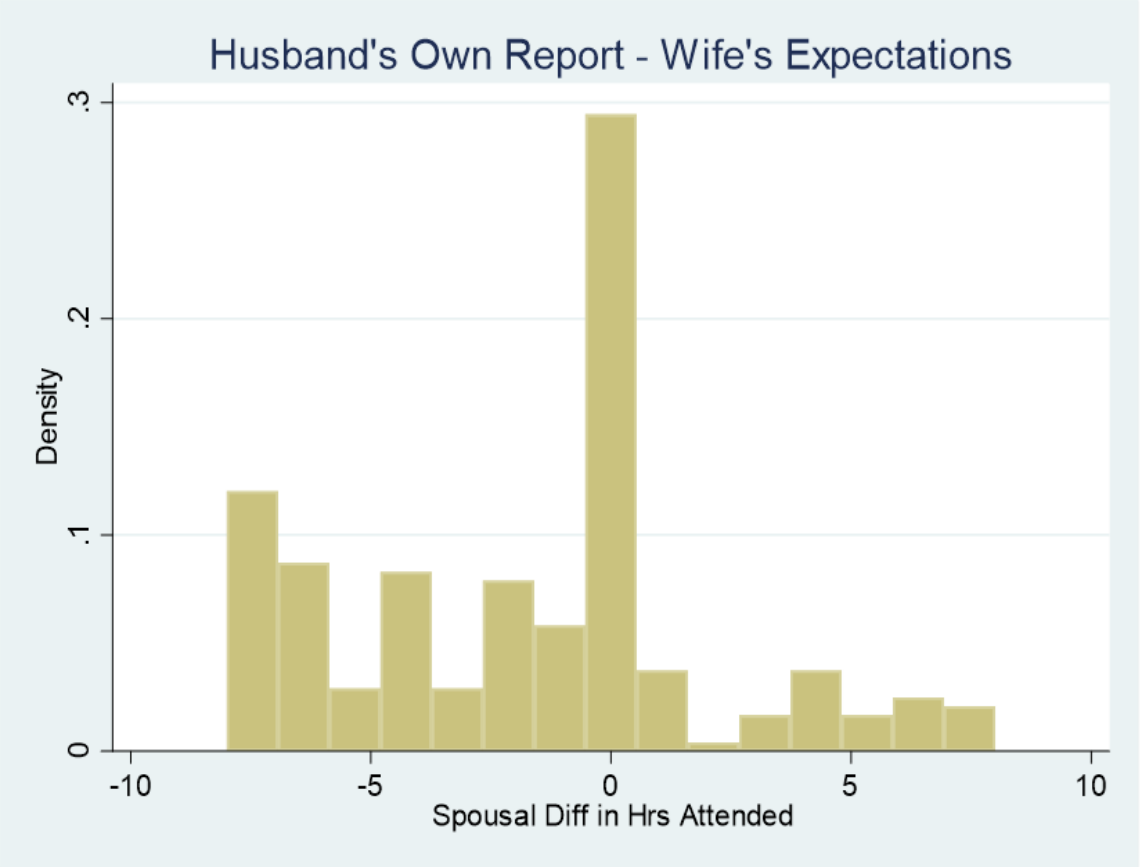


Figure 3.17 shows visibly more weight on the left side of the distribution, suggesting that the wives in our sample are consistently reporting more hours for their husbands to attend the training than the husbands' stated preference. Again, we see that about 30% of couples agree about how many hours the husband should attend the seminar, but unlike in the case of women's attendance above, 52% of men report fewer hours than their wives' preference. Only 17% report more hours than their wives' preference. The mean of this distribution is –1.68, compared to the mean for women of –0.87.

In both cases, on average, the husband and wife would like their spouse to attend more hours than the spouse would like to attend. This difference, however, is twice as large in the case of the husband (husband's own report minus wife's expectation). In the case of the husband, the wife has a much higher expectation of his attendance time relative to his own stated preference for attending.

In this section we explore why these differences exist and why it might be that, on average, the difference is much greater in the case of the husband than the wife. We explore three possibilities: (1) the experimental assignment explains these differences, (2) attitudes about time use agency explain these differences, and (3) time use obligations, such as paid and unpaid work, explain these differences.

Table 3.14. Average difference between own preference and spouse's preference for training attendance by experimental group

								t-test Difference		
			Public (1)		Private (2)		Negotiation (3)	Public – Private (4)	Public – Negotiation (5)	Private – Negotiation (6)
		N	Mean/ SE	N	Mean/ SE	N	Mean/ SE			
a	Wife's own report minus husband's expectation	76	-0.75	73	-0.932	77	-0.948	0.182	0.198	0.017
			[0.463]		[0.427]		[0.393]			
b	Husband's own report minus wife's expectation	75	-2.307	76	-1.737	75	-1.013	-0.57	-1.293**	-0.724
			[0.439]		[0.434]		[0.424]			
c	Spousal difference in hours attended (total)	151	-1.523	149	-1.342	152	-0.98	-0.181	-0.543	-0.362
			[0.325]		[0.305]		[0.288]			

Note: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

As in the earlier histograms, we see that in each experimental group and for both spouses, the number of hours for which an individual reports they would like to attend the training is less than the number of hours their spouse would like them to attend. This difference is much smaller for women, suggesting that there is a group of women who are interested in attending more than their husbands want them to (also seen in the histogram). It seems that the experimental assignment has little effect on generating this difference, except in the case of the difference between the husband's own report minus wife's expectations in the negotiation group and the public group. We see that when spouses negotiate, they get closer to an agreement on how many hours a husband should attend the training.

One notable difference is the direction in which negotiation pushes husbands versus wives. When husbands and wives are able to negotiate about how much a wife should attend, the difference becomes more negative, meaning that either the wife reduces the number of hours she is stating she will attend, or the husband increases the number of hours he wishes she would attend. In the case of husbands, the opposite is true. Thus, when spouses negotiate, they get closer to an agreement.

Our objective in exploring these differences is to start to understand their origin and to hopefully better understand the results we see here and in the experimental regression analysis. Table 3.15 suggests that it is unlikely that the experimental assignment is the primary driver of these differences for women.

Table 3.15. Factors explaining differences between preferred hours of training attendance

	(1)	(2)
Variable	**Wives**	**Husbands**

a	Private	-0.391	0.591
		(0.664)	(0.622)
b	Negotiation	-0.335	1.351**
		(0.636)	(0.641)
c	Paid work/own production	0.046	0.013
		(0.065)	(0.072)
d	Unpaid work	0.167**	0.115
		(0.082)	(0.115)
e	Supervisory care	-0.058	-0.159
		(0.086)	(0.142)
f	Leisure total	0.033	0.016
		(0.038)	(0.042)
g	Self-efficacy	-0.12	-0.036
		(0.154)	(0.131)
h	Decision-making	-0.048	-0.053
		(0.070)	(0.069)
i	Critical consciousness – Time	0.067	0.06
		(0.125)	(0.117)
j	Critical consciousness – Who should work	0.011	-0.045
		(0.153)	(0.196)
k	Voice	0.158	-0.034
		(0.122)	(0.136)
l	Works for pay	0.797	0.302
		(0.574)	(0.643)
m	Age	0.022	-0.044
		(0.028)	(0.030)
n	Secondary education	-0.356	0.749
		(0.614)	(0.578)
o	Tertiary education	-0.047	0.918

		(1.205)	(1.184)
p	Household size	-0.049	0.013
		(0.202)	(0.204)
q	Constant	-4.279	-0.12
		(3.732)	(3.413)
r	Observations	226	226
s	R-squared	0.051	0.06

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

Interestingly, we see that the factors that explain differences between wives and husbands' differences vary. In the case of a wife, we see that an increase in unpaid work hours reduces the discrepancy between her husband's and her own preferred amount of time to attend. We need to interpret this positive sign with caution. It means that women with more unpaid work hours are likely to report more hours that they would like to attend the seminar and/or their husbands report fewer hours. Specifically, for each hour of unpaid work performed by a woman, the difference between her and her husband's preferences increases by 0.167 hours (row d, column 1).

We also see that for husbands, it is the experimental assignment of being in the negotiation group, compared to the public group, that has a significant effect on the difference between a husband's stated preference and the number of hours that his wife wishes him to attend (row b, column 2). A positive sign here means that couples in the negotiation treatment are coming closer together in terms of their preferences for the husband's attendance at the training. Again, a bit more analysis suggests that this finding is primarily driven by being assigned to the negotiation group, resulting in a higher number of hours that husbands report they are willing to attend when spouses can negotiate. Effectively, we can think of this as a form of intra-household bargaining, where a wife bargains her husband's attendance hours up.

The importance of the findings presented in this section lies in the fact that unpaid work may be seen as a barrier to women's attendance at a professional seminar. Women who perform a high amount of unpaid work seem to increase the number of hours they would like to attend the seminar. This may be a result of the framing of the seminar to gain skills for the paid labour market. Their husbands modestly decrease the number of hours they report that their wives should attend, possibly because they see the amount of unpaid work their wives are performing. These results point, potentially, to an unmet demand for entry points (such as a training seminar) into the labour market by women who do a lot of unpaid work.

3.4. Multivariate Analysis for Time Use and Agency Attitudinal Questions

In this section we use a multivariate analysis to explore how stated levels of agency for women affect the amount of paid work, unpaid work and leisure time allocated within a day. These results should not be interpreted as causal but rather as correlations between reported attitudes about agency and time use. This is an important distinction because it is possible, for example, that being involved in paid work may affect the attitudes a woman holds about certain types of agencies (reverse causality). Our research design in this section prevents us from developing a causal model.

However, we can still learn something about the correlation between agency attitudes and time use allocation, since we are able to control for age, education and other factors in our multivariate analysis. These results refer only to women in our sample.

Table 3.16. Regression results explaining hours of work and school with agency attitudinal scores for women

		(1)	(2)	(3)
	Variable	**Paid work (hrs)**	**Learning**	**Employed (0/1)**
a	Self-efficacy	-0.115	-0.006	-0.015
		(0.153)	(0.006)	(0.020)
b	Decision-making	0.128*	-0.003	0.01
		(0.073)	(0.002)	(0.010)
c	Critical consciousness – Time	-0.014	-0.012	0.003
		(0.136)	(0.008)	(0.015)
d	Critical consciousness – Who should work	0.174	-0.006	0.031
		(0.183)	(0.008)	(0.022)
e	Voice	0.278**	-0.006	0.023
		(0.116)	(0.009)	(0.020)
f	Number of household members	-0.127	-0.003	-0.02
		(0.227)	(0.005)	(0.025)
g	Age	-0.034	0.002	-0.007*
		(0.034)	(0.001)	(0.004)
h	Secondary education	-0.744	0.03	-0.122
		(0.651)	(0.036)	(0.079)

i	Tertiary education	-0.238	0.002	-0.011
		(1.048)	(0.021)	(0.133)
j	Disability household	-1.366*	-0.022	0.024
		(0.732)	(0.028)	(0.150)
k	Household has kids < 10 years old	-0.172	-0.019	-0.065
		(0.63)	(0.04)	(0.08)
l	Observations	226	226	226
m	R-squared	0.05	0.029	0.056

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by respondents for paid and productive work. All regressions include bootstrapped standard errors in parentheses.

We see in Table 3.16 that women who report having high levels of control over their time and who feel they can ask a family member for help with household chores are more likely to work (rows b and e, column 1), although this is a correlation and as such it could also be that the reverse is true.

Table 3.17. Regression results explaining hours of unpaid work with agency attitudinal scores for women

		(1)	(2)	(3)	(4)	(5)
	Variable	**Unpaid domestic services for household and family members**	**Unpaid caregiving services for household and family members**	**Unpaid volunteer, trainee and other unpaid work**	**Female unpaid work (sum of cols. 1–3)**	**Supervisory care**
a	Self-efficacy	0.269**	-0.06	-0.001	0.207	0.094
		(0.116)	(0.072)	(0.028)	(0.144)	(0.180)
b	Decision-making	-0.061	0.018	0.001	-0.042	0.124
		(0.055)	(0.044)	(0.022)	(0.073)	(0.082)
c	Critical consciousness – Time use	0.047	-0.025	0.036	0.058	0.076
		(0.063)	(0.057)	(0.047)	(0.123)	(0.128)
d	Critical consciousness – Who should work	-0.147	-0.247***	0.038	-0.356**	-0.574***
		(0.094)	(0.095)	(0.033)	(0.153)	(0.137)

e	Voice	-0.165*	-0.061	0.035	-0.191	0.059
		(0.087)	(0.057)	(0.025)	(0.123)	(0.124)
f	Number of household members	0.025	0.202*	0.036	0.263	(0.220)
		(0.137)	(0.112)	(0.044)	(0.225)	(0.235)
g	Age	0.002	-0.032*	-0.001	-0.031	(0.045)
		(0.018)	(0.018)	(0.003)	(0.024)	(0.031)
h	Secondary education	0.406	-0.16	0.12	0.366	(0.880)
		(0.453)	(0.325)	(0.179)	(0.464)	(0.585)
i	Tertiary education	-0.561	-0.361	-0.121	-1.043	0.537
		(0.551)	(0.451)	(0.108)	(0.756)	(1.225)
j	Disability household	-0.904	1.333*	-0.19	0.239	(0.280)
		(0.636)	(0.725)	(0.130)	(1.167)	(1.267)
k	Household has kids < 10 years old	-0.716*	1.418***	0.006	0.709	3.217***
		(0.386)	(0.245)	(0.116)	(0.517)	(0.601)
l	Observations	226	226	226	226	226
m	R-squared	0.092	0.264	0.024	0.107	0.205

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively. The dependent variables are hours reported by respondents for unpaid work. All regressions include bootstrapped standard errors in parentheses.

In Table 3.17 we see that when women hold strong opinions about who should work and about their ability to ask family members for help, they engage in less unpaid labour (row d, column 2). We note that women who agree with the statement that women experience more time constraints do less supervisory care. The one contrary result is in the case of self-efficacy, where we see that those who report a strong sense of self-efficacy also engage in more unpaid domestic work, but not caregiving work (row a, columns 1 and 2). We see that having young children in the household decreases women's hours of unpaid domestic services but increases their hours of unpaid caregiving (row k, columns 1 and 2). Further, women who are already working, time-constrained, and more likely to report that women should work, are also the ones engaged in less supervisory care (because they are working).

Table 3.18. Regression results explaining hours of leisure activities with agency attitudinal scores for women

		(1)	(2)	(3)	(4)	(5)

	Variable	Socialising and communication, community participation and religious practice	Culture, leisure, mass media and sports practices	Self-care and maintenance	Travelling and waiting	Leisure (sum of cols. 1–4)
a	Self-efficacy	0.198	0.367**	0.016	0.003	0.584**
		(0.207)	(0.161)	(0.077)	(0.059)	(0.262)
b	Decision-making	0.222***	0.039	-0.110*	0.056*	0.207
		(0.086)	(0.065)	(0.065)	(0.031)	(0.133)
c	Critical consciousness – Time use	-0.171	-0.042	-0.094	-0.026	-0.333*
		(0.148)	(0.121)	(0.075)	(0.039)	(0.196)
d	Critical consciousness – Who should work	-0.187	-0.287*	-0.036	0.064	-0.446*
		(0.214)	(0.151)	(0.098)	(0.049)	(0.266)
e	Voice	-0.616***	-0.143	-0.019	-0.036	-0.815***
		(0.158)	(0.157)	(0.069)	(0.045)	(0.197)
f	Number of household members	-0.196	-0.644***	-0.158	0.043	-0.955***
		(0.228)	(0.173)	(0.107)	(0.064)	(0.309)
g	Age	0.02	0.024	-0.022	-0.015**	0.007
		(0.028)	(0.025)	(0.018)	(0.007)	(0.048)
h	Secondary education	-0.346	0.789	-0.24	-0.188	0.015
		(0.636)	(0.651)	(0.297)	(0.174)	(1.098)
i	Tertiary education	1.528	0.652	-0.438	0.67	2.412
		(1.055)	(1.252)	(0.481)	(0.423)	(2.016)
j	Disability household	0.46	-1.976	0.976	0.557	0.016
		-1.065	-1.256	-0.868	-0.592	-2.163
k	Household has kids < 10 years old	-2.129***	-0.938*	-0.332	0.14	-3.259***
		-0.63	-0.568	-0.35	-0.179	-0.905
l	Observations	226	226	226	226	226
m	R-squared	0.197	0.129	0.079	0.099	0.177

Notes: ***, ** and * indicate statistical significance at the 1%, 5% and 10% levels, respectively.

If we look at aggregated leisure time, we see in Table 3.18 that women who are more likely to agree that women can be the main earners and less likely to agree that women can only be secondary earners, have more time constraints and lower levels of leisure overall (row d). Further, if a woman reports strong agreement with the statement that she can ask a spouse for help (voice), she also tends to have less leisure time (row e). This result may be driven by women who are already working and feel that they can ask for help but also have less leisure time. Women who report higher levels of control over their daily schedule (decision-making) report more hours of socialising but fewer hours of self-care, and more hours of travel (row b).

3.5 Limitations

Several potential limitations should be considered when interpreting the results of this study. First, the size of the sample is relatively small and only covers two urban areas (Greater Jakarta and Greater Surabaya). Therefore, it is not indicative of the Indonesian population. Moreover, the size of the sample does not afford us an understanding of all relationships of interest, as small samples may lead to conservative estimates.

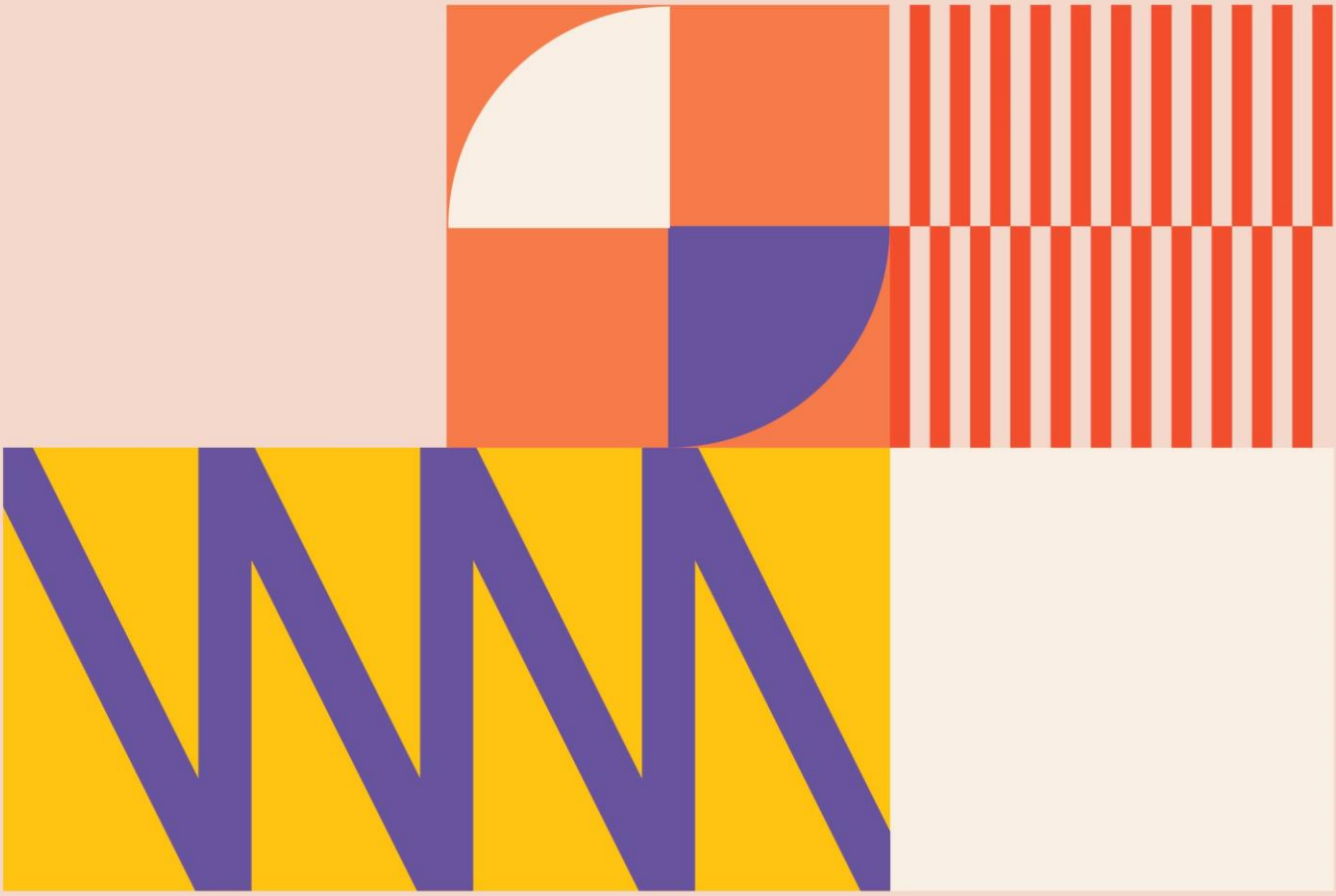
Second, although we applied multistage sampling and randomisation of respondents and diary days, there is an overrepresentation of older respondents. Only 5.5% of respondents are aged 29 and younger, while more than 70% are aged 40 and older. This is because older couples are relatively less mobile and may have been more available for interviews during the data collection process. There are two implications of this sample structure. First, the results may not fully reflect the gender gap in time use for paid and unpaid care work among younger workers with young children at home. Second, the overrepresentation of older couples in the sample may have influenced the average time spent on unpaid domestic and care work by women. As older couples tend to have grown-up children who need less intensive care, this may result in potentially conservative estimates of time spent on unpaid care work.

Third, one should not interpret the regression analysis describing time use under the assumption that attitudinal scores as causal. While the regressions demonstrate correlations between attitudinal scores and time use allocations, they do not suggest that increasing agency levels measured through attitudinal scores will *cause* differences in time use. It is possible that the causal relationship is the other way around - time use affects people's attitudes about their agency. For example, women involved in more unpaid work may be more likely to agree that they have more control over such work. This type of 'reverse causality' invalidates any causal interpretation of the results.

Lastly, we are concerned that there could be spill-over effects in the experimental design, particularly for those couples assigned to the private group. Given that many of the dwellings were quite small, it was not always possible to fully separate the spouses, so it is possible that a spouse could overhear the other's answers. This would distort the difference between public and private treatments in particular.

Section 4

Lessons and Programmatic Implications



4. Lessons and Programmatic Implications

The aim of this section is to answer the three central policy/program questions raised in section 1.2.3. The first question was whether women bear most of the unpaid care and domestic responsibilities, and how that may influence their engagement in paid work. We addressed this question by looking at how respondents allocate their time, their attitudes toward time use, and their agency in the use of time. The second question focused on identifying the most important policy levers and programmatic priorities to promote equitable participation in the economy for both men and women. In our study design and analyses, we considered five factors – the load of unpaid care work one bears, potential income-generating opportunities, intra-household negotiation, prevailing social norms, and one's education level. Finally, in response to the third question, we suggest how to efficiently collect time use data by utilising and fine-tuning instruments and generating insights on practicalities.

Policy/program question 1: Why do women participate less in the formal labour market than men?

In this study, we analysed the time spent on care and domestic work, its impact on economic activities, and agency over the allocation of time dedicated to paid work. Our study uncovered significant gaps and imbalances in the way that paid and unpaid work are distributed between men and women. Men spend more time on paid employment activities, while women spend almost three times as much time on unpaid domestic work and engage in more unpaid active caregiving services. Women also spend twice as much time as men on supervisory care, and even more when they have children under the age of 5. The uneven sharing of unpaid work leaves women with less time to engage in the formal labour market.

Our study also used attitudinal survey results to examine how much agency women and men exercise over time use. This revealed that in addition to care responsibilities, attitudinal barriers also discourage women from participating in the paid labour market. There is broad agreement that men should provide for the family and women are responsible for domestic work. Women tend to report higher levels of self-efficacy and decision-making regarding domestic and care responsibilities, but men are more likely to ask for help with these duties. Women also exhibit less agency when it comes to deciding to work for pay. These results likely reflect the fact that women have either internalised the view that domestic and caregiving activities are their responsibility or have less access to alternative care arrangements. Men's low level of decision-making agency may be attributed to their engagement in paid employment, the schedule of which is externally established.

Nevertheless, our survey of attitudes suggests that there is some divergence between men and women's views on women's participation in the labour market. Overall, women are more aware and readier to engage in paid work, but male attitudes do not match women's willingness. A larger share of women than men disagree or strongly disagree with the view that men should be the main breadwinners, and a larger share of women also think that

women can work as the main earners. Importantly, we note that women who agree with the statement that women experience more time constraints tend to do less supervisory care.

Attitudes are also correlated with educational levels. Generally, a higher level of education is associated with greater acceptance of women's paid employment. Yet, the current labour allocation among families penalises women in terms of unpaid work and self-care (especially sleep). Favourable male attitudes toward women working increase as men's education increases. Men with a higher level of education also report a higher critical consciousness about women working more and sleeping less.

Our lab-in-the field experiment addressed three questions related to negotiation of time allocation and decision-making in the face of a new/hypothetical economic opportunity: 1) how women and men negotiate about time allocation; 2) whether their behaviours conform to societal expectations; and 3) whether their decisions change depending on the presence of their spouse or the need to negotiate with them.

Our analyses reveal that the amount of time women and men are willing to allocate for training varies depending on whether they must decide on their own, reveal their response to their spouse and/or negotiate with their spouse. In general, when spouses can negotiate with one another, their preferred hours of training increase for both women and men.

Women reported the lowest hours of attendance in the private treatment group, that is, when their husbands were unaware of their answers. This may reflect an entrenched idea about who determines paid work-related activities, as we saw in the attitudinal section. Both men and women see women as the primary caregivers and men as the decision-makers regarding paid work.

However, we see in the negotiation assignment that when men and women speak openly with each other, women decrease the number of hours that men should attend, and men increase the number of hours they report their wives should attend. Similarly, men reported the highest preferred number of hours of training when they negotiated with their wives, but they wished for their spouse to attend for longer. In the negotiation treatment group, men were 'bargained up' for their own attendance and their wives'. This result is robust across all types of households and indicates that spouses can negotiate around the norms that each spouse may assume in the private or public treatment groups.

Another key finding was that care constraints, education and the number of hours of unpaid work are interlinked. Care constraints are important. When looking at all households, both men and women tend to want to attend more hours of training if the scenario mentions that their spouse is around to take care of the children. Higher education is associated with a higher number of hours of training wished for among both men and women.

Additionally, the results indicate that the number of hours that respondents would like to attend was lower than what their spouses preferred for them, though this difference tends to be greater among women. For women, these gaps are driven by the amount of unpaid care work (i.e. for each unpaid work hour performed by women, the difference increases by 0.167

hours). Women with more unpaid work hours are likely to report a higher number of hours they would like to attend the seminar, while their husbands report fewer hours for them, which drives up the gap. For men, the gaps can be attributed to being in the negotiation treatment group. A positive sign of the regression coefficient result suggests that couples in the negotiation treatment are coming closer together in terms of their preferences for the husband's attendance at the training.

Policy/program question 2: What are the most important policy levers and programmatic priorities to promote equitable participation in the economy?

This study highlights the gender disparity in unpaid care work and how it affects the economic opportunities available to women, especially those with young children. The differences in household responsibilities borne out by time use statistics, as well as results from the lab-in-the field experiment where a hypothetical economic opportunity was offered, suggest that unpaid domestic and care work is a barrier to women's participation in paid work. Simply increasing women's labour force participation will be complicated by the high opportunity cost of this participation. A comprehensive program that aims to support women's roles in the paid labour market is needed. Below we outline the possible dimensions through which such initiative can be enacted:

- **Educate.** One way to change attitudes is through education. In our study, both men and women who had completed more years of education than their peers shared more egalitarian views on unpaid and paid work and on gender roles within the household. That being said, education is a long-term investment that could forge the shift in gender norms and its associated roles. Encouraging and normalising the sharing of unpaid domestic work between adults within the household, regardless of gender, is an important step to improving FLFP.
- **Conduct campaigns.** Campaigns that target attitudes about women entering the paid labour force are also important. Women, particularly those who are engaged in high levels of unpaid labour, demonstrated an unmet demand for training that could lead to paid employment. However, there was a disparity between men and women's attitudes toward women working for pay, with women being more likely to strongly agree or agree that women can be the main income earners. Realization by both men and women about these attitudinal barriers is a necessary component of a suite of interventions that are needed to increase FLFP.
- **Invest in the care ecosystem.** Investing in a high-quality care ecosystem can help reduce the burden on women and have high payoffs for children, families and the economy. Without care infrastructure developed to support their participation, caregivers, who are often women, face a dual burden of managing care responsibilities and participating in the labour market. However, Indonesia faces significant childcare gaps, hindering children's development and women's economic participation. Refer to Prospera et al. (2022), UNESCAP (2022) and ADB et al. (forthcoming) for more insights and recommendations.

- **Prioritise production, analysis and use of high-quality data on time use.** As the saying goes, we cannot improve what we don't measure. Production and utilisation of good time-use data is beneficial to inform policies that advance gender equity and close gender gaps. Table 1.1 lists a number of time use studies conducted in various countries; their recommendations range from overarching strategies to justify the provision of more child and elderly care, extended maternity leave and flexible working arrangements, to evaluating the impact of family leave on time use of families and household income earning strategies. Therefore, investing in the timely collection and analysis of time use surveys, adding attitudinal survey questions where feasible, will be vital.

Policy/program question 3: How can BPS efficiently administer data collection on time use and agency in the near future as part of Sakernas?

This pilot contributed to the development of international guidelines for collecting standardised time use statistics on women's unpaid work, and to the testing and refinement of a set of instruments that can eventually be rolled out through Sakernas. Results of the pilot suggest that the data collected is of good quality. The pilot also provided insights on how the data collection can be operationalised. However, it should be noted that the pilot was carried out in urban areas only, so testing the instrument in rural areas is necessary prior to a national rollout to assess understanding, acceptance and precision of the instrument for all populations. Based on the urban pilot, the following steps are recommended to ensure efficient collection of time use data.

Account for both paid and unpaid work. One of the most important methodological conclusions from our work is that using only a narrow definition of work (one based on monetary compensation) results in an egregious mismeasurement of total household labour. This mismeasurement is to the detriment primarily of women because it undervalues the important contributions they make to the household, their communities and the overall economy. This conclusion is in keeping with the broader evidence base and has historically provided the rationale for undertaking time use measurement via comprehensive, diary-based surveys.

Account for multitasking in measures of time use. Understanding the multitude and interconnectedness of activities that people, particularly caregivers, engage in is critically important when reporting time use statistics and designing policies meant to support caregivers. Indeed, this has been the justification for many detailed time use surveys in recent years. Failing to account for the multitude of services provided in the course of unpaid work (care, home and own-use production) will result in inefficient and potentially ineffective policies. Increasing the labour force participation of workers active in the care economy is more complicated than simply providing substitute care. All the tasks undertaken by such workers must be considered, including those performed simultaneously.

Consider the distinction between supervisory and active care. One of the most important simultaneous activities is supervisory care, that is, being 'on-call' for those needing care. While one could introduce policies to support or substitute for active care, providing options for supervisory care is more complicated, given that it is often performed as a secondary activity. Moreover, people may be more reluctant to pay for services that substitute for supervisory care.

Methodologically, this type of care is difficult to measure and is often prone to underreporting. Defining it as paid work and assigning a cost to it is difficult. This highlights the value of careful qualitative research, such as the earlier cognitive testing conducted prior to the pilot to assess respondent comprehension. Cultural understanding must also be established in order to properly word the questions/prompts and ensure that respondents respond appropriately to probes on supervisory care, such as those included in the recovery section of the pilot questionnaire.

Include a dedicated 'recovery' section to directly probe for supervisory care responsibilities. Creating a recovery section is important, as it helps to ensure that all simultaneous unpaid care time is accounted for. In our study, the recovery section provided important inputs for measuring unpaid care work. Lessons from the study's cognitive testing round were in line with existing evidence that participants are not always conscious that they were minding or keeping an eye on children or other dependents while simultaneously performing other activities. For this reason our study retained this feature of the ILO pilot module, despite the additional respondent burden.

Apart from the core components that must be present in the survey tool itself, the pilot study identified several important factors that should be kept in mind when rolling out time use surveys in the future, as outlined below.

Selection of respondents. Collection of time use data may be challenging as it requires a balanced sample of diary days. Respondents may not be available at certain times, so enumerators may need to collect the data on additional days. One way to circumvent this issue is by working together with local authorities prior to the data collection and reserve a number of respondents who will be notified about the survey prior to the fielding to ensure their availability.

It must be noted that our pilot survey focused on households with couples because it aimed to better understand the dynamics of time use and agency within households. However, the requirement to interview both spouses complicated the logistics of the survey. Given the diversity of households and household arrangements in Indonesia, it may be sufficient to randomly select one member per household when conducting time use surveys at national level.

Training of enumerators. Training of enumerators is essential to ensure quality, accuracy and reliability of the collected data. Enumerators need to be thoroughly familiar with the

survey instrument, including questions, response categories and skip patterns, and know how to administer it consistently and in a standardised manner. Additionally, they should have a good understanding of time use concepts, such as primary and secondary activities, and be detail-oriented. Apart from these specific requirements of time use surveys, enumerators must have excellent communication skills to establish rapport with respondents, manage time efficiently to complete the survey within the allocated time, adhere to ethical principles, and follow up with respondents to clarify any incomplete or ambiguous responses to ensure data quality.

Time required. This pilot study consisted of the following modules: labour participation module, time use survey, attitudinal survey, and lab-in-the field experiment. The time needed to complete the time use survey ranged from 10 to 25 minutes per respondent, while the attitudinal survey and the agency experiment together took 10 to 20 minutes per respondent. Because a lab-in-the-field experiment may be too complicated to administer during a national labour force survey, delivering an attitudinal survey alongside a time use module may be a more practical approach. This would avoid overburdening both enumerators and respondents while still providing valuable insights into prevailing attitudes and agency regarding time use.

Language and translating 'supervisory care'. Considering Indonesia's ethnic and linguistic diversity, it is sensible to adapt the survey module when necessary, especially when the survey collects information in rural areas. Lessons from the cognitive testing phase suggest that it may be helpful to check the semantics that prevail among the target respondents with regard to describing and measuring time. For example, in West Kalimantan, 'yesterday' can refer to any day prior to today. The phrase commonly used to refer to the previous day is 'last night' (*semalam* in Bahasa Indonesia). Using prayer times as time markers may also be helpful, considering that the majority of Indonesians are Muslim.

In sum, this study finds that women do more work in the care economy than their male counterparts, and that these unpaid work duties are a significant barrier to their participation in the paid labour force. However, both attitudes and care responsibilities are notable barriers to FLFP. Without appropriate care infrastructure development, along with simultaneous efforts to normalise women's participation in the paid labour force, only modest gains in FLFP can be made.

References

Abraham, K. G., Helms, S., & Presser, S. (2009). How social processes distort measurement: The impact of survey nonresponse on estimates of volunteer work in the United States. *American Journal of Sociology*, *114*(4), 1129–1165.

Abraham, K. G., Maitland, A., & Bianchi, S. M. (2006). Nonresponse in the American time use survey: Who is missing from the data and how much does it matter? *International Journal of Public Opinion Quarterly*, *70*(5), 676–703.

Addati, L., Cattaneo, U., Esquivel, V., & Valarino, I. (2018). *Care work and care jobs for the future of decent work.* International Labour Organization. https://www.ilo.org/global/publications/books/WCMS_633135/lang--en/index.htm

Alonso, C., Brussevich, M., Dabla-Norris, E., Kinoshita, Y., & Kochhar, K. (2019). *Reducing and redistributing unpaid work: Stronger policies to support gender equality* (IMF Working Paper, October 15, 2019). International Monetary Fund. https://www.imf.org/en/Publications/WP/Issues/2019/10/15/Reducing-and-Redistributing-Unpaid-Work-Stronger-Policies-to-Support-Gender-Equality-48688

Asian Development Bank. (2019). *The CAPI effect: Boosting survey data through mobile technology*. https://www.adb.org/publications/capi-survey-data-mobile-technology

Asian Development Bank. (Forthcoming). *Investments in Childcare for gender equality in Asia and the Pacific*

Badan Pusat Statistik. (2021). Hasil Sensus Penduduk Indonesia 2020, Berita Resmi Statistik No. 7/01/Th. XXIV, 21 Januari 2021.

Bell, A., Ward, P., Tamal, M. E. H., & Killilea, M. (2019). Assessing recall bias and measurement error in high-frequency social data collection for human-environment research. *Population and Environment*, *40*, 325–345.

Buvinic, M., & King, E. M. (2018). *Invisible no more? A methodology and policy review of how time use surveys measure unpaid work*. Data2x. https://data2x.org/wp-content/uploads/2019/05/Data2X-Invisible-No-More-Volume-1.pdf

Cameron, L., Suarez, D. C., & Rowell, W. (2018). *Female labour force participation in Indonesia: Why has it stalled?* (Melbourne Institute Working Paper No. 11/18). Melbourne Institute: Applied Economic & Social Research. https://melbourneinstitute.unimelb.edu.au/publications/working-papers/search/result?paper=2938388

Chang, L., & Krosnick, J. A. (2003). Measuring the frequency of regular behaviors: Comparing the 'typical week' to the 'past week'. *Sociological Methodology*, *33*(1), 55–80.

Data2x. (2018). *Invisible no more? Country case studies*. https://data2x.org/wp-content/uploads/2019/05/Data2X-Invisible-No-More-Volume-2.pdf

Devercelli, A. E., & Beaton-Day, F. (2020). *Better jobs and brighter futures: Investing in childcare to build human capital*. World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/35062/Better-Jobs-and-Brighter-Futures-Investing-in-Childcare-to-Build-Human-Capital.pdf?sequence=5&isAllowed=y

Donald, A., Koolwal, G., Annan, J., Falb, K., & Goldstein, M. (2020). Measuring women's agency. *Feminist Economics*, *26*(3), 200–226.

Eissler, S., Heckert, J., Myers, E., Seymour, G., Sinharoy, S., & Yount, K. M. (2021). *Exploring gendered experiences of time use agency in Benin, Malawi, and Nigeria as a new concept to measure women's empowerment* (IFPRI Discussion Paper 2003). International Food Policy Research Institute. https://ebrary.ifpri.org/digital/collection/p15738coll2/id/134275

Ferrant, G., Pesando, L. M., & Nowacka, K. (2014). *Unpaid care work: The missing link in the analysis of gender gaps in labour outcomes* (OECD Development Centre, December 2014). https://www.oecd.org/dev/development-gender/Unpaid_care_work.pdf

Folbre, N. (2006). Measuring care: Gender, empowerment and the care economy. *Journal of Human Development*, *7*(2), 183–199.

Folbre, N. (2018). *Developing care: Recent research on the care economy and economic development*. International Development Research Centre. https://www.iaffe.org/media/cms_page_media/788/Folbre_Nancy.pdf

Fricker, S., & Tourangeau, R. (2010). Examining the relationship between nonresponse propensity and data quality in two national household surveys. *Public Opinion Quarterly*, *74*(5), 934–955.

IDX, IBCWE, & BOI Research. (2022). *Sensus Perempuan dalam Tim Kepemimpinan Eksekutif di perusahaan IDX200*. https://www.ibcwe.id/riset/detil_riset/89

Ingen, E. V., Stoop, I., & Breedveld, K. (2009). Nonresponse in the Dutch time use survey: Strategies for response enhancement and bias reduction. *Field Methods*, *21*(1), 69–90.

International Conference of Labour Statisticians. (2013). *Resolution concerning statistics of work, employment and labour underutilization*. Adopted by the 19th ICLS (October 2013). https://www.ilo.org/global/statistics-and-databases/standards-and-guidelines/resolutions-adopted-by-international-conferences-of-labour-statisticians/WCMS_230304/lang--en/index.htm

International Finance Corporation. (2018). *Tackling childcare: The business case for employer-supported childcare in Sri Lanka*. ICF Women in Work Program. https://www.ifc.org/wps/wcm/connect/topics_ext_content/ifc_external_corporate_site/gender+at+ifc/priorities/employment/SriLanka_Tackling_Childcare

Juster, F. T., Ono, H., & Stafford, F. P. (2003). 2. An Assessment of Alternative Measures of Time Use. *Sociological Methodology*, *33*(1), 19–54.

Kabeer, N. (1999). Resources, agency, achievements: Reflections on the measurement of women's empowerment. *Development and Change*, *30*(3), 435–464.

Lagarde, C., & Ostry, J. D. (2018, November 28). *Economic gains from gender inclusion: Even greater than you thought.* IMF Blog. https://www.imf.org/en/Blogs/Articles/2018/11/28/blog-economic-gains-from-gender-inclusion-even-greater-than-you-thought

Prospera. (2022). *Designing better childcare for women, children, and inclusive recovery in urban Indonesia and beyond.*

https://childcare.prospera.or.id/assets/doc/Briefing%20note_Designing%20better%20childcare.pdf

Prospera, ADB, ILO, World Bank, IFC, UN Women, & Mobile Creches. (2022). *The future of care economy for women, children and quality jobs: Passing the baton and translating the G20 commitment into concrete actions* [Proceedings of a G20 side event on 11 October 2022 in Jakarta, Indonesia].

Schaner, S., & Das, S. (2016). *Female labor force participation in Asia: Indonesia country study* (ADB Economics Working Paper Series No. 474). Asian Development Bank. https://www.adb.org/publications/female-labor-force-participation-asia-indonesia-study

Setyonaluri, D., Nasution, G., Ayunisa, F., Kharistiyanti, A., & Sulistya, F. (2021). *Social norms and women's economic participation in Indonesia.* Investing in Women, Prospera, and Lembaga Demografi. https://investinginwomen.asia/knowledge/social-norms-womens-economic-participation-indonesia/

Shafik, M. (2021). *What we owe each other: A new social contract*. Random House.

Sijapati Basnett, B., Riyanto, U., Hadiarta, A., & Istanti, I. (2022). Towards an inclusive response to combat COVID-19 impacts on women and people with disabilities: Lessons from Indonesia for peer countries. In S. M. Indrawati, S. Nazara, T. Anas, C. F. Ananda & K. Verico (Eds.), *Keeping Indonesia safe from the COVID-19 pandemic: Lessons learnt from the National Economic Recovery Programme.* ISEAS Yushof Ishak.

Sinharoy, S., Cheong, Y. F., Seymour, G., Heckert, J., Johnson, E., & Yount, K. M. (2021). *The time-use agency scale: Development and validation of a measure for Ghana and beyond*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3986265

Singh, P., & Pattanaik, F. (2020). Unfolding unpaid domestic work in India: Women's constraints, choices, and career. *Palgrave Communications*, *6*(1), 1–13.

Smith, V. L. (1989). Theory, experiment and economics. *Journal of Economic Perspectives*, *3*(1), 151–169.

Tulving, E. (1983). *Elements of episodic memory.* Clarendon Press.

UN Department of Economic and Social Affairs. (2005). *Guide to producing statistics on time use: Measuring paid and unpaid work.* https://unstats.un.org/unsd/publication/seriesf/seriesf_93e.pdf

UN Statistics Division. (2022). *Minimum Harmonized Instrument for the Production of Time-Use Statistics*. https://unstats.un.org/unsd/statcom/53rd-session/documents/BG-3h-TimeUseStats-rev2-E.pdf

UN Women. (2021). *Measuring time use: An assessment of issues and challenges in conducting time use surveys with special emphasis on developing countries*. UN Entity for Gender Equality and the Empowerment of Women, UN Women. https://data.unwomen.org/publications/measuring-time-use-assessment-issues-and-challenges-conducting-time-use-surveys

UNICEF, UNDP, Prospera, & SMERU. (2022). *The social and economic impact of COVID-19 on households in Indonesia: A second round of surveys in 2022.* https://www.unicef.org/indonesia/reports/social-and-economic-impact-covid-19-households-indonesia

UNESCAP. (2021). *COVID-19 and the Unpaid Care Economy in Asia and the Pacific.* https://www.unescap.org/sites/default/d8files/knowledge-products/2021_Regional_Report_Covid19-Care-Economy.pdf

Wesselow, M., & Stoll-Kleemann, S. (2018). Role-playing games in natural resource management and research: Lessons learned from theory and practice. *Geographical Journal*, *184*(2), 298–309.

Wodon, Q., & de la Brière, B. (2018). *Unrealized potential: The high cost of gender inequality in earnings.* The Cost of Gender Inequality series, World Bank. https://openknowledge.worldbank.org/handle/10986/29865

World Bank. (Forthcoming). *Narrowing the gender gaps: The initial gendered impacts of the COVID-19 pandemic on Indonesia's labor market.*

YouGov & Investing in Women. (2020). *Gender equality matters 2020: Social norms, attitudes and practices of urban millennials in Indonesia, Philippines and Vietnam.* https://investinginwomen.asia/knowledge/gender-equality-matters-2020-social-norms-attitudes-practices-urban-millennials-indonesia-philippines-vietnam/


PROSPERA
Australia Indonesia Partnership
for Economic Development
INVESTING IN WOMEN
SMART ECONOMICS
AN INITIATIVE OF THE AUSTRALIAN GOVERNMENT
FEB
Lembaga
Demografi
Australian Government