Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Gender Statistics
Geneva, Switzerland, 10–12 May 2023
Item H of the provisional agenda
New approaches to measuring unpaid work and work-life balance

New approaches to modular time use measurement: Harnessing national labour force surveys for the measurement of unpaid care work

Note by the International Labour Organization*

Abstract

2023 marks a decade since the 19th International Conference of Labour Statisticians (ICLS) adopted a new resolution on international standards concerning "statistics of work, employment, and labour under-utilisation".

The new standards signified a major step change in the measurement of productive activities in official statistics and provide a strong conceptual framework and attendant reference definitions to underpin the production of statistics on own use provision of services / unpaid domestic and care work (UDCW) in national labour force surveys (LFS). However, measurement and analysis of UDCW remains something of a specialised field and presents distinctive measurement challenges, not routinely encountered in household sample surveys.

In this paper, we present recent work to operationalise the 19th ICLS standards to enhance the visibility of UDCW within national labour statistics and to support analysis of gender-based inequalities in divisions of paid and unpaid work, and total work time. Such data have relevance for a wide range of public policy areas, especially when collected alongside data on labour force participation and employment.

*Prepared by (Samantha Watson and Kieran Walsh)

> We report on ongoing work to operationalise the standards over a three-year
> programme of international and mixed-methods piloting, developed by the ILO, with
> the support of Data 2X, and in partnership with national statistics offices and
> research institutes in different regions.

I. Introduction

1. This paper presents ongoing work to operationalise the 19th International Conference of Labour Statisticians (ICLS) standards to enhance the visibility of own use provision of services / unpaid domestic and care work (UDCW) within national labour statistics.

2. Convened every five years since 1923 (with few exceptions), the ICLS is the internationally recognised standard-setting body for labour statistics[1]. In October 2013, the 19th ICLS adopted a new resolution on international standards "concerning statistics of work, employment, and labour under-utilisation"[2]. These standards superseded those adopted under the 13th ICLS Resolution "concerning statistics of the economically active population, employment, unemployment and underemployment", which had set the scope for labour statistics for over 30 years.

3. The new standards introduced an internationally agreed statistical definition of "work" as a reference concept. As such, work "comprises 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". The definition encompasses both paid and unpaid productive activities and applies regardless of the (in)formality or (il)legality of the sector and status of the work[3], or the economic unit in/for, which it is performed.

4. Within this definition, five separate and mutually exclusive "forms of work" are identified[4], with persons potentially occupying multiple work situations within a given reference period. The five forms of work are:

   a) Own-use production work comprising production of goods and services for own final use

   b) Employment work comprising work performed for others in exchange for pay or profit

   c) Unpaid trainee work comprising work performed for others without pay to acquire workplace experience or skills

   d) Volunteer work comprising non-compulsory work performed for others without pay

---

[1]The International Conference of Labour Statisticians (ICLS) is the recognized standard-setting body in labour statistics. The ICLS makes recommendations on selected topics of labour statistics, articulated in resolutions and guidelines. Once approved by the Governing Body of the International Labour Organisation (ILO), a specialized agency of the United Nations, these become part of the set of international standards on labour statistics.

[2]ILO (2013) "Resolution I: Resolution concerning statistics of work, employment and labour underutilization" ICLS-Resolution-I-[STATI-131114-1]

[3]Activities that do not involve production of goods or provision of services (e.g., begging and stealing), that relate to "self-care" (e.g., personal grooming and hygiene), or that fail to meet the "third-person criterion" (i.e., activities that cannot be performed by another person on one's own behalf, e.g., sleeping, learning and recreation) are excluded.

[4]ILO (2013) "Resolution I: Resolution concerning statistics of work, employment and labour underutilization" ICLS-Resolution-I-[STATI-131114-1], 3: 7

e) Other work activities (not defined in the resolution, but encompassing activities such as court mandated unpaid work)

5. The 19th ICLS signified a major step-change in the measurement of productive activities for national statistics. The revised standards enabled forms of work that had historically been omitted or marginalised to be made more visible, concepts of employment, unemployment, and labour under-utilisation to be refined, and international categories to be better aligned. One important outcome of these changes is the potential for much improved analysis of gender-based inequalities in labour force participation, employment characteristics, divisions of labour, and total work time.

6. The new standards are especially notable for extending the remit of labour statistics to encompass work activities that fall outside of the system of national accounts' (SNA) "production boundary" but within the SNA's broader "general production boundary"[5]. As a result, own use provision of services (also termed "unpaid domestic and care work - UDCW) is brought into the reference scope of activities for labour force statistics for the first time (as is direct volunteer work to provide services, via a separate sub-category).

Figure 1
19th ICLS forms of work framework, alignment to the system of national accounts (SNA)
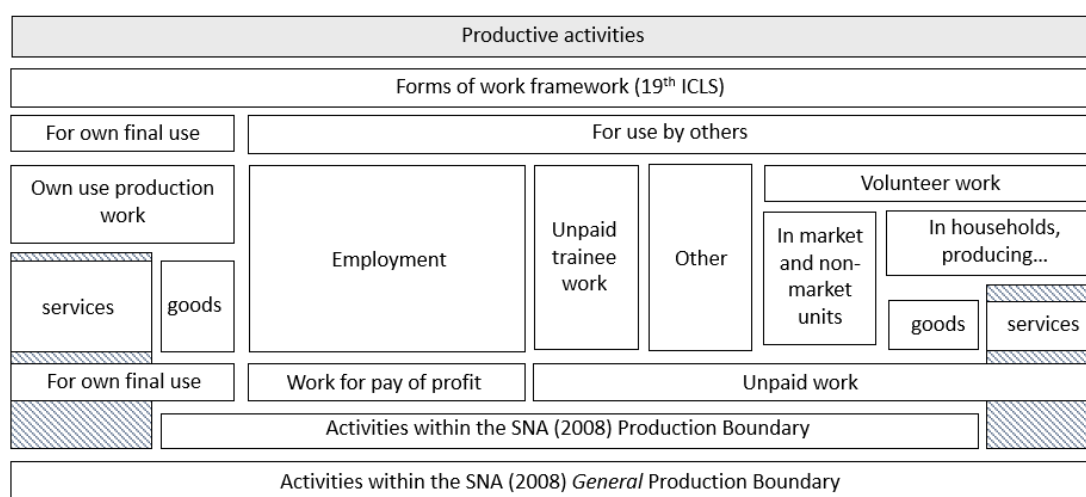
7. Although the 19th ICLS provides a strong conceptual framework and attendant reference definitions to underpin the production of statistics on UDCW, operationalisation of the standards has proven challenging. Reflecting on progress made since the 19th ICLS, the 20th ICLS (held in 2018) identified a "critical need for guidance and methodological development related to time-use methods", noting the "many challenges… [raised by] data collection…analysis, and use of the data generated"[6].

8. Responding to this need, the ILO initiated a programme of work to support the production of statistics on own use provision of services / unpaid domestic and care work through the intermittent inclusion of a dedicated questionnaire module in national labour force surveys

---

[5]The 19th ICLS concept of work is aligned to the General production boundary as defined in the System of National Accounts 2008 and its concept of economic unit that distinguishes i. market units (i.e., corporations, quasi-corporations and household unincorporated market enterprises); ii. non-market units (i.e., government and non-profit institutions serving households), iii. households that produce goods or services for own final use.
[6]ILO (2018), "Report III: Report to the Conference", ICLS-20-2018-3-Report III-[STATI-181106-1], 27:126

(LFS), with a focus on low- and middle-income countries (LMICs). The remainder of this paper discusses the key phases in the design and operationalisation of these tools.

II.   Measuring own use provision of services / unpaid domestic and care work

A.   Own use provision of services / unpaid care and domestic work

9.   Own-use provision of services / UDCW encompasses a wide range of activities, including caring for children and ill, disabled, or aging relatives, routine housework (cooking, cleaning, shopping, maintaining the dwelling), minor household repairs and decoration, and household management. Around the world, women and girls perform the major share of this work. The existence - and persistence - of gender inequalities in the division of UDCW has long been argued to be both a cause and effect of women's political, economic, and social marginalisation, with the effect most pronounced for the poorest in society (lacking options to outsource activities to the market and/or invest in 'time-saving' domestic appliances).

10.   The 19th ICLS specifies the production of key statistical indicators (headcounts, participation rates, and volume measures) for own use provision of services, according to "activity cluster"[7]. The term "activity cluster" is not defined in the 19th ICLS but clarification is provided in a separate report[8]. Four separate activity clusters are classified under own use provision of goods are as follows[9].

(i) household accounting and management, purchasing and/or transporting goods

(ii) preparing and/or serving meals, household waste disposal and recycling

(iii) cleaning, decorating, and maintaining one's own dwelling or premises, durables and other goods, and gardening

(iv) childcare and instruction, transporting and caring for elderly, dependent or other household members and domestic animals or pets, etc.

11.   This is consistent with the wider UN conceptual basis for time use statistics on unpaid domestic and care work (e.g., as specified in SDG indicator 5.4.1, and the domain approach utilised in the International Classification of Activities for Time Use Statistics (ICATUS-16)). The "cluster" approach permits "direct care" activities (iv) to be distinguished from "indirect care" activities (i – iii) in the resulting indicators – in line with conventions in the wider research literature. Here, "direct" care work refers to active interactions between a care-provider and a care-recipient (such as feeding, bathing, dressing, providing medical care, accompanying places), as well as "passive" / "supervisory" responsibilities expressed in caregiver proximity and availability to intervene in case of need. "Indirect" care work refers to the provision of services (such as cooking, cleaning, laundry, household

---

[7]ICLS/19/2013; page 15, paragraph 74b
[8]ICLS/19/2013/2; Page 23, paragraphs 97 - 100
[9]ICLS/19/2013/2; page 15, paragraph 22c

maintenance and management) which underpin daily life, and form the pre-conditions for direct caregiving[10].

12. The concept of "passive" or "supervisory" care articulates dimensions of caregiving that tend to be most prone to under-reporting. Often taken-for-granted by respondents as a background constant, passive care is argued to correspond to a particular "state of mind"[11], expressed in being present, attentive, available, watchful[12]. Examples of passive care include "on call" supervisory responsibilities for children and/or ill, injured, frail, or differently abled adults[13]. Understood as a state of being, rather than of doing, time spent on passive care is especially prone to omission or misrepresentation when time use is expressed as a series of discrete activities.

13. The production of official statistics on own use provision of services / UDCW is a growing priority, with relevance for a wide range of public policy areas, especially when collected alongside data on labour force participation and employment. These data can directly inform the formulation and implementation of policies to increase women's labour market participation, to promote gender equality in the workplace, to address gender pay-gaps, decent work deficits, excessive work burdens, and time poverty. Such data may also reveal opportunities for investments in public infrastructure and basic services (e.g., piped water and sanitation, electricity, cooking fuel, public transportation), to directly reduce or redistribute UDCW.

14. Over the longer term, successive rounds of data can inform monitoring, evaluation and cost-benefit analysis of policy changes. The data may also inform valuation exercises, which can permit transitions from household provision of services to market-based provision to be monitored, with important implications for the interpretation of national GDP figures, and for international comparability of the same.

B.   Time use measurement: Available approaches

15. Time use surveys (TUS) are widely recognised as the main measurement approach for statistics on unpaid domestic and care work, but they can be extremely resource intensive to implement. Historically, this has deterred widespread take-up in national statistical programmes, particularly in low- and middle-income countries.

16. Two main approaches are available for survey-based time use measurement, termed "diary" and "stylised"[14]. A variety of different formats are available within each broad approach. In

---

[10] International Labour Organisation (2018:6), *Care work and care jobs for the future of decent work*, ILO, Geneva

[11] Budig M, Folbre N. "Activity, Proximity or Responsibility? Measuring Parental Childcare Time". In Folbre, N and Bittman, M, eds. Family Time: The Social Organization of Care. New York: Routledge; 2004. p. 51–68

[12] Folbre, Nancy. 2006. "Measuring Care: Gender, Empowerment, and the Care Economy.". *Journal of Human Development*, 7(2): 183–99.

[13] Carrasco and Serrano (2011) draw attention to the parallels with certain occupations, where a part – and sometimes the major part – of the working time is acknowledged to involve a state of being "on call", and is recognised, recorded, and remunerated indivisibly from more "active" aspects of the work. They offer the example of fire fighters. Cristina Carrasco & Mònica Serrano (2011) "Lights and Shadows of Household Satellite Accounts: The case of Catalonia, Spain". Feminist Economics, 17:2, 63-85, DOI: 10.1080/13545701.2011.573483

[14] Approaches used outside of survey contexts, such as immersive observation, experience sampling methods (ESM), and deployment of wearable technologies, are omitted from discussion as beyond scope.

recent years, 'hybrid' diary instruments, which combine aspects of each, have been the subject of growing interest. This section introduces the main features of these different instrument formats, highlighting the features of design and implementation selected for attention in the development of the ILO's modular LFS pilot instruments.

17. Diary approaches to time use measurement are characterised by the chronological reporting of time use over the 24 hours of a day, sometimes for multiple days. The respondent records (if self-administered) or reports (if interviewer-administered) how they spend (if contemporaneous) / spent (if retrospective) their time, from a designated moment, conventionally 4am or 12midnight, or from the moment that they wake/woke up. The exercise is sometimes repeated for multiple days. Within this broad approach, several diary formats have been developed. 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 divided into, usually equal15, intervals of between five minutes and 60 minutes, ready to be populated with the respondents' corresponding activity/ies.

18. Formats also vary according to whether activities are open-coded or pre-defined. In the former, activities are recorded verbatim, in the respondents' own words, subject to coding 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 pre-specified 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 "multi-tasking", i.e., activities undertaken simultaneously or over-lapping, and/or contextual information such as location, presence of others, 'beneficiary', remuneration, use of ICT, and/or affect.

19. In a stylised 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 either a seven-day week or a 24-hour day. While diary formats record the timing, sequencing, and duration of activities, stylised formats provide only participation and total duration. Stylised questions may be framed in terms of a specific reference period, e.g., "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, e.g., "How many hours a day (or a week) do you usually spend doing (activity X)?"16. The term "stylised" originated in this latter framing. The resulting time use estimates were "stylised" in that they referred to a hypothetical construct – the "typical" day, week, month, or other reference period – rather than a concrete, actual, reference period17. Contemporary usage of

---

15In some cases, longer intervals are assigned to night-time hours, during which most respondents are presumed to be sleeping, to condense the diary exercise.

16It is generally agreed that "typical" period questions place greater cognitive demands on respondents than do 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.

17Juster, F. Thomas., Ono, Hiromi., and Stafford Frank P. (2003) "An Assessment of Alternative Measures of Time Use", Sociological Methodology, 2003;33: pp. 19-54, Available from: https://www.jstor.org/stable/1519852

the term "stylised" has expanded to incorporate "actual" alongside "usual" framings. The distinctiveness of the approach is instead defined in opposition to the diary format[18].

20. Stylised approaches vary in breadth of coverage. At one end of this spectrum are "stylised analogues of time diaries"[19]. The scope of activities covered in such "stylised analogues" is deliberately comprehensive, the goal being to capture – at varying levels of detail – all activities the respondent performed for a given reference period[20]. The comprehensive list of activities included in stylised analogues permit a level of activity coverage comparable to diary approaches. At the other end of the spectrum are short question series', which forego a \full accounting of respondents' time use, to restrict investigation to a limited number of activity classes.

21. Hybrid diary instruments include direct question items, familiar from 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 which 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, utilising a longer reference period to supplement a core diary, to good effect.

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

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

---

[18]Terminology relating to time-use measurement instruments remains somewhat unstandardised. Within the peer-review literature, stylised approaches are sometimes restricted to investigations of "usual" or "typical" time use and sometimes applied expansively. Practice varies considerably. The most recent international guidelines (UN 2005: 15) applies the expansive definition (i.e., encompassing "actual" and "usual").
[19]See UN 2005:15, Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid Work, Department of Economic and Social Affairs, Statistics Division, ST/ESA/STAT/SER/93, United Nations, New York
[20]See UN 2005:16, Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid Work, Department of Economic and Social Affairs, Statistics Division, ST/ESA/STAT/SER/93, United Nations, New York

C. Time use measurement: Strengths and limitations of available approaches

24. Diary instruments are hypothesised to map more closely onto autobiographical memory. The "chronological reporting" feature, combined with the proximity and concreteness of the reference period, are believed to limit scope for measurement error. Data generated by stylised approaches have been found to be more susceptible to error, and thus less accurate, than diary-based data21. This is largely attributed to the greater cognitive burden imposed as respondents are obliged to recall and aggregate numerous discreet instances of a particular activity class, often over a longer period than that required by diary instruments22. There is evidence that aggregation error is exacerbated when distinct activities are collapsed as higher-level domains (whereby, for instance, a single question asks about time spent on routine housework in place of separate questions on constitutive tasks, such as cooking, serving food, cleaning dishes, etc.)

25. Providing further support for the hypothesis that diary instruments impose lesser cognitive burdens, the sequential recall of activities characterising diary approaches has been found to result in shorter duration interviews relative to short stylised sequences in some settings. As a result, diary-based estimates have historically served as a benchmark against which to appraise the performance of stylised instruments23. Concordance studies comparing diary and stylised time-use estimates have produced several common findings.

26. Firstly, with few exceptions24, stylised estimates have been found to over-estimate time spent on UDCW (variously defined) relative to diary estimates. Several explanations have been put forward to account for this tendency, including random error, failures of respondent recall, confusion about the activities constitutive of UDCW, or the specified reference period, and double counting of simultaneous activities.

27. There is evidence that stylised approaches encourage respondents to produce exaggerated "usual time" estimates, even when a narrow reference period is specified, owing to the greater cognitive burdens imposed by a genuine 'recall and sum' strategy - resulting in a higher level of measurement error, in comparison with diary approaches. In comparison to other activity domains, UDCW has been shown to be particularly susceptible to reporting error when stylised questions are used. This is because such work tends to be undertaken at irregular intervals, which vary in duration, and in the case of indirect care work (i.e., care work that does not involve direct, relational, or interactive caregiving), its timing is "discretionary", to the extent that it is not typically subject to externally imposed schedules25.

21e.g., Bianchi, Suzanne M., Melissa A. Milkie, Liana C. Sayer, and John P. Robinson. (2000). "Is Anyone Doing Housework? Trends in the Gender Division of Household Labor." Social Forces 79:191-228

22Schwarz, Norbert (2007): "Retrospective and Concurrent Self-Reports: The Rationale for Real-Time Data Capture". In: Stone, Arthur; Saul Shiffman, Audie Atienza, and Linda Nebeling (Eds.) (2007). The science of real-time data capture: Self-reports in health research. New York: Oxford University Press

23E.g., Press and Townsley 1998, Robinson and Godbey 1997, Marini and Shelton 1993, Niemi 1993

24Bonke (2005), using data for Denmark, finds the reverse to be true, with stylised estimates underestimating time spent on unpaid domestic work, relative to diary estimates. Notably, this study is one of only a few to compare stylised and diary estimates for the same individuals. This is not the case in a majority of studies.

25E.g., Kitterod, Ragni Hege., Lyngstad, Torkild Hovde., (2005) "Diary Versus Questionnaire Information on Time Spent on Housework? The Case of Norway" Electronic International Journal of Time Use Research

28. Stylised data on UDCW time are also widely perceived to be more vulnerable to measurement error originating in social desirability bias than are diary data26, with the weight and direction of its effects varying dependent on the social context27. Social desirability bias is often invoked to explain why stylised estimates are inflated for certain population groups but not for others. Though, again, the empirical evidence is mixed. This relates to a second common, though not universal28, finding in the comparative literature, whereby the inflation of stylised estimates relative to diary estimates is found to vary with respondents' characteristics, introducing systematic error.

29. A number of studies have found that women's stylised estimates inflate UDCW time to a greater extent than men's, and parents to a greater extent than non-parents, though the weight of the effect varies. The gap widens further with greater irregularity in UDCW hours, and with the amount of UDCW reported as a secondary activity in the benchmark diary data29. There is also some evidence that the reporting gap varies by age group or cohort, with stylised data producing equivalent or near-to-equivalent estimates for some age groups but not others30. Overestimation bias (whereby the duration of more preferred activities is over-estimated, and the duration of less preferred activities under-estimated) has similarly been found to undermine the validity of stylised time use estimates in some settings31.

30. While the existence of a reporting gap across TU measurement methods is widely accepted in the methodological literature, the explanations for it are still the subject of debate. Another seam in the methodological literature has questioned the interpretation that stylised approaches inflated estimates of UDCW time are evidence of upwards bias in stylised time use measures, relative to diaries. It has been suggested that a positive gap between stylised and diary estimates of UDCW time may be explained, at least in part, by the latter's under-estimation rather than the former's over-estimation. There is evidence, too, that a majority of the difference in diary and stylised estimates may be attributable to random rather than systematic error, with both types of estimate prone to large random errors32.

31. Diary data are not error-free, they may exhibit recall or reporting error, and allocated "diary days" may, by chance, be unrepresentative of respondent's normal time use. Diary-based

---

26whereby respondents over-estimate time spent on socially desirable activities and omit or under-estimate time spent on converse activities.

27 Press and Townsley (1998) attribute over-reporting of UCDW time to respondents' attitudes to gendered divisions of labour and wider social norms concerning the same. Press, Julie E., and Eleanor Townsley. (1998). "Wives' and Husbands' Housework Reporting: Gender, Class, and Social Desirability" Gender and Society 12:188-218

28For instance, Baxter and Bittman (1995) find the gap between stylised and diary estimates is similar across sub-groups, meaning stylised estimates, though biased, can still support ordinal scaling. Baxter, I, and Michael Bittman. (1995). "Measuring Time Spent on Housework: A Comparison of Two Approaches." Australian Journal of Social Research 1:21-46

29See Kan 2008 for a review, Kan, Man Yee., (2008) "Measuring Housework Participation: The Gap between "Stylized" Questionnaire Estimates and Diary-based Estimates." Social Indicators Research 86:381-400

30For instance, Kitterød et al (2005) find a high level of concordance between the stylised and diary estimates overall, and when broken down by sex. Kitterod, Ragni Hege., Lyngstad, Torkild Hovde., (2005) "Diary Versus Questionnaire Information on Time Spent on Housework? The Case of Norway" Electronic International Journal of Time Use Research

31Otterbach, S., & Sousa-Poza, A. (2010). How accurate are German work-time data? A comparison of time-diary reports and stylized estimates. Social indicators research, 97(3), 325-339

32Kan, Man Yee., Pudney, Stephen (2008) "Measurement Error in Stylized and Diary Data on Time Use" Sociological Methodology, 2008;38, pp. 101-132; http://www.jstor.com/stable/20451151 JSTOR

estimates of UDCW time have been shown to be sensitive to instrument design, coding, and data analysis decisions. Diary-based estimates of UDCW time have been shown to be particularly sensitive to different treatments of temporal sequencing and simultaneity, with a tendency for respondents to relegate UDCW tasks to the status of a secondary task when performed alongside other activities and/or for short-duration activities to be omitted. The latter tendency will be exacerbated by longer (pre-specified) episode slots.

32. When diaries permit only a "main" or "primary" activity performed in each episode to be recorded, UDCW may be under-identified. Where respondents are permitted or prompted to report simultaneous activities (and when simultaneous UDCW time is included in the subsequent analysis), the gap between diary and stylised estimates of UDCW time narrows. One hypothesis is that stylised questions prompt respondents to include both simultaneous and isolated UDCW time in their summed estimates (which also accounts for the tendency for summed activity time to exceed the 24 hours of the day / 168 hours of the week), whereas diary estimates of UDCW time skew towards the latter unless secondary activities are actively probed for. A reporting bias towards time spent on "activities", at the expense of time occupied with responsibilities (such as passive child or elder care), has similarly been hypothesised to obscure 'true' time spent on UDCW.

33. Differences in diary and stylised instruments' treatment of short breaks or "interruption tasks" has also been implicated as a reason for the former's lower estimates of UDCW time relative to the latter. Where interruptions in the continuity of UDCW tasks (e.g., short rest breaks) may be separately coded in diaries (so inflating leisure or self-care values), respondents tend to ignore such breaks or interruptions when producing stylised estimates of their UDCW time. The latter treatment accords with guidelines for estimating paid work time, in which such short breaks and interruptions are included in paid work time[33]. The inclusion of a 'secondary activity/ies' field in diary instruments has been shown to reduce the reporting of short-term, interruption activities, which are otherwise recorded as the "main" or "primary" activity[34].

34. The empirical evidence base is also limited by its geographical concentration. Most studies to assess concordance of stylised and diary generated data have utilised data for a limited number of high-income countries, primarily North America, Australia, and Europe.

35. Rather than assigning one class of instruments superiority or benchmark status, the existing literature provides insights into possible explanatory mechanisms for divergent estimates, providing a basis for improving the power and performance of both approaches[35].

III.   Priority topics for piloting

36. The pilot time use measurement tools were developed to permit the evaluation of alternative measurement approaches designed to reduce the response burden and resource intensity of traditional "full diary" approaches, while maintaining data quality. Piloting focussed on

---

[33]See Rydenstam, Klas (2001): Paid and Unpaid Work. The Apples and Pears of Time Use Statistics. Available from http://www.ssb.no/english/about_ssb /conference/iatur _2001/rydenstam.pdf

[34]Kitterod, Ragni Hege., Lyngstad, Torkild Hovde., (2005) "Diary Versus Questionnaire Information on Time Spent on Housework? The Case of Norway" Electronic International Journal of Time Use Research

[35]See Schulz & Grunow (2012), who argue for closer attention to the reasons why estimates "differ so markedly", Schulz, F., & Grunow, D. (2012). Comparing diary and survey estimates on time use. European Sociological Review, 28(5), 622. https ://doi.org/10.1093/esr/jcr03 0

several key evidence gaps in modular time use measurement, with the aim of refining the questionnaire design as well as informing key aspects of survey administration / fielding:

37. In relation to survey fielding, the workability of alternative schemes to minimise non-response was a key site of interest. The twin requirements for direct reporting of time use (i.e., prohibition of proxy response), and pre-assignment of reporting day (to avoid biased estimates of population-level time use) presents a challenge to maintaining sufficiently high response rates. This presents risks for data quality, via the introduction or inflation of nonresponse bias. In a modular design, this may risk undermining response rates for the parent survey in addition to the time use module.

38. 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". The random assignment of respondents to designated diary days directly conditions the survey participation day (i.e., the day immediately following the diary day). Sample units assigned to report on Monday's time use must be surveyed on Tuesday, those assigned to report on Tuesday must be surveyed on Wednesday, and so on.

39. While it is relatively straightforward to extend the sample design to obtain a probability sample of days of the week (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 sampled individuals will be unavailable, unable, or unwilling to participate in the survey on their assigned day.

40. Various strategies have been proposed to reduce the challenges imposed by the designated diary day feature. Some strategies retain the feature in modified form, while others relax or even dispense with it completely. Each strategy involves trade-offs in exposure to selection bias, measurement error, and the complexity and costs of field operations. Two strategies were selected for pilot testing. The first combined postponement of up to seven days (maintaining the original day of week assigned), with some restricted substitution (substitute days were probabilistically assigned at the sampling stage, when sampled households were probabilistically assigned to days of the week. The second strategy commenced identically, with households probabilistically assigned to an initial day of the week, but more flexibility was permitted at the pending recovery stage, meaning interviews could be rescheduled flexibly. In both cases, interviewers were directed to make an initial appointment and to schedule the timing of interviews on the designated day around the respondent's availability. The sample was randomly assigned to one of the two pending recovery conditions.

41. In relation to the module design, a key area of focus was the measurement of simultaneity / multi-tasking, and the performance of alternative measurement strategies to reduce known under-reporting of unpaid domestic and care work (via a mix of context-based items, items to record simultaneity, and dedicated probing / recovery items specifically targeting passive care time).

42. In addition, gaps and/or redundancies in the pre-coded listing and contextual items at the data input and analysis phases were a focus for the refinement of the tools, more broadly, as was the performance of the 15-minute fixed episode approach to recording the timing and duration of activities in the hybrid light diary tools. More generally still, respondents comfort with "clock time" and the fungibility of alternative temporal frameworks was explored.

43. The pilot strategy proceeded from intensive qualitative research (rapid ethnographic assessment and cognitive interviewing) to randomised survey experiments (utilising a

between subject design). The separate study phases were intended to inform and evaluate key decision points in the design of the different module formats, the enumerator guidelines, and broader features of sample design and field operations. Piloting was undertaken in India (Haryana state) and Lesotho. For the survey experiment phases in each country, selected geographies were purposively selected and then stratified by urban, rural, and "remote" locations. Within strata, primary sampling units, households, and individual respondents were purposively selected. A balanced design was adopted, with randomisation to experimental arm at the level of PSU (50% of PSUs allocated to each arm).

IV.   Instrument development

44. Based on a review of country practices, and taking account of the existing evidence base, three alternative time use measurement modules were developed for testing. Two modules adopted the hybrid light diary format – with variability limited to their treatment of simultaneity. The third module was a stylised diary analogue. All tools underwent several phases of refinement, across the different phases of piloting.

45. All three pilot time use modules were designed for face-to-face interviewer administration. A CAPI (computer aided personal interviewing) tool was developed in CS Pro software. Use of a CAPI, together with the pre-coding of activities and the use of fixed episodes of time, results in a much faster turn-around time from data collection to analysis and dissemination. It also has potential to minimise data entry error. The pilot time use modules included several features to reduce erroneous or missing data fields (including restrictions on data entry format, automated updates and warning flags activated for incongruous entries). The pilot time use modules were embedded within a wider pilot survey reproducing standard LFS items, was intended to contextualise the questions contained in the TU modules for respondents and support meaningful analysis of TU module performance within a national LFS. Table 1 summarises the key features of the time use modules developed for testing.

Table 1:
Key features of the pilot time use modules

Feature	Summary details
Mode (all three instruments)	Face-to-face interviewer administered
Administration (all three instruments)	CAPI, CS Pro
Reference period (all three instruments)	One 24-hour day, "yesterday" from 04:00 to 04:00
Reference day assignment (all three instruments)	Probabilistic assignment of sampled households to days of week in advance
Pending recovery strategy (Experimental group one)	Postponement (+7 days) or substitution (pre-specified substitution day)
Pending recovery strategy (Experimental group two)	Convenience-based substitution (no specified substitution day, but day after designated day barred)
Respondent sampling (within household)	One eligible couple-dyad (probabilistically sampled in households with more than one eligible couple (Country one) Or Probabilistic sampling of one household member (Country two)
Activity timing (light diary instruments 1&2)	Fixed episodes (96 x 15-minute episodes)

Feature	Summary details
Activity coding (light diary instruments 1&2)	Respondent narrates diary day verbatim. Interviewer selects from 50 pre-coded activities + "other, specify" (aligned to the UN International Classification of Activities for Time Use Statistics 2016 (ICATUS-16) coding scheme)
Activity coding (stylised instrument)	27 direct question items on participation in and (conditional on participation) volume of time spent on select activities (aligned to ICATUS-16
Contextual items (light diary instruments 1&2)	Five conditionally activated contextual items are targeted for pilot testing (location, co-presence, beneficiary, job linkage, income generation)
Treatment of supervisory / passive care time (light diary instruments 1&2)	Dedicated recovery series activated on completion of the diary day. Separate items target supervision / passive care of children and adults
Treatment of simultaneity / multi-tasking (light diary instrument 2 only)	Multiple activities recorded as occurring simultaneously (i.e., during one or more of the same 15-minute episodes).

46. In the case of the hybrid light diary instruments, the time use modules record everything the respondent reported that s/he did on her/his "diary day". The "diary day" is divided into 96 consecutive and non-overlapping fixed episodes of 15-minutes, from 4am on the day preceding the interview until 4am on the day of the interview.

47. The initiation of the diary day at 4am is a standard convention in time use research (an alternative convention is to begin at 12midnight). Generally – but not always – respondents are asleep at 4am so the diary catches the start of the waking day. The time use module begins by asking respondents to recall what they were doing at 4am. Once the activity is recorded (selected from the drop-down 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 end-time entered for the prior activity. The end-time of the activity is recorded to the nearest 15-minute interval on a drop-down menu, which automatically updates to exclude episodes prior to each activity start time. The adoption of fixed episodes in place of open episodes represents a trade-off between different aims. A fixed episode approach limits scope to record activities lasting less time than the episode (or exaggerates their duration if recorded). However, it dispenses with the need for interviewers to manually enter end times in hours and minutes (removing an important potential for interviewer error) and simplifies data cleaning and analysis considerably.

48. These pilot time use modules are designed to comprehensively record respondents' time use for one 24-hour hour day. The modules are programmed with a drop-down menu of 50 pre-coded activities, along with an option for "other, specify" (Code 51) featuring an open field to record activities that do not "fit" into one of the 50 specified codes. Codes are aligned to the UN International Classification of Activities for Time Use Statistics – 2016 (ICATUS-16) at the two or three-digit level, and all nine of the ICATUS major divisions are represented (ICATUS-16 is a hierarchically organised classification scheme with 165 groups classified into 56 divisions and 9 major divisions). 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 under-counting (such as unpaid domestic work and unpaid care work, and fewer codes assigned to domains less prone to measurement error and/or of lower substantive priority for the measurement objectives, where less disaggregation is demanded at the output stage.

49. For purposes of comparison, one of the two pilot diary instruments permitted only a single activity to be recorded for each fixed episode. The other 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 multi-tasking to be recorded, given the large body of pre-existing research indicating this strategy supports reporting of otherwise unacknowledged unpaid care work. For instance, the respondent may report that while doing paid work or unpaid housework (e.g., cleaning, cooking, doing 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 that they were grazing livestock while weeding an adjacent field or harvesting fruit in a nearby orchard. Or that they were cooking an evening meal, while cleaning up the kitchen, while supervising a child to make sure s/he is doing on her/his homework. Interviewer training highlighted 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. The trade-offs in information gain versus respondent burden, interview duration, and the complexity of the resulting data were a priority area for comparison.

50. In addition to the record of daily activities, "contextual information" is recorded to support correct assignment of 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.,). There is evidence that contextual items may also aide respondent recall, particularly items probing location and co-presence of others. The pilot time use module included five conditional context items (the final instruments is expected to have fewer contextual variables, based on the results of piloting). Context items are activated only when a coded activity is eligible for classification to multiple higher-level domains.

51. The five conditional context items included in the pilot instruments are: location[36]; co-presence / with whom; for whom / beneficiary; job linkage; income generation

52. The activation of the contextual items is tailored to the activity reported. For time spent sleeping, only location is recorded (the omission of the "with whom" item for sleep time may result in an under-estimation of passive care time, however, the question can be received as intrusive and undermine the interview interaction at an early stage. The recovery series on passive care time does permit passive care time to be recorded during times when the respondent was asleep). For all other activities, "location" and "with whom" are recorded. The contextual items "for whom", "job linkage", and "income generation" activate only when required to correctly classify activities to higher-level domains.

53. Common to both hybrid light diary tools, a "supervisory" or "passive" care recovery series is designed to correct for known under-reporting of supervisory or background / on-call unpaid caring responsibilities. This series prompts respondents to recall time when s/he was minding, watching over, or supervising "dependents", but not directly interacting with them. The passive care recovery series asks first about whether and if so when the respondent had supervisory / passive care responsibilities for children household or family members (anyone aged under 17 and younger). These questions are then repeated with reference to adult household or family members (aged 18 and over) who require assistance or help from others

---

[36]The *location* item includes a quality control feature to limit under-reporting 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.

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 day has been fully filled out. A separate roster permits the timing, sequencing, and duration of supervisory / passive care episodes to be recorded in 15-minute fixed episodes, along with the corresponding diary activity/ies.

54. Finally, a short quality audit series asks respondents to confirm – and revise if desired – the timings of key points throughout the diary day (wake up time / bedtime / mealtimes).

55. In the case of the stylised time use module, a diary analogue approach was adopted. This means that instrument covered – at varying levels of detail – all major "activity domains". A total of 27 activities were listed to develop an accounting of respondents' time use for the reference day. The intention was to assess concordance with the diary tool across activity domains, while also minimising risk for social desirability bias inherent in asking about time spent on a single domain of interest. The stylised module asked respondents if they had participated in each activity in turn during the preceding day, and – if so – how long they had spent on it. Duration was recorded in minutes (if less than 60 minutes) or hours and minutes (if more than 60 minutes). Contextual items (excepting that specified in the wording of the activity) and simultaneity are not compatible with the stylised approach, marking a key site of divergence from the diary modules. Additionally, only participation and volume measures are recorded, since sequencing, frequency, and timing of activities are not compatible with the stylised format. Effort was made to separately account for active and passive care time in the question wording.

56. Common across both the stylised and light hybrid diary tools, a short "typical day" series asks respondents to report whether the diary day reported in the survey is unusual in any way. This series permits any apparent anomalies in the resulting data to be contextualised. Similarly, all respondents are asked what the time is at the commencement of the individual interview as a means to assess their familiarity and ease with clock time. The interviewer records the response as well as the means by which the respondent told the time (e.g., by reference to a watch, clock, mobile phone or electronic device, an external schedule or timetable, the position of the sun, etc.,).

V.   Conclusions

57. Current international standards set out a coherent and inclusive framework for the measurement of own use provision of services / unpaid domestic and care work. Not only are these standards harmonised to the system or national accounts, they are also consistent with a subsequent, major advance in the statistical measurement of time use, the International Classification of Activities for Time Use Statistics (ICATUS 2016)[37] adopted by the UN Statistical Council in 2017. ICATUS 2016 provides a standardised framework for the international harmonisation of time use statistics via a detailed hierarchical activity classification scheme.

58. Time-use surveys are the main data source for statistics on own use provision of services. Such surveys are highly specialised and present distinctive measurement challenges, not routinely encountered in household sample surveys.

---

[37]UNSD (2019) International Classification of Activities for Time Use Statistics 2016, Department of Economic and Social Affairs, Statistics Division, Statistical Papers, Series M: No 98, United Nations, New York

59. The ILO's programme of work to develop and test a range of time use measurement modules, suitable for integration into national labour force surveys in LMICs is not designed to be prescriptive. For that reason, two separate instrument formats – the hybrid light diary and the stylised questionnaire have been developed, tested, and refined on an iterative basis. Similarly, different approaches to field operations, compatible with different approaches to modular attachment and survey administration have been tested. The objective is to provide a toolkit to inform country practices, and national adaptation.

60. Analysis of the pilot data is still ongoing, but this study is expected to result in important contributions to emerging best practice on modular time use measurement.