# ILO LFS PILOT STUDIES EXPERIMENTAL FIELD TESTS:
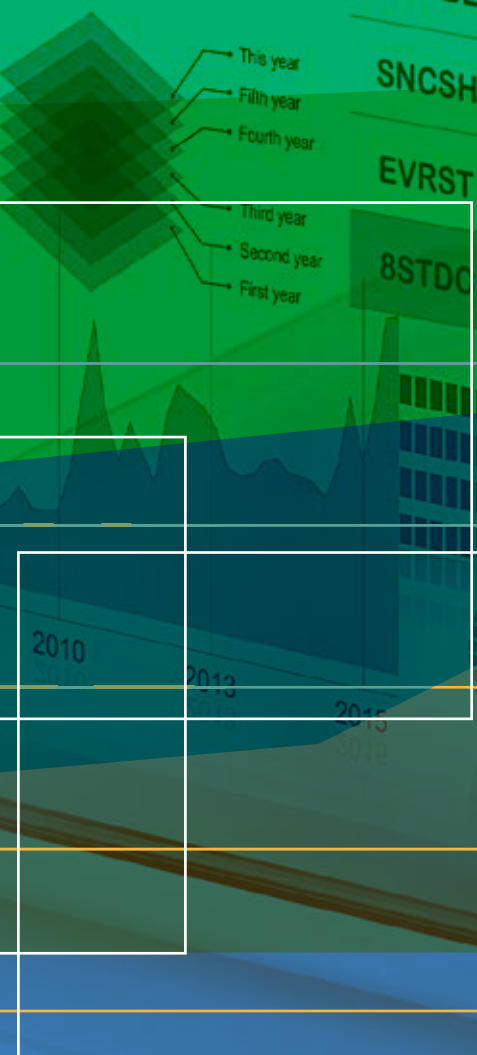
Methodology, process and outcomes

**Elisa M. Benes, Kieran Walsh**

**May 2018**

International Labour Organization

STATISTICS
Department
of Statistics

# ILO LFS pilot studies - experimental field tests:

## Methodology, process and outcomes

---

---

---

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACRONYMS

| | |
|---|---|
| **ICLS** | International Conference of Labour Statisticians |
| **LFS** | Labour Force Survey |
| **CI** | Cognitive interview |
| **NSO** | National Statistical Office |
| **PSU** | Primary Sampling Unit |
| **CSPro** | Census and Survey Processing System |

# *ACKNOWLEDGEMENT*

# I.  BACKGROUND

1. The latest international recommendations on how to measure some of the key headline labour market indicators, including the labour force participation rate, employment-to-population ratio and unemployment rate, are contained in the *Resolution I concerning statistics of work, employment and labour underutilization* adopted in 2013 by the 19th International Conference of Labour Statisticians (ICLS). These standards introduced a number of important advances that will impact the way work and labour force statistics are collected and disseminated by countries around the world in the years to come. To support their wide implementation, the 19th ICLS called on the ILO to "*conduct further conceptual and methodological work including testing*" and develop "*technical manuals and model data collection instruments*" aligned with the latest standards (ILO, 2013).

2. As follow-up, in 2015, the ILO launched a global project of labour force survey (LFS) pilot studies. The Project had as main aim to develop and test alternative survey questionnaires to collect statistics on high priority topics such as employment, labour underutilization (comprising time-related underemployment, unemployment and the potential labour force), and own-use production work, in line with the 19th ICLS standards. The ultimate objective is to develop evidence-based guidance to support countries in adopting the new standards.

3. The background to the studies, measurement objectives and overall methodological approach are described in a separate report in this series. The cognitive testing phase, along with findings on various specific topics are also the subject of dedicated reports in this series. The focus of this report is to provide a more detailed description of the field testing stage of the process including the design, implementation and lessons learned. All reports can be found on a dedicated page on the ILO website.[1]

---

[1] See: http://www.ilo.org/stat/Areasofwork/Standards/lfs/WCMS_484981/lang--en/index.htm

# II. FIELD TEST METHODOLOGY

4. As described in the report covering the cognitive testing phase the five initially developed model questionnaires were updated based on the findings from the cognitive interviews (CI). Field tests were then completed using the updated questionnaires applying the methodology described below.

5. A purposive sample design was used for the field tests. The target design involved a relative oversample of rural versus urban households in a ratio of 60:40. Furthermore countries were advised, where possible, to use four strata (rural without market, rural with market, peri-urban and other urban) reflecting degree of urbanisation in their sample selection. This sample design reflected the desire to ensure questions were adequately tested with populations where the greatest impact of the changes in concepts was expected, e.g. those in rural areas with limited access to markets as well as those where mixed agricultural production (for sale and for own final use) could be prevalent. This created an additional challenge for the design of the test questionnaires given the expectation of relatively lower education levels in those settings. This purposive approach had the main objective of yielding sufficient number of respondents in situations that posed the greatest challenges for the implementation of the new standards. It was not meant to yield results that would be generalizable to the wider population, but rather inform the design of questionnaires to enable their use in such contexts. At the same time, to assess the adequacy of the questionnaires to more urban contexts, a reasonable proportion of urban sample (40%) was envisaged in the design.

6. With regards to target areas, countries were advised to focus on a limited number of districts or regions where the proposed sample design could be achieved and taking into account the prevalence of small-scale farming and/or fishing activities.

7. A total target sample size of 400 households per questionnaire was set. It was proposed to use a split sample design whereby each questionnaire was administered to a set of households with the same sample features. This was chosen to minimise incomparability arising from differences in the achieved samples. The proposed sample design is illustrated in Table 1. However, countries had some flexibility in implementing the design to reflect differences in information available on sampling frames across countries, size of PSUs etc.

**Table 1.** Illustration of proposed sample design for LFS field tests (per questionnaire)

| Strata name | Urban (40%) | | Rural (60%) | | TOTAL |
| | A | B | C | D | |
| | Other urban areas | Peri-urban Areas | Rural with market | Rural without market | |
| PSU | 6 | 10 | 12 | 12 | 40 |
| Total # of hh | 60 | 100 | 120 | 120 | 400 |

8. Field tests were organised over two waves, with each sampled household being visited twice at different times is the year. This was designed to allow some comparison of outcomes at different periods of activity. Specifically, countries were asked, to the extent possible, to pick two different points of the agricultural cycle to assess stability of measurement (e.g. first interview during the planting season and second interview during the harvesting season).

9. Bearing in mind likely non-response and attrition between wave 1 and wave 2 of field collection, countries were asked to oversample in wave 1 so that at least 400 households per questionnaire would be achieved in wave 2. In deciding on the appropriate number of wave 1 households countries took into account typical national non-response rates in similar household surveys. Substitution of households within the same PSU was allowed in wave 1 but not wave 2.

10. In addition to the assigned model questionnaires and proposed sample design, countries were given guidance on a variety of related implementation issues. For example, countries were advised to plan for multiple visits to selected households to minimise the level of proxy response. In addition, countries were advised to assign separate teams to each of the questionnaires. This strategy was meant to support the training of interviewers and supervisors so that they would become closely familiar with a single model approach and to limit possible contamination between the model questionnaires due to interviewer practices. Beyond these specific instructions, much of the survey implementation followed national practices for the LFS or other similar household surveys.

11. To assure the methodology was applied as intended a substantial technical support programme was implemented as part of the pilot study process. This was provided through a mix of documentation, remote support and missions to provide technical assistance and training.

12. In advance of the first wave of the field test in each country a mission was organised to provide training and support for the field implementation process. Typically, this involved providing training and support for final preparations during the week before the field tests commenced and some supervision and support during the early days of the interviews. As considered necessary, missions were also organised in advance of or during wave 2 field tests, but in all cases ongoing support was provided remotely. The observation of interviews in the field provided an invaluable source of insight to the ILO and offered the opportunity for intensive interaction between the ILO and the implementing NSOs. The missions and ongoing interactions have added to the quantitative evidence yielded by the field tests in generating the conclusions presented in the other reports of the pilot study project.

13. In addition to providing support through missions and remote technical assistance, the ILO provided documentation and tools to assist countries. In particular, interviewer manuals were provided, which were nationally adapted as needed. The ILO also provided a data entry tool developed in CSPro for countries who did not have the facility to develop their own dedicated data entry tool in the project timelines.

14. The remainder of this report is dedicated to discussing the actual implementation of the field tests and reflections of the ILO on the process.

# III. IMPLEMENTATION OF THE FIELD TESTS

15. The ILO gathered information on field tests through a metadata template which was completed by the countries. This metadata template covered a variety of aspects of the implementation of the field tests including:

    a. Study description – covering basic information such as sample districts selected, period of the agricultural cycle covered by each wave;
    b. Basic concepts – such as the definition of household applied, definition of household member used etc.;
    c. Sample design – description of strata used and achieved sampled;
    d. Field operations – size and number of teams used etc.;
    e. Data collection – dates, mode, languages used etc. ;
    f. Data entry and processing – description of system and process involved;
    g. Summary assessment – a summary comment by the NSO on the quality of the data collected indicating any issues of concern etc.

16. In addition to the metadata template, supplementary information on the implementation of the field tests was collected in a variety of ways. Countries were asked to collect information from interviewers on their experience of using the questionnaires, e.g. noting any areas of difficulty for respondents or interviewers. This information was shared with the ILO and further supplemented by observation of ILO staff through technical assistance missions and regular ongoing contact between ILO and the partner countries. The full set of information was an invaluable resource in assessing the implementation of the field tests but also provided excellent qualitative information during the analysis phase.

17. The information on implementation, along with initial conclusions from the pilot studies, were shared with the countries at a validation workshop which was held in Geneva during November 2016. Detailed discussions were held to inform subsequent analysis on the different topics covered by the pilot study and provide further feedback as needed. As regards the implementation process, it can be noted that countries were satisfied overall with the process and the quality of the information collected. They also reported positive feedback from interviewers on the experience of using the questionnaires in the field.

18. The field test process and the profile of the achieved samples are discussed in the following sections. The purpose of presenting this information is to provide context for the analysis presented in other reports in the series by giving a sense of the sample of households and individuals upon which the analysis is built.

## A.    Timeline of main stages of the field tests

19. The cognitive testing phase was completed in October 2015. The field tests commenced in countries from November 2015 and continued until October 2016. Timing of field work varied across countries reflecting various issues such as timing of the agricultural cycle, timelines needed to complete contracting processes and receive funding, etc. The period of field collection for waves 1 and 2 in each country are presented in Table 2 in broad order of the dates when wave 1 interviews

were conducted. The duration of fieldwork was typically 3 to 4 weeks but was shorter or longer in a limited number of cases due to national circumstances. As can be seen on Table 2, only wave 1 could be completed in 3 of the pilot countries due to unexpected difficulties with availability of funding to support implementation.

**Table 2.** Field test implementation dates

| Country | Wave 1 | Wave 2 |
|---|---|---|
| Vietnam | Nov 2015 | Mar – Apr 2016 |
| Kyrgyzstan | Nov 2015 | Apr 2016 |
| Moldova | Nov – Dec 2015 | Mar – Apr 2016 |
| Philippines | Nov – Dec 2015 | May 2016 |
| Peru | Nov – Dec 2015 | Jun 2016 |
| Namibia | Feb – Mar 2016 | May – Jun 2016 |
| Tunisia | Feb – Mar 2016 | N/A |
| Ivory Coast | Apr – May 2016 | N/A |
| Ecuador | May 2016 | Jul 2016 |
| Cameroon | Sep - Oct 2016 | N/A |

20. Through the metadata template completed by countries it was possible to observe differences in implementation in practices across countries (see Table 3). For example, countries used different size teams, had different ratios of supervisors to interviewers and there were differences in duration of training, etc. These differences depended on country resources and context and would not be expected to influence analysis in any meaningful way. Nevertheless, in all cases, all interviewers had prior survey experience and tended to have worked directly with the national LFS. For the most part, the low supervisor-interviewer ratios allowed for close follow-up and quality assurance of the completed interviews. Likewise, except for in Moldova, where experienced LFS interviewers - permanent staff of the NSO - were used, all pilot countries organized a multi-day training to cover the pilot survey contents, question-by-question, and highlight key operational guidance.

**Table 3.** Characteristics of field teams for ILO pilot studies

| Pilot country | Team composition* | | | Personnel type | Background | Training duration (days) |
|---|---|---|---|---|---|---|
| | No. supervisors | No. interviewers | S : I ratio | | | |
| Cameroon | 3 | 15 | 1 : 3 | Short-term & permanent staff | Prior survey experience | 6 |
| Ecuador | 2 | 6 | 1 : 3 | Short-term staff | Prior survey experience | 4 |
| Ivory Coast | 4 | 12 | 1 : 3 | Short-term staff | Prior survey experience | 6 |
| Kyrgyzstan | 1 | 5 | 1 : 5 | Short-term & permanent staff | Prior survey experience | 2 |
| Moldova | 2 | 15 | 2 : 15 | Permanent staff | LFS interviewers | 1 |
| Namibia | 3 | 12 | 1 : 4 | Short-term staff | Prior survey experience | 5 |
| Peru | 4 | 12/13 | 1 : 3/4 | Short-term staff | Prior survey experience | 6 |
| Philippines | 2 | 8 | 1 : 4 | Short-term staff | LFS interviewers | 3 |
| Tunisia | 2 | 6 | 1 : 3 | Short-term staff | Prior survey experience | 3 |
| Vietnam | 3 | 6/7 | 1 : 2/3 | Short-term staff | Prior survey experience | 3 |

*per model questionnaire

21. The difference of greatest practical importance was the requirement to use different languages across countries (see Table 4). The ILO provided materials in English, French, Russian and Spanish. For many countries these materials had to then be translated to their national language or, depending on the region of the fieldwork, another local language. In some cases, it was necessary to translate to multiple national languages (Philippines, Moldova, Kyrgyzstan) and in further cases some live translation to local languages was also required (Namibia and Ecuador). Live translation in particular required additional training and practice for interviewers as they were required to work with a paper version in one language but interact with the respondent in the local language. This created risks of inconsistency which were mitigated to the extent possible through the training process.

22. Table 4 also shows the regions where the testing took place. All countries except Moldova selected between 1 and 3 regions or districts for interviewing. The limited number of regions covered in combination with the purposive samples of limited size highlights that the samples could not be used to generate representative estimates, either at region or state level.

**Table 4.** Languages used by country and location of field testing

| Country | Language(s) of questionnaire | Live translation | Region of testing |
|---|---|---|---|
| Cameroon | French | N/A | Yaoundé, Mkometou and Nkolmekok |
| Ecuador | Spanish | Kichwa | Imbabura and Esmerelda |
| Ivory Coast | French | N/A | Abengourou, Agnibilékro and Béttié |
| Kyrgyzstan | Kyrgyz, Russian | N/A | Oshskaya and Chuyskaya |
| Moldova | Romanian, Russian | N/A | Chisinau + 19 other districts |
| Namibia | English | Oshiwambo | Ohangwena |
| Peru | Spanish | N/A | Departamento de Lima |
| Philippines | English, Cebuano, Tagalog | N/A | Cebu |
| Tunisia | Arabic | N/A | Sidi bouzid and Beja |
| Vietnam | Vietnamese | N/A | Thai Binh and Bac Giang |

## B.    Achieved sample - overall

23. The analysis presented in this section is based on achieved samples in wave 1. The size and profile of the achieved samples inevitably varied by country. Given that the proposed sample design focussed on degree of urbanisation as a key stratification variable that is a useful starting point to assess how achieved samples corresponded to the proposal.

24. Figure 1 shows how the achieved samples were actually distributed across strata. As a first point we can note that not all countries had the information necessary to use all 4 strata for sample selection. In Ivory Coast, Namibia and Tunisia it was not possible to subdivide PSUs between *Rural with market* and *Rural without market*. Additionally, in Kyrgyzstan it was not possible to stratify beyond urban and rural.

**Figure 1.** Sample households by place of residence (%)
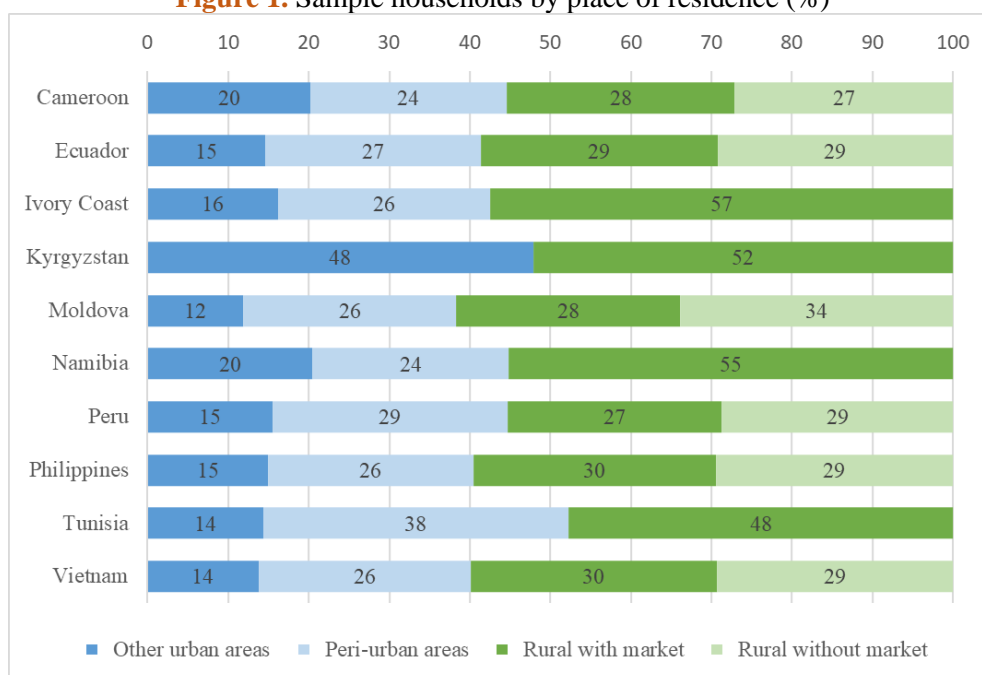


*The percentages presented are the average across the two model questionnaires in the country during the wave 1 fieldwork. The variation across model questionnaires within each country was minimal. Percentages may not add to 100 due to rounding.

25. The sample distributions by and large broadly aligned with the proposed sample design with relative oversampling in rural areas. The most notable exception was Tunisia where 48% of the sample was in rural areas but a further 38% was in peri-urban areas. As discussed further below, given that Tunisia had the highest achieved sample of any country (over 500 households for both questionnaires) this divergence from the proposed design is not expected to have any impact on analysis as sufficient rural households were still covered.

26. The total number of households and number of survey respondents varied across country and model questionnaire as shown in Table 5. We can note that overall 8,703 households were visited across the 10 pilot countries. The total number of respondents was 33,789 of which 23,187 were of working age.

27. Looking at number of households per questionnaire it can be observed that, at least for wave 1, the target sample of 400 households was achieved in almost all cases. The only case with substantially less than 400 was model 4 in Namibia (368 households for model questionnaire 4). A number of countries achieved samples at least 10% larger than the target (Ecuador, Peru, Philippines, Tunisia and Vietnam).

28. The sample size of 400 household per country had been chosen with the broad target of achieving samples of 1000 individuals of working age per model tested with the aim of having sufficient numbers of persons employed, not employed and own-use producers to allow for basic analysis regarding how well the alternative model questionnaires worked to capture them. This was for the most part achieved. Only Moldova achieved a slightly lower number of respondents in both models tested, whereas Ivory Coast and Namibia achieved slightly lower numbers in one of the models tested. Overall, the number of individuals of working age varied from 880 in Moldova (model 3) to 1,628 in Tunisia (model 2). Within countries the samples were typically very similar in size with

the greatest difference between models seen in Namibia where the sample of households for model 4 being approximately 10% smaller than the sample for model 1.

Table 5. Sample characteristics by country and model questionnaire, wave 1 of field tests

| Country | Model | Total households | Total survey respondents | Total respondents of working age* | % Working age respondents | Average household size |
|---|---|---|---|---|---|---|
| Cameroon* | M1 | 432 | 1865 | 1171 | 63% | 4.3 |
| | M5 | 398 | 1657 | 1040 | 63% | 4.2 |
| Ecuador | M3 | 450 | 1764 | 1217 | 69% | 3.9 |
| | M5 | 437 | 1700 | 1172 | 69% | 3.9 |
| Ivory Coast* | M1 | 409 | 1647 | 989 | 60% | 4.0 |
| | M3 | 397 | 1774 | 1060 | 60% | 4.5 |
| Kyrgyzstan | M2 | 390 | 1642 | 1108 | 67% | 4.2 |
| | M3 | 406 | 1722 | 1161 | 67% | 4.2 |
| Moldova | M3 | 391 | 1076 | 880 | 82% | 2.8 |
| | M5 | 418 | 1130 | 925 | 82% | 2.7 |
| Namibia* | M1 | 411 | 1976 | 1152 | 58% | 4.8 |
| | M4 | 368 | 1663 | 970 | 58% | 4.5 |
| Peru* | M3 | 450 | 1507 | 1121 | 74% | 3.3 |
| | M4 | 446 | 1486 | 1105 | 74% | 3.3 |
| Philippines | M2 | 493 | 2025 | 1329 | 66% | 4.1 |
| | M3 | 480 | 1879 | 1233 | 66% | 3.9 |
| Tunisia | M2 | 504 | 2141 | 1628 | 76% | 4.2 |
| | M3 | 511 | 1980 | 1505 | 76% | 3.9 |
| Vietnam | M3 | 448 | 1527 | 1178 | 77% | 3.4 |
| | M4 | 465 | 1628 | 1243 | 76% | 3.5 |

29. Notwithstanding the differences in size of the achieved samples, the profiles of the samples achieved across the two questionnaires within each country were quite similar with regards to the percentage of working age respondents and average household size. The percentage of working age respondents was essentially identical for both model questionnaires in all countries, while minor variations in average household size were found. Not surprisingly, both these features varied heavily across countries reflecting differences in the socio-economic contexts. For example, the average household size varied from below 3 in Moldova to close to 5 in Namibia.

30. Taking into consideration the feedback from countries and the overall size and distribution of the achieved samples, it can be concluded that the samples provide a reasonable basis for the comparative analysis which the studies needed to support. However, inevitably actual cell sizes generated could be small in cases, for example breakdowns of unemployment by age group, etc. This was borne in mind during the analysis phase and is taken into account in the presentation of results in the other reports in this series.

# C. Achieved sample – weighting and characteristics

## 1. Weighting of data

31. While the samples were not designed to be representative, it nonetheless remained important that within countries differences in the achieved sample across the two questionnaires did not excessively influence the analytical process. If the distribution of respondents by age and sex varied substantially between the two model questionnaires used in the same country, this could impact conclusions, for example about whether one questionnaire was more likely to identify employed people than another.

32. As already noted at an aggregate level, the percentage of respondents of working age and average household size was very similar across the two model questionnaires within each country. In addition to this, the ILO assessed the distributions of the sample by a variety of demographic and livelihood related characteristics to identify if sufficient differences were noted which could influence analysis. The outcome of this assessment was that differences across questionnaires within countries were by and large minor (see description of sample characteristics below).

33. Nonetheless the ILO decided to implement a basic weighting scheme to further reduce any impact of sample differences. This approach involved taking the average of the achieved samples for the two questionnaires in each country for key demographic variables and using this average to weight both questionnaires. Only information from the achieved samples for the pilot studies were used as the intention was to remove differences in the profiles of the achieved samples, not to weight to an external known population.

34. The variables chosen as control variables for the weighting were strata (degree of urbanisation), sex and age. The weights generated by this procedure are described by the equation below:

$$w_{m,i}^c = \frac{\varphi_{m,s,x,a,i}^c}{\sum_{i=1}^{n_m^c} \varphi_{m,s,x,a,i}^c} n_m^c$$

Where $n_{m,s,x,a}^c$ is the sample size for country "c", model questionnaire "m" strata "s", sex "x" and age group "a" and $n_m^c$ is the sample size for country "c" and model questionnaire "m"; and

$$\varphi_{m,s,x,a,i}^c = \frac{\sum_m n_{m,s,x,a}^c}{2\, n_{m,s,x,a}^c}$$

35. Given that the distribution of the samples across the two questionnaires in each country was quite similar, the weights generated were generally close to 1. The weights were generated in such a way as they did not alter the total number of respondents, i.e. if there were 1,200 respondents of working age for model x in country y then the weights would be rescaled to ensure that the post weighting sample total would remain 1,200. In this way, only the internal distribution of the sample was altered through the weighting procedure.

36. Table 6 shows the impact of the weights on the distribution of the sample by age and area. With few exceptions, the impact by age was relatively minor. However, in a few cases relatively large impacts were seen among the oldest age groups reflecting that these age groups were often relatively small within the overall sample and thus weighting could have a proportionally large impact. To illustrate this, we can take the case of respondents aged 65 and over in Ivory Coast. For model 1

there were 41 respondents in this age group (4.2% of all working age respondents). However, in model 3 there were 81 respondents in this age group representing 7.5% of working age respondents. The weighting applied a relatively large positive weight to this group for model 1 and a lower weight for model 3 to achieve consistency in the distribution.

37. Looking at the distribution by area we can note that, as explained earlier, not every country used all four strata. For a number of countries, we can see reasonably large weighting impacts representing some imbalances in the achieved samples, for example 28% for model 4 in Namibia but also above 20% in Ecuador and Ivory Coast.

**Table 6.** Impact of weighting on sample distribution by age and area

| Country | | Age groups | | | | | | | Area | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 15-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65+ | | Other urban areas | Peri urban areas | Rural with market | Rural without market |
| | | % difference (unweighted minus weighted) | | | | | | | | | | |
| Cameroon | M1 | 0 | -1 | -5 | -5 | 10 | 13 | | 6 | -9 | 13 | -9 |
| | M5 | 0 | 2 | 5 | 6 | -12 | -15 | | -7 | 11 | -15 | 10 |
| Ecuador | M3 | 1 | 3 | -1 | -1 | -5 | -3 | | -26 | 21 | -1 | -6 |
| | M5 | -2 | -2 | 1 | 1 | 5 | 3 | | 27 | -22 | 1 | 7 |
| Ivory Coast | M1 | -2 | -4 | 0 | -2 | 8 | -31 | | 24 | -21 | -1 | |
| | M3 | 2 | 4 | 0 | 1 | -7 | 29 | | -22 | 19 | 1 | |
| Kyrgyzstan | M2 | 3 | 2 | -2 | -2 | 8 | 0 | | 13 | | -7 | |
| | M3 | -5 | 0 | 2 | 3 | -8 | 0 | | -12 | | 7 | |
| Moldova | M3 | -6 | -6 | -7 | -4 | 14 | 14 | | 9 | 2 | -4 | 4 |
| | M5 | 6 | 5 | 5 | 5 | -13 | -13 | | -8 | -2 | 4 | -4 |
| Namibia | M1 | -2 | -2 | -1 | 6 | 5 | -5 | | -23 | 4 | 6 | |
| | M4 | 2 | 5 | 0 | -6 | -5 | 5 | | 28 | -4 | -7 | |
| Peru | M3 | 1 | -2 | -3 | -2 | 6 | -2 | | -3 | 5 | 1 | -7 |
| | M4 | 0 | 1 | 3 | 2 | -6 | 2 | | 3 | -5 | -1 | 7 |
| Philippines | M2 | -2 | -1 | 0 | -1 | 1 | 5 | | 0 | 3 | 2 | -5 |
| | M3 | 2 | 1 | 0 | 0 | -1 | -5 | | 0 | -3 | -2 | 6 |
| Tunisia | M2 | 2 | 2 | 2 | 2 | -1 | -6 | | -6 | -2 | 6 | |
| | M3 | -2 | -2 | -2 | -2 | 2 | 6 | | 6 | 2 | -6 | |
| Vietnam | M3 | 6 | 3 | 0 | 0 | -4 | 1 | | 2 | 12 | -13 | 5 |
| | M4 | -6 | -2 | 0 | 0 | 4 | -1 | | -2 | -11 | 12 | -5 |

## *2. Sample characteristics - demographics*

38. The data presented in the remainder of this section reflects the average profile across the two model questionnaires in each country post weighting. Only one set of figures will be presented per country. This can give an insight into how well the samples supported the type of analysis presented in the other reports in this series.

39. Figure 2 shows the distribution of the working age respondents in the sample by sex in each country. We can observe that, as would have been hoped, the samples were quite evenly distributed by sex in all countries. For all countries except Tunisia and Ecuador (equal representation of men and women) the percentage of women in the sample was slightly higher than men. Namibia had the highest percentage of females recorded (58%).

**Figure 2**. Sample distribution by country and sex (% of working age population)
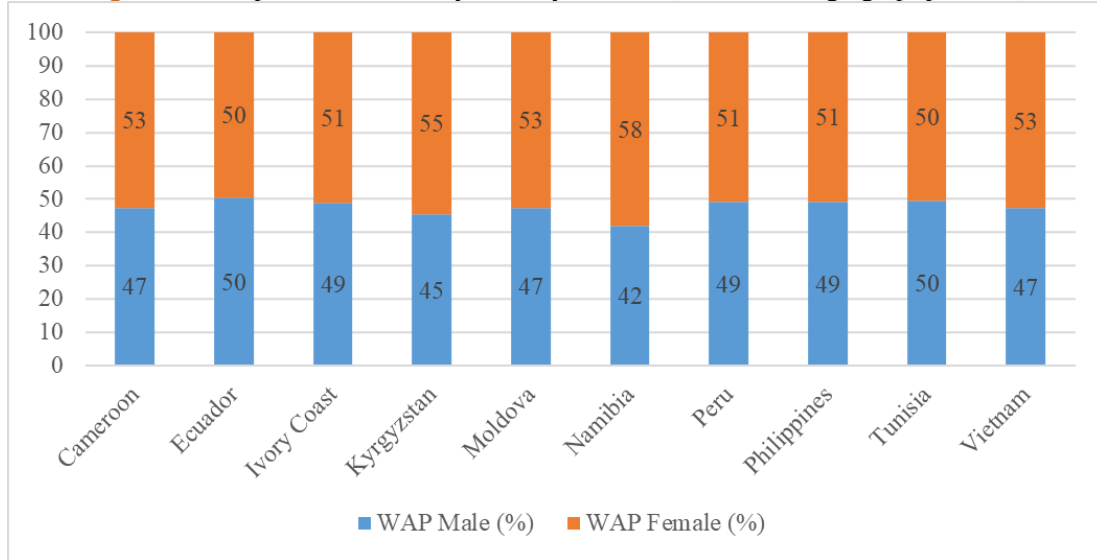
40. The sample distribution by age broadly matched the population pyramid of the countries in question (see Figure 3). In all countries respondents were spread widely across all age groups in the working age population. The highest percentages of youth (15 to 24 years old) respondents were recorded in Namibia (35%) and Ivory Coast (31%) while the lowest percentages were found in Vietnam and Moldova (both 15%). By contrast Moldova had the highest percentage of respondents aged 55 or above (38%).



**Figure 3.** Sample distribution by country and age group (% of working age population)
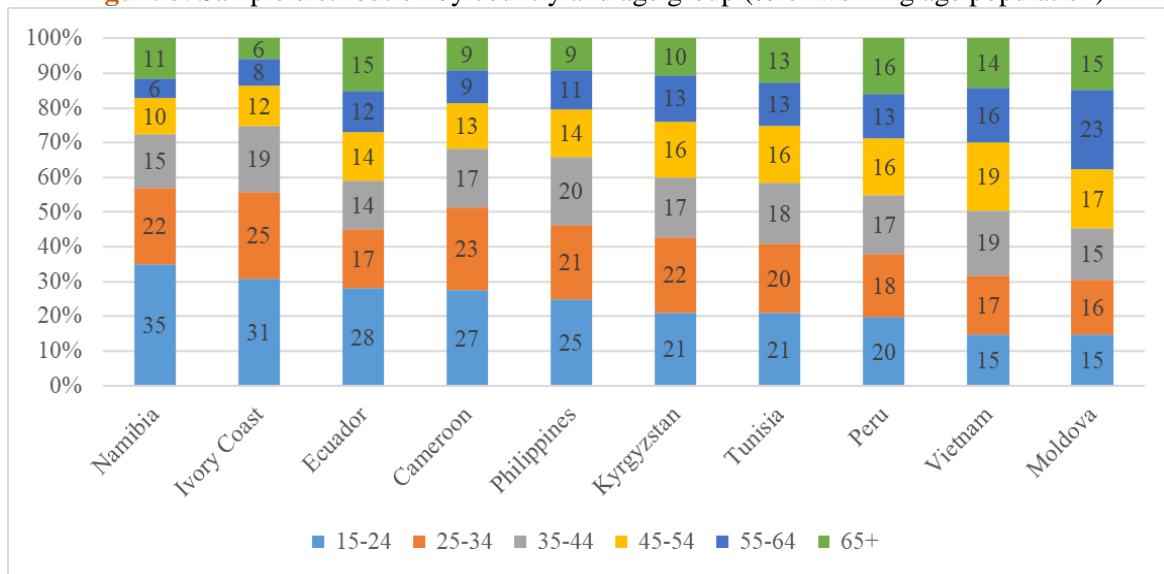
41. When looking at both sex and age together we see similar age profiles for men and women of working age in each country. Taking Peru as an example we see an almost identical distribution of men and women across age groups with minor differences in youngest and oldest age groups but identical percentages of respondents aged 30 to 64. This pattern was very consistent across countries (see Figure 4).

**Figure 4.** Sample distribution by country, sex and age group (% of working age population)
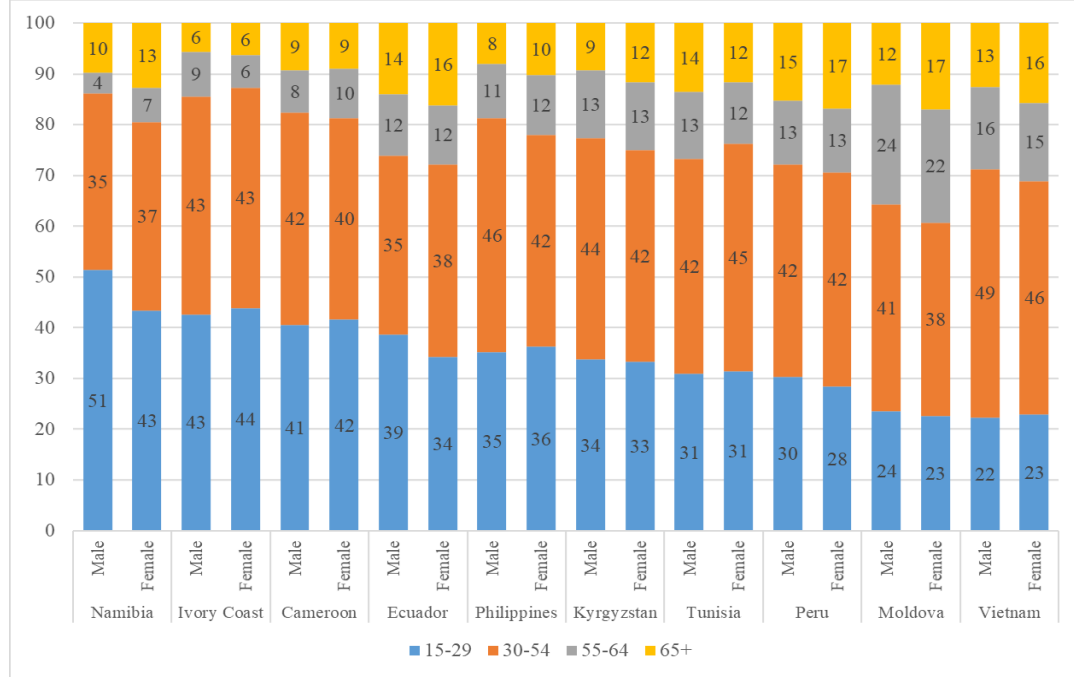


42. Another demographic characteristic of interest for the studies was the level of education of respondents. This is particularly true given its inevitable link with questionnaire comprehension. Given the intention to target rural areas most heavily, there was an expectation of relatively lower levels of education on average than may have been seen with a broad representative sample. Figure 5 shows that the situation varied across countries, again as would be expected. The largest group of respondents had primary or less as their highest level of education attained in 5 of the countries with the highest percentage found in Ivory Coast (75% of respondents of working age). At the opposite end of the scale a very small number of respondents in Moldova or Kyrgyzstan had achieved only primary education or less, with the large majority of respondents having achieved a secondary education, and a sizeable group having post-secondary education. Given the focus on rural areas and the different country contexts (i.e. access to formal education) these distributions are not surprising.

**Figure 5.** Sample distribution by country and highest level of education attained (% of working age population)
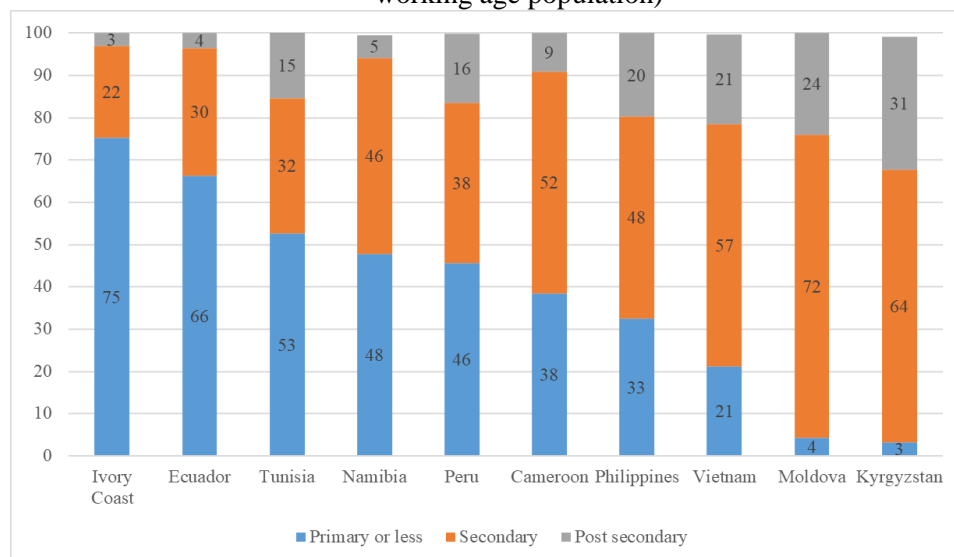
43. One conclusion which can be highlighted is that the distribution of respondents on key demographic variables (age, sex, educational attainment) shows sufficient spread to enable a wide range of analysis. Where differences are seen across countries they are not surprising given the areas targeted for the conduct of the pilot studies. Notwithstanding this, care was still needed in analysis given potentially small cell sizes generated when the sample was disaggregated by multiple characteristics.

### 3. *Sample characteristics – household livelihood*

44. The household level questionnaire used for the pilot studies included questions on household's resources and sources of livelihood. As with demographics, this information is useful to assess the types of households ultimately covered by the sample in different countries. As one example, bearing in mind some of the key measurement objectives, it was important to cover a reasonable number of households which reported foodstuff produced by members as one of the sources of livelihood. Only in this way could a reasonable assessment be undertaken of how different questions on the boundary between employment and own use production of goods worked.

45. Questions on individual and household livelihood had been covered by the cognitive tests. This had shown some comprehension difficulties among respondents regarding the term 'livelihood' (as translated to different languages) and in particular showed difficulty for respondents in reporting to a question of this type without predefined read-out response categories. Based on this experience the questions on household livelihood used for the field tests were developed. Respondents were initially asked the following question:

*"In the last 12 months, which of the following sources of support did the household have?"*

   a. *Income or other pay from a household business, including from farming or fishing*
   b. *Income or other pay from employment of household members including yourself*
   c. *Foodstuff produced by the household from farming or fishing*
   d. *Remittances from abroad*
   e. *Support from other households in the country*
   f. *Income from properties, investments or savings*
   g. *Private or state pension or other Government support*
   h. *Charity from NGOs or other charitable organisations*
   i. *OTHER (SPECIFY):_____*

46. For those indicating more than one source of household support a follow up question was asked to identify which of the sources was considered (subjectively) to be the main source of support.

47. This combination of questions enabled a range of analysis. As a starting point the number of sources of support available to households can be analysed. Figure 6 shows that mixed sources of livelihood were most common in Moldova (85% of sampled households) followed by Ecuador (67%). At the opposite end of the scale 87% of sampled households in Tunisia reported only one source of livelihood. Among the remaining pilot countries, the distribution between mixed and single sources households was relatively more balanced.

**Figure 6.** Sources of household livelihood by country
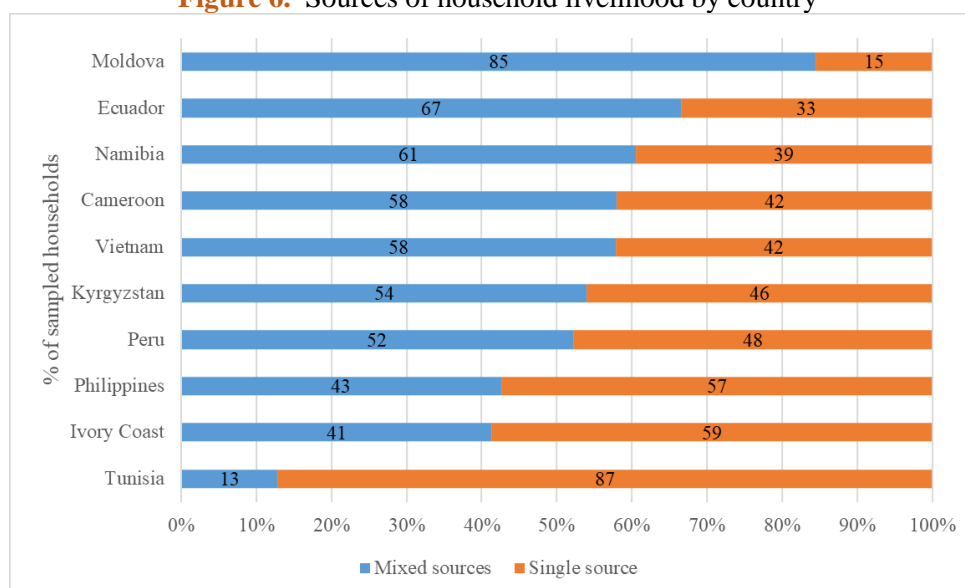
48. Expanding on this we see that the average number of sources of livelihood ranged from 1.11 in Tunisia to 2.47 in Moldova (see Table 7). Other than these two countries the averages all lay in the range 1.41 to 1.94.

49. Table 7 also presents the prevalence of different types of source of livelihood. Given that households can have multiple sources of livelihood the columns for each country do not sum to 100%. As with many other issues covered by the pilots the pattern of livelihood sources varied significantly across countries. For example, the proportion of households reporting income from a household business (category A in Table 7) as a source of livelihood ranged from 15% in Tunisia to 60% in Ivory Coast. When income from employment of household members (category B in Table 7) is also taken into account then it becomes clear that the majority of sample households in most countries relied on some form of employment income.

**Table 7.** Type of source of livelihood by country (% of sample households)

| Country | Type of source of livelihood | | | | | | | | Avg No. |
|---|---|---|---|---|---|---|---|---|---|
| | a | b | c | d | e | f | g | h | |
| | *% of sample households* | | | | | | | | |
| Moldova | 16 | 51 | 66 | 24 | 9 | 20 | 61 | 0 | 2.47 |
| Ecuador | 44 | 53 | 38 | 1 | 13 | 1 | 38 | 1 | 1.89 |
| Namibia | 23 | 35 | 52 | 3 | 32 | 3 | 44 | 3 | 1.94 |
| Cameroon | 44 | 42 | 62 | 5 | 16 | 4 | 3 | 1 | 1.77 |
| Vietnam | 48 | 63 | 21 | 2 | 10 | 4 | 25 | 1 | 1.74 |
| Kyrgyzstan | 19 | 50 | 24 | 21 | 7 | 1 | 46 | 2 | 1.68 |
| Peru | 46 | 59 | 24 | 3 | 13 | 7 | 15 | 1 | 1.67 |
| Philippines | 50 | 61 | 10 | 5 | 8 | 1 | 10 | 4 | 1.50 |
| Ivory Coast | 60 | 45 | 23 | 2 | 9 | 2 | 1 | 0 | 1.41 |
| Tunisia | 15 | 75 | 2 | 2 | 1 | 1 | 14 | 2 | 1.11 |

***Table key: Type of source of livelihood***
  a. *Income or other pay from a household business, including from farming or fishing*
  b. *Income or other pay from employment of household members including yourself*
  c. *Foodstuff produced by the household from farming or fishing*
  d. *Remittances from abroad*
  e. *Support from other households in the country*
  f. *Income from properties, investments or savings*
  g. *Private or state pension or other Government support*
  h. *Other (including charity)*

50. In the case of foodstuff produced by the household (category C) all but two countries had more than 20% of sample households reporting this as a source of livelihood with half or more of households in Moldova (66%), Cameroon (62%) and Namibia (50%) in this situation. The exceptions were Philippines (10%) and Tunisia (2%). In the case of Tunisia in particular, this could create some difficulty in the analysis of outcomes on own-use production of goods and the boundary between employment and own-use production. This is taken into account in the development of the thematic reports on those topics.

51. Across the other sources of livelihood, we can see that relatively a large number of households reported a pension or other Government support in a number of countries with up to 61% in the case of Moldova. This was at its lowest among sample households in Cameroon (3%) and Ivory Coast (1%). Remittances were uncommon in all countries except Moldova (24%) and Kyrgyzstan (21%). Support from other households in the country was highest in Namibia (32% of sample households) but was reported by less than 10% of households in half of the countries. Very few households in any pilot country reported 'other' sources of support, which included charity.

52. While Table 7 shows the diverse sources of livelihood relied upon by the households, the surveys also asked for the main source among those with more than one source. As shown in Figure 7, employment income was the most common main source of livelihood for sample households in all countries albeit ranging from 41% of sample households in Namibia to 90% in Philippines. Employment income was particularly predominant in Philippines, Tunisia (85%), Peru (84%), Ivory Coast (83%) and Ecuador (81%). As such, even where households had multi-faceted strategies to generate their livelihood, employment income was in most cases considered the main source.

**Figure 7.** Main source of household livelihood of sampled households by country
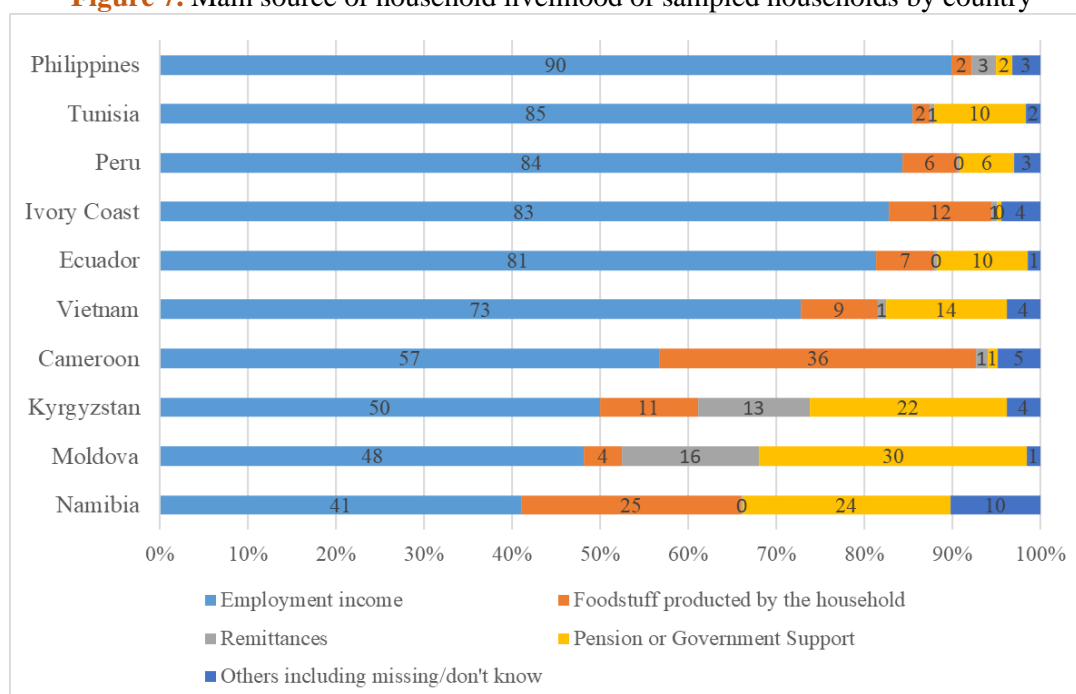


53. Cameroon (36%) and Namibia (25%) had the highest proportions of households reporting foodstuff produced by the household as a main source of livelihood. Among other sources, pensions or other Government supports were a relatively common main source in Moldova (30%), Namibia (24%)

and Kyrgyzstan (22%), while remittances were most common in Moldova (16%) and Kyrgyzstan (13%).

54. In summary, we can conclude that it was common for the sample households in the pilot countries to have multiple sources of livelihood, and substantial numbers of households relying on employment income or own-produced foodstuff were found in addition to the various other sources. For the most part, this is expected to provide a good base for analysis across the key topics covered by the pilots, such as measurement of employment and own-use production.

## 4. Sample characteristics - household resources

55. In addition to sources of livelihood, the surveys included information on different types of household resources and characteristics. In all model questionnaires questions were included on the main sources of energy and water. Additionally, in model questionnaire 1 there were questions on access to agricultural land and ownership/use of farm or work animals. Household level information on these resources was also deemed important to support validation of reports of participation in own-use production of goods, in particular, fetching water, collecting firewood, farm-work and animal husbandry.

56. Sources of energy, which required labour from household members to collect it (such as wood, animal dung etc.), were the dominant source of energy for household in Cameroon, Ivory Coast, Namibia, and the Philippines, (between 74% and 81% of households). For Ecuador, Moldova, Peru, Tunisia and Vietnam the predominant source of energy for cooking was natural gas, ranging from 68% of sample households in Peru to 97% in the case of Tunisia (see Figure 8). Nevertheless, except for Tunisia, a small share household (ranging from 9% to 30%) in these pilot countries also indicated relying on wood or charcoal from wood as their main source of energy. It is worth noting that the energy mix is dominated by conventional sources such as gas, oil and wood with little usage of renewable energy such as solar energy.

**Figure 8**. Sample households by country and main source of energy
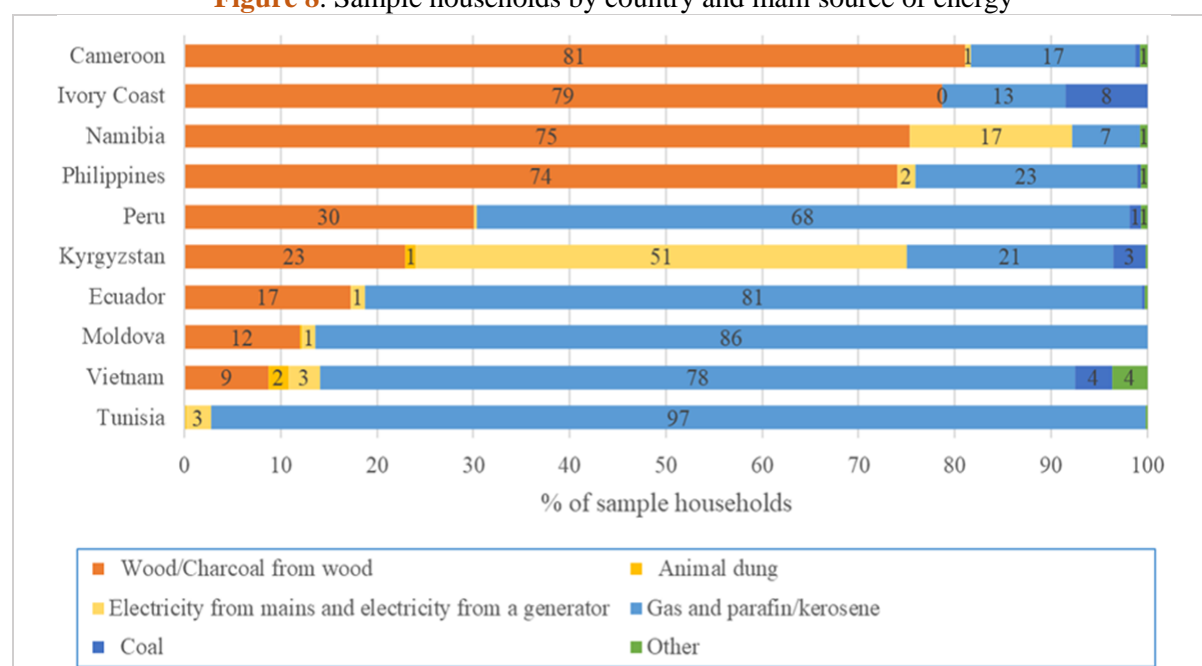
57. For main source of water respondents were able to indicate various sources. For analytical purposes those have been combined into sources which are inside the dwelling (e.g. piped water or a well within the dwelling/compound), sources outside the dwelling that require labour for their acquisition (e.g. a well, lake, river, piped source outside the compound) and others. Whereas in Vietnam, Peru, Moldova and Tunisia more than half of households had sources of water inside the dwelling, it was not the case in the remaining countries (see Figure 9). In Cameroon and Ivory Coast for instance, 78% and 69% respectively of the households' main source of water for cooking and drinking was located outside the dwelling. Among other things, this has implications for likely findings on own-use production activities such as collection of water.

**Figure 9.** Main sources of household water by country
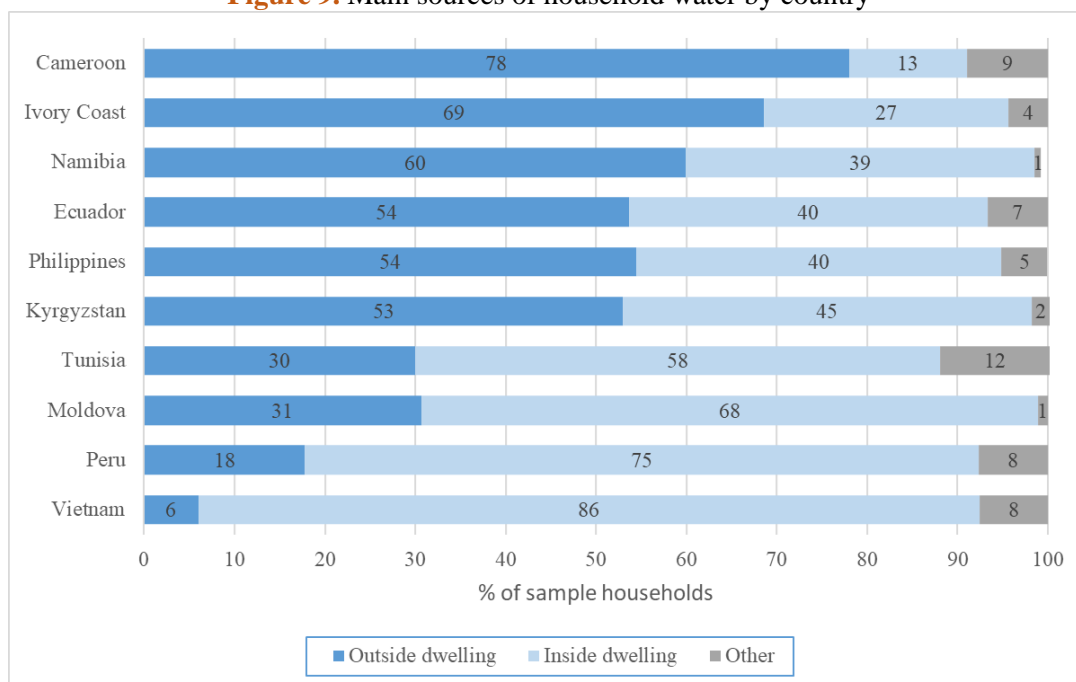


58. Model questionnaire 1 was designed with a particular focus on capturing work in farming or fishing early in the questionnaire, whether it was employment or for own-use. This approach was chosen to contrast with questionnaire designs which focussed more heavily on capturing employment in early sections of the questionnaire and measuring own-use production activities later. As part of this approach, additional household level questions were included in model 1 regarding access to agricultural land of household members and ownership and use of farm or work animals. As model questionnaire 1 was only used in 3 of the pilot countries (Cameroon, Ivory Coast, Namibia) the analysis of these variables only covers those countries.

59. For access to agricultural land the question asked was *"Does any member of this household currently own, rent or have access to any land that can be used for agriculture?"*. This question reflects an interest in being able to identify households where mixed agriculture (some sold, some consumed) s part of the livelihood strategy. The question wording was chosen to reflect that the type of access to land may vary in reality and it was not desired to limit the focus only to households with specific types of access (e.g. owning their own plots). Figure 10 shows a relatively uniform situation across the 3 countries with between 77% and 80% of sample households reporting having access to agricultural land. Notwithstanding this similarity, as shown in other analytical reports in

this series, the types of activities undertaken, e.g. mix between employment and own-use production did vary across country.

**Figure 10.** Access to agricultural land by country
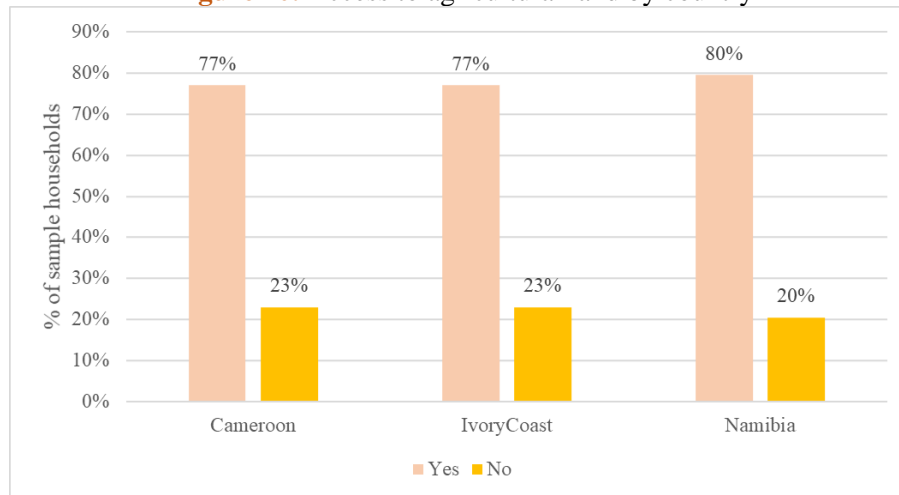


60. Similarly, it was common for the sample households to report ownership of animals as shown by Figure 11, ranging from 64% in Namibia to 85% in Ivory Coast. The combined conclusion from questions on land access and animal ownership is that agricultural activity was sufficiently common in the sample households to achieve the overall intention behind the sampling design of the pilots in these 3 countries, namely to cover sufficient farming households to assess the operation of questions on employment and own use production in that context.

**Figure 11.** Ownership of farm or work animals in sample households by country
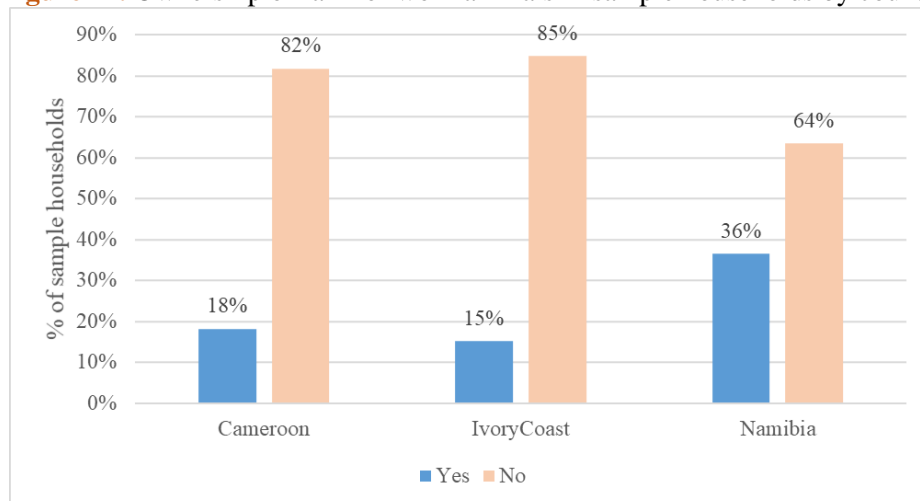


## D. Interview characteristics – proxy response and duration of interview

61. In this final section we look at the characteristics of the interviews achieved. In the case of proxy interview countries were advised to adopt strategies, such as multiple visits and setting appointments, to minimise proxy levels to the extent possible. Nonetheless, proxy response was not prohibited in order to potentially enable some assessment of proxy effects but also to ensure the

pilots were completed in a manner relatively similar to typical implementation of household surveys in the countries. Unsurprisingly, and in line with the typical experience in the countries in question, the level of proxy response differed. Proxy response was 40% or more in 4 of the pilot countries and as high as 47% in Vietnam and Namibia. The lowest levels of proxy response were recorded in Philippines (11%) and Ecuador (8%).

**Figure 12.** Proxy respondents as % of working age population (WAP) by country
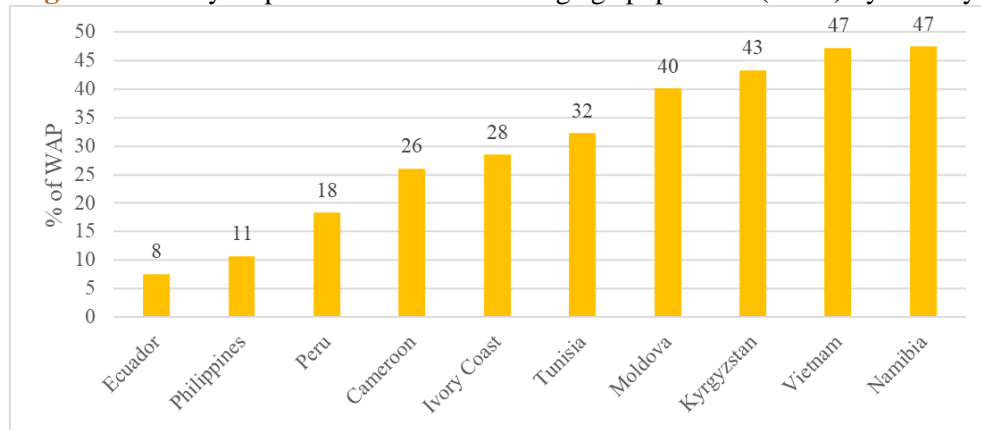


62. Some variation in personal interview duration across model was expected given the different approaches taken to identify the employed, and to capture own-use production work. The overall situation across the five models is shown in Figure 13. Some care is needed in interpreting this information at aggregate level given that different questionnaires were used in different countries and some variation could be found across countries. Nevertheless, the average durations documented are in line with expectations, with model 3 and model 4 showing longer duration than other models. This reflects the fact that models 1, 2 and 5 had some elements of design intended to shorten the questionnaire sequences for certain groups, whereas this was less so the case for models 3 and 4. Overall, the average durations of the interviews remained below 20 minutes, which is within the usual average duration of national LFS interviews.

**Figure 13.** Average duration of personal interview by model
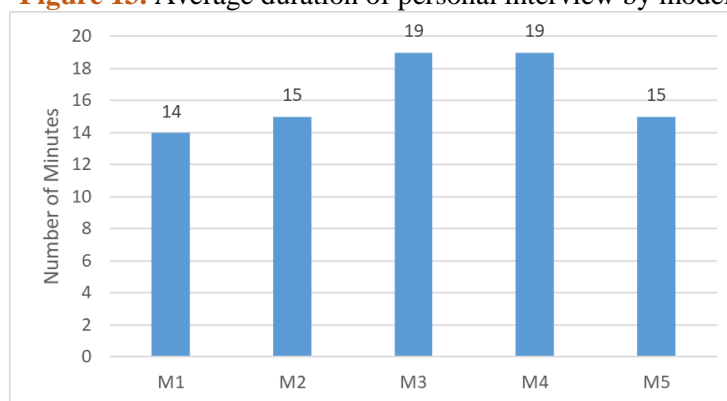


63. The situation in different countries is reflected in Figure 14. For the most part, very similar average durations were found across models and where differences are not found they are not unidirectional. For example, in 3 countries model 2 and model 3 were used (Kyrgyzstan, Philippines and Tunisia). In two of those cases, model 2 had a shorter interview duration which was the generally expected outcome. However, the opposite was observed in Tunisia where we can also note the shortest durations were found of any country (4 minutes for model 3 and 6 minutes for model 2). The short

duration of interview in Tunisia can at least partially be linked to relatively lower proportions of the working age population reporting own use production of goods as compared with the other pilot countries.

64. The longest interview duration was recorded in Vietnam (in excess of 25 minutes on average for both models). This can be linked to particularly high levels of employment recorded, as highlighted, in the thematic report on employment which would increase interview duration. For the large majority of countries and models, average interview duration was somewhere in the range 10 to 20 minutes.

**Figure 14.** Average duration of personal interview by model and country
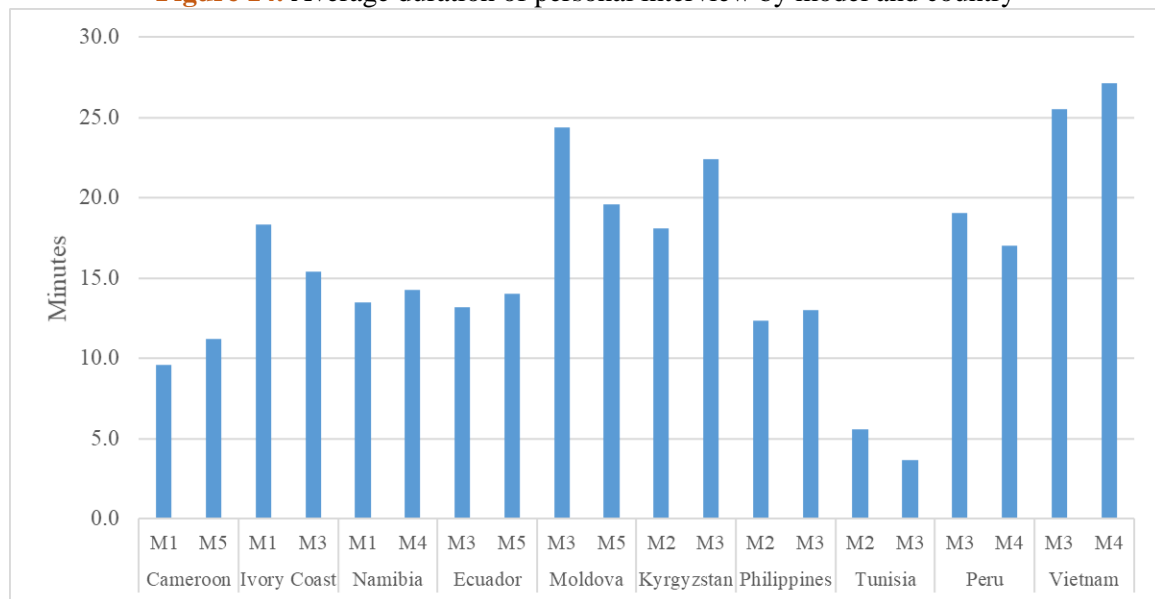


65. Additional analysis of interview duration by respondent characteristics showed no major patterns of interest which would influence analysis of the pilot study outcomes. For example, average duration for males and females was similar. However, predictably longer interviews durations were recorded for people who reported more activities (e.g. employed and an own use producer) than others, such as those outside the labour force.

66. In summary, while a simple aggregate level analysis suggests that interview duration is longest for models 3 and 4, this clearly depends on the respondent profile and other factors. As such, it should not be presumed that the main differences in design of model 3 and 4 versus others would automatically raise interview duration – this will depend on the other content of the interview, etc. As with all things this should be assessed through pre-testing of any questionnaire.

# IV. REFLECTIONS ON THE FIELD TESTS PROCESS

67. The bulk of this report has been dedicated to describing how the implementation of the field tests process advanced and to understanding how the sample characteristics could influence the analysis presented in other reports. It is also useful to reflect on the implementation and draw some lessons for similar future activities, such as those which might take place at national level.

68. An important point to highlight is that very careful consideration is needed of the appropriate scale of tests to take place. The selection of 10 partner countries for the pilot studies undoubtedly has some benefits in ensuring wide coverage of studies and thus as wide as possible applicability of the results. However, the process of working with 10 countries simultaneously also created substantial pressures which did impact on the outcomes. These included:

   a. Technical and financial support were spread very thinly. On the technical side, this made it difficult to provide the level of support which would ideally have been targeted leading to some inconsistencies in implementation which created difficulties during data processing and analysis. On the financial side, the level of funding anticipated was not ultimately available. This meant that wave 2 of field collection could only completed in 7 of the 10 countries. Among the 7 countries who completed wave 2, funding had to be sought during the process in some cases and some NSOs had to make an additional financial contribution to the process.

   b. Timelines were significantly impacted. Planned timelines for the cognitive phase were broadly maintained. However, the field-testing phase was initially planned to be completed by mid-2016 for all countries and both waves. As a knock-on effect of the resource issues mentioned above, timelines were pushed back and wave 1 of field testing was only completed in the final country in October 2016 with knock on effects on the analytical phase.

   c. Substantial time was required to complete data processing. While efforts were made to standardise the format of the data to be submitted, this was not fully achieved and the number of countries involved meant the differences became substantial. This resulted in data processing taking longer than planned. For example, it was intended to create unique identifiers as a combination of some of the identification variables. These identifiers would, among other things, allow easy linking of data between wave 1 and wave 2. However, this did not come about in a number of countries and ongoing contact was required to resolve this. Another issue was that not all countries ended up having the resources to develop a data entry system as planned. To fill this gap, the ILO engaged a consultant to develop data entry applications in CSPro. As this was unplanned, it added time and cost to the process but did generate benefits with regards to the format of data received. Overall substantially more time was required than initially planned to process the datasets to the extent needed to enable analysis.

   d. The analytical process was further complicated by the large number of datasets involved. There are few if any precedents for a pilot study process of such a scale, organised with a fixed methodology with the purpose of cross country and within country comparison. Even

on a smaller scale the analytical challenge would have been substantial, but on the scale completed the process of deriving clear and solid conclusions becomes incredibly difficult as not all observed patterns will be repeated in the same manner in all countries.

    e. The design envisaged the two waves of fieldwork taking place at different peak periods of agricultural activity. Given the varied situation across countries (different crops, climate etc.) and the tight overall timelines for the project, it was not possible to achieve this in all cases. For example, in Kyrgyzstan and Moldova there was serious time pressure to commence the first wave in November 2015 in order to avoid the very low winter season which commences in November/December each year. This required very quick turnaround of preparation of materials following the cognitive testing. As a consequence of issues of this nature, very different timelines had to be adopted across countries and in some cases part of the interviewing took place in low seasons. This in particular could influence longitudinal analysis of the two waves of data.

69. In addition to the number of countries the range of measurement objectives certainly added complexity. The questionnaires were quite extensive, covering employment, labour underutilization, own use production work and various other aspects. This added to the efforts needed in questionnaire design, development of supporting materials, training, data processing and analysis.

70. On reflection it could have been preferable to focus on a smaller number of pilot countries and questionnaires/issues for comparison. This would still have enabled valuable analysis to have been completed but would have lowered the level of complexity and burden involved, thus allowing outputs to be generated more quickly and creating greater room for follow up studies on topics requiring additional assessment. It would also have eased the planning process to enable closer adherence to the study design. The process would have remained complicated but certainly easier to manage.

71. While it was important to standardise elements of the methodologies across countries, it was equally important to maintain flexibility in implementation to ensure the studies were meaningful for the country in question. This flexibility extended across a range of issues such as national adaptation of questionnaires to include appropriate national examples or wording. These adaptations were necessary but also difficult to manage as it needed to be ensured that any changes didn't substantially alter the flow of the questionnaire or meaning behind questions. The slightly different timing across countries made this all the more complex. The lesson learned from this is that additional time and resources should be planned to manage adaptations to questionnaires and the many other processes where national specificities had to be accounted for.

72. Looking at the more positive lessons learned, we can note that the broad methodological approach provided a very good basis for assessment of questionnaires. Combining qualitative (cognitive) and quantitative (field tests) methods enabled quite in-depth analysis and both contributed substantially to the process. The process of cognitive testing in particular requires some familiarisation but was very positively received by the participant countries with many planning further use in the future.

73. The importance of building in supervision and feedback into the testing process cannot be overstated. Some of the analytical conclusions drawn have been heavily supported by observations

in the field, both during cognitive and field interviews. In addition, feedback from the interviewing teams is a very valuable resource and gathering this feedback should be part of any testing process.

74. When combining the above, possibly the most important overall lesson learned is to carefully plan the scope of studies from the beginning and match this scope with sufficient resources and time. If implementing multiple stages of testing (e.g. cognitive and field tests), sufficient time must be allowed between stages to enable changes to be made. For any agency planning a co-ordinated cross-country set of studies additional time should be allowed for all the national adaptations and efforts needed to try and achieve harmonisation of outputs for analytical purposes.

75. Finally, the importance of adequate testing of questionnaires cannot be clearer. While different testing methods can be implemented, there are serious risks of adopting questionnaires without national testing on some scale. While ILO and others can produce model questionnaires, it was evident for all countries that different types of adaptation are necessary which can ultimately only be confirmed through a robust development and testing process at the national level. Any country planning to implement new household surveys or update existing ones should plan for tests in sufficient time to update questionnaires and other materials and processes before live implementation. As part of a case to build support and gain the resources needed it can be emphasised that good testing processes generate invaluable information for data users who can otherwise be left with breaks in series with little if any ability to explain or understand their source. While testing cannot necessarily explain every source of difference, when properly designed, it does massively improve the possibilities to understand any changes seen in indicators as well as generally adding to the quality assurance of the survey process.

# V.  OVERALL CONCLUSIONS

76. The purpose of this report is to describe the pilot study project methodology and the implementation of the field test phase of the studies. This acts both as a record of the objectives, design and process involved, and provides context for those interested in the types of sample achieved and how that might impact the outcomes of the studies.

77. The overall scale of the studies was quite large, involving 10 pilot countries and over 23,000 respondents of working age in wave 1 of the field collection. The range of settings involved was also wide, from heavily urbanised areas in middle income countries to remote rural areas without access to markets and different types of agricultural and other economic activities. This offers a substantial base upon which to undertake the analysis required by the measurement objectives. It also creates challenges of trying to undertake a comparative analysis with so many elements.

78. The proposed sample design was broadly observed in all the countries with any differences considered unlikely to impact analysis. Furthermore, the differences between achieved samples for the two model questionnaires in each country were typically minor when assessed against key demographic variables.

79. To minimise the analytical impact of any differences in achieved sample between the two model questionnaires covered by each country, a simple stratified weighting scheme was used by the ILO. The relatively minor nature of differences in the samples meant that the weights were in general close to 1. The weighting scheme did not attempt to benchmark against any external known population, rather just corrected for differences across the achieved samples. This reflected the general approach to the pilot studies whereby it was not attempted to generate representative estimates, rather undertake comparative study between the model questionnaires.

80. While the proposed sample design was purposive, reflecting the measurement objectives of the pilot studies, it was nonetheless important to achieve samples which covered a wide cross-section of different segments of the population, i.e. men and women, different age groups, education levels etc.

81. The samples achieved across countries varied substantially both in terms of size and distribution across types of households and individuals interviewed. These differences very much reflected differences in the national context such as population pyramids, access to education and so on. However, critically for the analysis phase of the studies, all countries selected field sites that matched the desired profiles for the tests. In addition, the samples in all countries were varied enough to support a wide range of analysis.

82. Inevitably, there will be cases where achieved samples do not allow detailed disaggregation of findings by all desired characteristics. For example, as highlighted in the report, the small sample in Tunisia of households reporting foodstuff produced by household members as a source of livelihood could impact the ability to undertake a detailed analysis of differences between models 2 and 3 of measurement of own-use production of goods and related issues for that country. The sample will continue to be adequate for various other measurement objectives. Where considered of significance, the ILO will highlight specific cases where sample sizes were considered a limitation for the analysis in the different thematic reports published alongside this report.

83. Finally, while specific lessons can be learned from the exercise, one key general message which must be taken is that testing of questionnaires is a critical process which should be planned for when implementing a new survey or adapting an existing one. The specific plan will depend on available resources and information demands.

# VI. REFERENCES

ILO. (2013). Resolution I concerning statistics of work, employment and labour underutilization. *19th International Conference of Labour Statisticians* . Geneva: ILO.

ILO. (2018). *LFS Pilot Study Programme*. (ILO, Producer) Retrieved from ILO: http://www.ilo.org/stat/Areasofwork/Standards/lfs/WCMS_484981/lang--en/index.htm