Did you sleep here last night? The impact of the household definition in sample surveys: a Tanzanian case study.

Tiziana Leone, LSE
Ernestina Coast, LSE
Sara Randall, UCL

Abstract
Household sample surveys are integral to planning for development in most poor countries and there is growing demand for data to inform development strategies. The practicalities of data collection require a social unit to be defined, generally referred to as a household, although households as defined by survey practitioners may differ considerably from the social units that many people live in. Most survey and census definitions of a household rely on some combination of 3 factors: sleeping in the household the night before the interview; eating from a common cooking pot; and sharing economic resources. Using Tanzania as a case study, this paper firstly analyses the impact of different household definitions on key socio-demographic indicators. Secondly it explores how to overcome the limitations of the definition both in the data collection and analysis. The aim is to highlight shortcomings of household data and to investigate the possible impact that the outcomes might have on policy-making. This study uses the 2004 Tanzanian Demographic and Health Survey (DHS) (n=9735 households) and primary in-depth (n=52) case study interviews with Tanzanians in four different settings. Analyses use sensitivity analysis to plot possible scenarios for a set of socio-demographic indicators (dependency ratio, sex ratio, median education level, mean household size, and sex of household head), indicators which are used frequently as proxies or correlates of development. Results show that the household age and sex structure change considerably if the approach of the definition changes. For example both the dependency ratio and the proportion of female-headed households increase if more stress is given to the sleep in definition. This study has two implications. Firstly, that household survey instruments might be adjusted to better capture a range of lived realities. Secondly, information that will help survey analysts to better understand and interpret household survey data.

Background

Household sample surveys are integral to planning for development in most poor countries and there is growing demand for data to inform development strategies. This demand received further impetus from the Millennium Development Goals (MDGs). The practicalities of data collection, analysis and policy formulation require a social unit to be defined, generally referred to as a household, although households as defined by survey practitioners may bear little resemblance to the social units that many people live in. Little attention is paid to the issue of how the survey 'household' was defined and what this might mean for interpretation. Most survey and census definitions of a household rely on some combination of 3 factors: sleeping in the household the night before the interview; eating from a common cooking pot; and sharing economic resources.

There are two key levels of challenges: at individual level and at household level. At individual level there is a range of fuzziness which could influence whether the individual is included or excluded from the household. This could be due to double membership or to absence/presence near the time of the survey or simply to the nature of their staying in the household such as borrowed children or older relatives which migrate from household to household.
At household level the issue is mainly linked to the structure in terms of what is considered to be as a household from the survey point of view and from the

household members one. The fuzziness at this level could often lead to missing households, splitting of de-facto ones and merging of others.

Using Tanzania as a case study, this paper analyses the impact of different household definitions on key socio-demographic indicators (e.g.: dependency ratio, sex ratio, median education level, mean household size, and sex of household head).

Data and methods

This study uses two datasets. Firstly, secondary analysis of the 2004 Tanzanian Demographic and Health Survey (DHS) (n=9735 households). Secondly, analysis of primary in-depth (n=52) case study interviews with Tanzanians in four different settings.

The DHS relies on a cooking pot definition of the household. From DHS Tanzanian final report p.9: "For the purpose of the 2004-05 TDHS, a household was defined as a person or a group of persons, related or unrelated, who live together and share a common source of food". The DHS collects information on both members and visitors of the household that slept there the night before and those members who have been absent for less than three months. It is therefore possible to consider both the de facto and the de jure population.

By examining differences in the reported de jure and de facto population, we investigate the impact on socio-demographic indicators of different definitions of the survey household. This distinction cannot highlight discrepancies within the survey but it allows in principle to investigate the implications of the de facto approach used in the Tanzanian census.

For the analysis of the DHS data in particular we considered the following factors: What happens to indicators at the household level if done on de jure or de facto basis? For range of indicators we compared and contrasted de jure and de facto. We further explore two types of households which are known to be problematic both for survey measurements and for policy issues: single headed households and female headed households. Using Tanzanian Socio Economic Database as guide for key development indicators which form the basis for policy making and DHS as source we considered the following indices:

- Household size
- Sex household head
- Number of people per sleeping room
- Age distribution household head
- Dependency ratios

Household size is typically used as an indicator associated with aspects of household welfare. Female-headed households are, for example, typically poorer than male-headed households. Larger households are generally associated with greater crowding in the dwelling, as well as poverty and unfavourable health conditions.

The second data source comes from a fieldwork conducted by us in Tanzania in four different settings. Case study interviews in a range of settings with Tanzanians about the membership of their household and the residence, production and consumption of household members (Figure 1).

a. Dar Es Salaam: Interviews (n=24) were undertaken in two different low income areas of Dar Es Salaam using four experienced interviewers from a university-based consultancy. Participants were organised in advance to expect us, and told the basic aims of

our interviews, through personal contacts and local leaders, resulting in high levels of cooperation. Our fieldwork supervisor selected a range of different household types and circumstances, and the interviews involved one of the authors working with an experienced interviewer. Most interviews were recorded and all were in Swahili.

b. Maasai agropastoralist community: Interviews (n=8) were collected using a university educated interpreter from that community who had previously worked there as a research assistant for an anthropologist who was also present, having worked on the literature review for this project. Both knew the households and respondents well and provided supplementary information for triangulation. A further interview was undertaken with one Swahili household living in this area. Given the size and complexity of Maasai households these interviews covered the equivalent of about 20 DHS defined households

c. Rufiji community Interviews (n=20) were collected by a research student with in-depth knowledge of the community and its setting[1].

We developed our interview approach out of cognitive interviewing (Willis, 2004), making it clear from the beginning that there were no clear right or wrong answers, that we wanted to explore and understand diverse living arrangements and encouraged respondents to talk about individuals whose status with respect to that living unit might be ambiguous. Often two or three respondents were present for the interview and frequently younger members reminded older males about the existence of various children. We emphasised that this was not part of a larger survey and was just a small, one-off study. Groups a and b were given a small remuneration for their time.[2] At the beginning of each interview the interviewer explained that we were particularly interested in everyone who belonged to that '*kaya*' and that we wanted to know who they thought were members of their *kaya,* whether actually present or not[3]. Using a household grid similar to those used in major household surveys (but with considerable space for written comments) we listed all the members in the household as reported by the respondent and not as dictated by a predefined definition, After obtaining a list of *kaya* members we asked about relationships and marital status, location of absent people, details of people who had slept there last night, sleeping and eating arrangements both the previous night and more generally, sources of income and support and links with other households either through providing or receiving economic support. Limited questions on asset ownership were asked. This information was meant to establish the discrepancies between some of the key concepts usually conceived to define the household such as eating and/or sleeping together and what interviewees consider as THEIR household. The open and discursive questioning also allowed for discussion around household members whose status was ambiguous even for the respondent. Reasons for such ambiguity were discussed as were situations of membership of multiple households.

Our interviews were not strictly cognitive interviewing which is designed to test respondent's understanding of survey questions. However building on

---

[1] Need to complete

[2] In all cases respondents were extremely happy to participate and in the two research sites where PhD students had undertaken long term work in the community they both commented on the willingness and interest of the participants compared to other forms of data collection. Group c was not remunerated because the student concerned was still working in the community and did not want to generate future demands for money.

[3] We explained that we were interested in those who they thought were part of their local unit and not just those people registered as being the 'kaya' members on the 10 cell unit lists.

the ideas behind cognitive interviewing respondents were encouraged to reflect on what made people a member or not of their household and whether these were economic, emotional, supportive or other ties and obligations.

Figure 1. Tanzania fieldwork sites
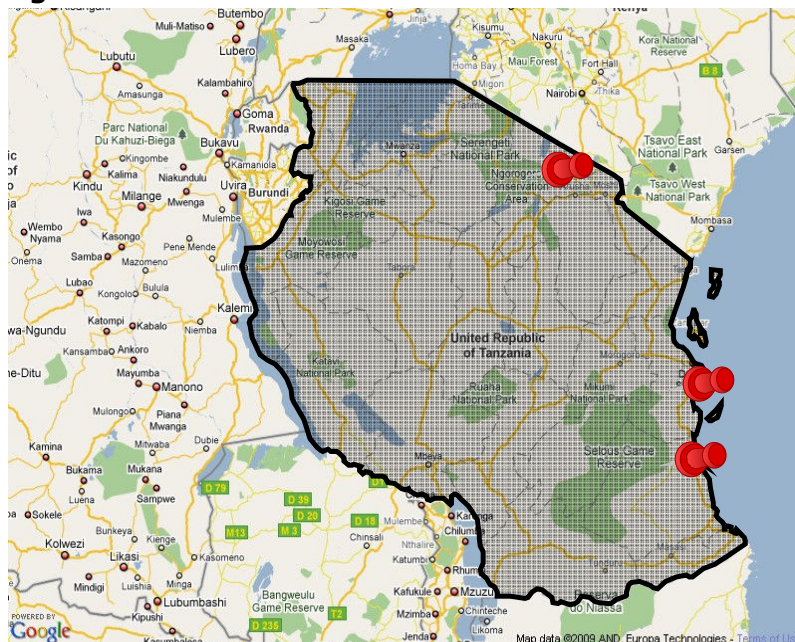
We analyse these data using sensitivity analysis to plot possible scenarios for a set of socio-demographic indicators (dependency ratio, sex ratio, median education level, mean household size, and sex of household head), indicators which are used frequently as proxies or correlates of development. We also considered whether each household would have made it into either the DHS or the census and in which shape.

Results

DHS data analysis
The household level of the DHS data allows to distinguish between the de jure members and the de facto ones. In the de jure the DHS includes the people who usually live in the household but have not slept there the night before. In the facto they include visitors who do not usually live in the household but slept there the night before.
Possible examples of people not sleeping in the household might be polygamous husbands, or husbands working elsewhere temporarily in the field, boarding school children, people absent due to temporary trips to visit relatives or hospitalisations as long as the absence is shorter than three months.

The description of the overall sample and the subsamples for those who did not sleep there the night before but listed de jure, female headed household, and single person households is reported in the appendix. For each subsample we considered age, sex, relationship to head of household and residence.

In total 700 households (7.2%) are single person household. They are usually concentrated in urban areas and have a higher percentage of female headed household. Female headed households are 1164 (11.3%) of these 557 are reported to be currently married/living together. Of these 199 have their husband living elsewhere and 421 are in polygamous unions. They are usually poorer

(higher percentage on low wealth index) and with a higher percentage of female members.

In table 1 we explore the key indicators for three subsamples of the DHS and in particular the de jure, de facto and the female headed household. The key indices that seem to be most affected by the different samples compositions seem to be the dependency ratios and the mean age. The former in particular is higher for de facto and female headed households possibly due to a higher presence of children or due to the fact that in the de facto men of working age are often absent due to work related reasons. Further analysis of the various scenarios which will include single person households and polygamous unions ones will be conducted.

Table 1 DHS subsample scenarios

	% included	mean age sample	Dependency ratio	Mean years of HH education	% female population	Age head of HH	household size	# people per room
Total Sample		25.05	1.26	4.88	51	44.73	5.13	2.48
de jure	96.3	25.09	1.35	-	50.8	44.68	4.94	2.3
de facto	93.2	21.74	1.55	4.33	51.8	39.99	4.78	2.4
Female HHH	11.3	23.37	1.47	3.01	61.2	45.59	4.58	2.42

Table 2 shows the key indicators of the sample collected in the Tanzanian fieldwork. We also report how the indicators would have looked like if we apply the DHS definition or the census definition to the data. Firstly 83 people out of 576 would not have made it into a survey if we were to apply the DHS definition. In addition 152 individuals would have been recorded elsewhere in the census (a few would have been recorded in the Kenyan census). In this case the differences are more striking as they represent the 'real' population. Firstly the number of households would have doubled should we apply the DHS definition and more than doubles with the census one. This is mainly due to the enlarged structure of households as considered by the respondents which might include more than one housing structure. The link might obvious in terms of socio and economic transfers but would not strictly represent a statistical unit for teh purposes of enumeration. Secondly the mean size of the households is drastically reduced if we apply a sleeping aor eating together definition. The number of single person households increases dramatically and so does teh eprcentage of female headed households. There is a need to further explore what are the possible implications of an inflation of single persons household (harder to enumerate) and female headed households (usually considered poorer and used as an indicator of deprivation). The dependency ratios increase when we apply DHS or census definitions due to the fact that working age men might have been counted in the fieldwork data. In addition older people might not be fully captured in the DHS or census data as the mean years of education shows a slight increase. In the fieldwork we found many elderly which move from household to household or which are counted in even though they leave in shacks.

The fieldwork data will further be explored and in particular we will analyse the implications of these results for the analysis of DHS data.

Table 2 Fieldwork scenarios

	Number of households	Number of individuals	mean size	Percentage female Headed Household	Single person HH	HHH mean years education	Dependency ratio
Fieldwork	52	573	11.23	27.5%	2%	6.67	1.11
DHS definition	104	490	5.86	41.9%	23.1%	7.17	1.20
Census definition	133	421*	5.64	46.3%	27%	7.18	1.27

A way forward?

The aim of this study is not to redefine the household but to consider how common household definitions might impact on both data analysis/collection and policy outcomes. We think that there is no need for a new definition and the effort should be both on flexible data collection and 'creative' data analysis.

Far from being exhaustive these are possible direction:

From the data collection point of view

- Collect information on who resides in the household as reported by the respondent before being selected for the main part of the questionnaire
- The DHS, for example, uses the households to select the individuals. The first part could be expanded to include more information
- Collect data in more sensible way that allows better configurations
- Include information on who slept there the night before, who ate and possibly on contributions to the household economy
- Relationship to household head
- Line numbers and relationship to each other
- Where possible and in particular for specialized surveys avoid assumptions of crisp boundaries – allow multiple membership of HHs and find ways to record it (e.g: Hosegood &Timaeus).

From the data analysis point of view:

- Education of users: more background material on the issues surrounding the impact of the household definition
- Careful interpretation of the results
- Non-technical language to educate policy makers on the interpretation of the data
- Methodological material available to users
  - Warnings from users' manuals
  - Make better use of the household recode of the DHS survey when analysing individual files
  - More methodological research into the use of households needed
  - There is a limited literature on the impact of the definition on the possible outcomes. Especially poverty mapping
  - Future research needed into how different types of respondents can influence the household's composition structure (e.g.: example of man not reporting wife's son).

Discussion and conclusions

The definition of the household used in surveys has an impact on socio-demographic variables. For example, if the cooking pot definition of household membership approach is replaced with a definition based on where an individual slept, the household age and sex structure change considerably. Both the dependency ratio and the proportion of female-headed households increase. Analyses of the in-depth case studies highlighted some of the processes behind

understanding these impacts, including how individuals self-define their household compared with survey-imposed definitions. In particular, high levels of individual mobility and complex socio-economic relationships between and within households emerge as key issues in explaining the difference between survey and respondent conceptualisations of the household.

There are three levels which need to be further explored and might not be identified by this analysis

1. Households who don't make it into the survey
2. Individuals that don't even make it into the de jure household
   o polygamous men
   o boarding schools
3. Individuals who live in households and don't make it into the survey

In addition we can identify an over-arching issue of individuals who can justifiably be members of multiple households such as:
- children (multiple carers and meals)
- young adults – mobility
- household head eg: polygamous man

This study has two implications. Firstly, that household survey instruments might be adjusted to better capture a range of lived realities. Secondly, information that will help survey analysts to better understands and interpret household survey data. Further analysis will highlight further potentialities with the DHS data analysis. In addition the study will explore the implications within a wider context for data collection in both developed and developing countries.

Appendix: Descriptive analysis DHS samples
Table 2 DHS Overall sample

Place of residence	%	Type Residence	%	Sex head of household	%	Wealth index	%
Capital, large city	4.4	Urban	22.2	Male	75.8	1	18.8
Small city	6.2	Rural	77.8	female	24.2	2	19.6
Town	11.4					3	19.1
Countryside	78.0					4	22.4
						5	20.1

Table 3 DHS subsample of people that did sleep in the household (but reported de jure)

Place of residence	%	Type Residence	%	Sex of household member	%			Wealth index	%
Capital, large city	3.7	Urban	23.6	Male	59.4			1	16.9
Small city	6.6	Rural	76.4	female	40.6			2	16.9
Town	13.3							3	18.2
Countryside	76.4							4	24.4
								5	23.6

Table 5 DHS sample single person household

Place of residence	%	Type Residence	%	Sex household member	%	Wealth index	%
Capital, large city	8.3	Urban	34.0	Male	58.4	1	18.1
Small city	10.1	Rural	66.0	female	41.6	2	21.9
Town	13.9					3	16.7
Countryside	67.7					4	17.1
						5	26.1

Table 6 DHS sample female headed household

Place of residence	%	Type Residence	%	Sex household member	%	Wealth index	%
Capital, large city	3.6	Urban	22.9	Male	38.8	1	22.3
Small city	6.7	Rural	77.1	female	61.2	2	20.2
Town	12.1					3	20.6
Countryside	77.6					4	20.0
						5	16.9