



Survey Households: an e-conference

November 2009

Household survey analysts neglect the household

Summary

Household surveys are the mainstay of micro-level data for developing countries. They are particularly important in resource-poor country settings where data cannot be produced by routine or administrative data systems. They are vital for poverty measurement and there has been an upsurge in the number of household surveys conducted post-1950. The rise in household survey data has been reflected in an explosion in published research that analyses household survey data.

Our question: Do end-users of household survey data ever consider the implications of the definition of the household for their analyses or results?

Our conclusion: Rarely

On what do we base this statement?

We searched major online research databases (JSTOR, Scopus, Web of Science) for articles published between 2003-2009 that used household survey data in their analyses, involving either analyses of primary [collected by author(s)] or secondary [collected by others] data. We accessed 2,432 articles and identified whether they considered the definition of the household used in their analyses.

Just 3% (n=64) of all these articles considered, to greater or lesser extent, what was represented by the household in the survey they analysed. We conclude that these academic and scholarly end-users of data exacerbate and perpetuate the problem of an uncritical approach to data produced by household surveys. Given the convenience created by harmonisation of major cross-national datasets such as the Demographic and Health Surveys (DHS) and the Living Standards Measurement Surveys (LSMS), we argue that such uncritical use of household survey data is increasing.

Questions:

1. Why do analysts of household survey data rarely consider what exactly it is that they are analysing?
2. How might analysts of household survey data be better trained in order to understand how statistical concepts such as “the household” map onto local realities?
3. What kinds of materials would help users to consider the shortcomings of household surveys?

Discussion

Participant A

1. Why do analysts of household survey data rarely consider what exactly it is that they are analysing?

Mostly people assume that household is a standard word although they know that this concept is differently defined over the time and across the world. In addition, scholars seem to be more interested in running statistic frequencies and models rather than starting to read the manual of the survey and the concepts used.

2. How might analysts of household survey data be better trained in order to understand how statistical concepts such as “the household” map onto local realities?

Comparison of definition and methods used to capture households and their implication on the outcome could contribute to understand the problem.

Participant B

I think that it is important to distinguish between definitions that are produced for purposes of survey administration and for purposes of analysis. Standard definitions of what constitutes a 'household' in surveys are produced mainly in order to facilitate administration of a standardised procedure in a practical and cost-effective manner. Without this, surveys would not be feasible. These 'process definitions' are not, generally, produced with the intention of creating an analytical variable. To create an analytical definition of the household which incorporates all the fuzzy boundaries and dynamics over time would require practically the entire questionnaire, leaving no space for other substantive research questions. In my view, this is why analysts often end up using crude definitions of 'the household', because that is all there is in the data set. Just like most survey variables!

Sara Randall

It is not a problem if this difference between the 'statistical household' (van de Walle) and fuzzy real households is recognised. Surely the problems arise:

- (a) out of the fact that the same word 'household' is used for both
- (b) because the analysts take the survey defined household (because there is nothing else) but they then analyse data and write up their analysis as though this unit were a good representation of reality.
- (c) report users, readers of papers, policy makers etc then see the analysis and the word 'household' and assume that the analysis refers to something approaching reality.

So should we find a new term instead of 'household' to be used for the survey 'statistical household' to reduce ambiguity and / or can we find ways of highlighting the mismatch to end users. At present there is little clarity and transparency.

I think there may be a further issue in some African (and quite probably elsewhere) contexts which is that the survey defined 'household' has become an entity in its own right - that respondents know what is required of surveys / censuses and who is to be

included - but that does not necessarily match well with any other aspects of daily economic life or domestic group organisation

Participant C

The notion of "household" changes whether it is for a census, or a survey for cross-sectional or longitudinal analyses. What makes analyses very difficult is the change of definition in time and space.

In Mexico, for surveys, the term "hogar" is used to designate a household. It is defined as all of the people who usually live in a particular residence and who share common expenses mainly for meals (and who can belong to the same family or not). For the census, there is a distinction between the extended household, two parent household, complex household, unrelated co-resident household, family household, indigenous household, nuclear household, [ménage no familial? We can't translate this], single parent household, and single person household.

In France, since 2005, the definition of a household, in the sense of the household surveys conducted by INSEE (French national institute of statistics and economic studies) has been substantially modified. A household is considered to be all of the people (related to each other or not) who regularly share a single place of residence (whether or not this is their main place of residence) and who have a common budget. The usual place of residence is the accommodation in which one usually lives. Therefore those belonging to the same household are people who share the same budget, that is to say, they bring in resources used in the expenses devoted to the life of the household and/or who simply benefit from these expenses.

In Spain, the definition has also been modified a little. The household is defined as the person or all persons who occupy the main family residence or part of the residence, and consume and/or share the meals or other goods sharing the same budget. It is considered that the household members are the persons who usually reside in the same surveyed unit, with the same budget (economic dependence) that is to say who share income and expenditure with the other persons who make up the unit and are present.

According to Eurostat, on the community level, a household is defined from the perspective of shared residence and of its organisation, comprising a person living alone or a group of persons not necessarily related living at the same address with a common domestic economy, that is to say generally sharing their meals or sharing a sitting or a living room. Not all countries stick strictly to this EU definition. The persons residing in the household at the time of the survey, the persons temporarily living in institutions (hospital, full time education, military service) or those absent for work or travel reasons are included in all countries. Nevertheless, in Denmark, people in hospitals are excluded. The two private household concepts are defined in the census recommendations, namely co-residence (concept of the dwelling-household) and co-consumption (concept of the economic-household), these are both used in different countries and different surveys.

The analysis of the household panels which follow individuals and observe the households to which an individual belongs seems a very attractive idea to me. The European households panel was conducted in several survey waves, but analysing the

data is very long and complex, especially in developing countries. Analysis becomes even harder when interpreting results when trying to follow up units because these units "move". Hence we can only use them for cross-sectional analysis.

I think that variables should be identified in order to be able to classify them afterwards according to their links:

- 1- ID of the censused family to whom a particular individual belongs
- 2- ID of the economic family to whom a particular individual belongs
- 3- ID of the household to whom a particular individual belongs
- 4- ID of the individual in question (from a certain age for children)

Ernestina Coast

Participant C, thanks for your interesting and diverse examples, the fraught issue of harmonisation for international comparability notwithstanding. The issue of the household over time is particularly interesting and focuses the issues that we are concerned with. For example, some would argue that there is no such thing as a longitudinal household. Indeed, if we look at the rules for membership of a household in most longitudinal studies, we see particularly clearly the issue of the "statistical household". That is, a household that is clearly defined from the perspective of the survey designer/collector in order to take account of change over time, but one that might have relatively less analytical use as time goes on.

Tiziana Leone

Participant C, your suggested IDs are very interesting. How do you see 1 different from 3? do you think the notion of family might help? Are you suggesting a record linkage of administrative, survey and census data?

Participant D

1. The concept of household is not clearly understood by some researchers. For example they do not distinguish between family and household. In a nutshell the definition of household has the following leads: (a) people must live together, (b) share economic resources, to the extent possible, (c) they commonly eat together. This is how some UN guideline defines a household especially those in developing countries e.g. in Africa. I am not sure academic researcher use the above household concept in interpreting their research findings. It is common to interchange the concept of household and family. However, the two in many cases may not overlap. In Africa, for example, there are extended families who may belong to different and many households.

2. Analysts should clearly understand and comprehend the definition of household, in use, before making inferences. They should review the literature on the definition of households. Because the definition may vary by regions and circumstances, it is important that comparative studies be carefully done. Again the analysts should clearly distinguish the concept of family and household. In many empirical studies in Africa, for example, the concept of family is more difficult to use in data analysis and interpretation.

3. Materials which users should use to critique household surveys would include publications on household surveys designs. These would indicate the sampling design up to the ultimate stage of household. Users should therefore examine the definition of household and the interpretation of results as they relate to households. Read UN

publications such as those under the Household Survey Capability Programme, in the 1970s., the Principles and Recommendations of Population and Housing Censuses, Revision 2. The significance of such material is that it will guide the user of household survey results whether they are correctly interpreted. For the household survey practitioners it is important to clearly define the concept of household and make rational interpretation of results.

Ernestina Coast

Participant A, based on your experience, what would be a good way to alert analysts of household survey data to the need to consider how the household has been defined in the survey(s) that they are analysing?

Tiziana Leone

It's true that a lot of the 1970s material do include those kind of methodological issues but going through more recent publications we increasingly find that less and less attention is given to the analysis' guidelines. More specifically we have more work on data quality and issues related to the analysis for the World Fertility Survey than the DHS. Another example is the UN manual on sampling surveys in developing countries. There is no focus on the issues surrounding the household. Possibly because cumbersome? How could we help analysts in analysing the data we already have taking into account its limitations when it comes to define the 'household'?

Ernestina Coast

Our research (and that of other people) shows that, frequently, eating + sleeping + sharing economic resources do not overlap neatly as a statistical definition of a household implies. In practical terms, the non-overlapping of these 3 activities means that interviewers "negotiate" household membership, which might well not reflect the clear definition given in manuals etc. Might a practical solution be to routinely collect de facto and de jure information around these 3 key dimensions, in order that analysts are able to apply a range of household memberships to their analyses?

Participant E

The household as "collective unit" has become so ubiquitous that for many it is used without thinking; and it has become the "analysis unit" that follows from it, the relevance of which we no longer think is even worth questioning.

For the analysis of existing data sources, and relying on the household, only a (socio, eco) knowledge of the concerned societies may allow to question the adequacy of the household to the socio-economic reality; which, as a matter of fact, is done all too rarely.

In the spirit and the extension of your project, a serious task just as a speech for the defence, has to be conducted, starting with concrete examples, bringing to light the problems and the pernicious effects which all this can produce in terms of action, and therefore making analysts and policy makers more aware...

Sara Randall

First I would like to thank all the participants in this section of the e-conference for your observations and points. What comes across clearly is that most of you are the people involved in setting up definitions (often for your own surveys) and that you are very aware of many of the problems we have outlined.

What is extremely significant is that NO-ONE has contributed to the third discussion group: ' how can academics / analysts convey their findings about the problems with 'household' to non-specialists?

Yet... we invited many of these non-specialists who were respondents in our research interviews to participate. We have tried to reach other such people through an extensive invitation list. No-one seems to be interested in this major problem from either the perspective of the technical experts nor from the user-community. This is the critical issue but what can we do about it?

Participant F

1. Because it would be possible that some people and the others consider that data at their disposition are assumed dependable and do not judge useful of days cause technical approaches used to collect data
2. The analysts of data have to participate in sessions of training by line; be part of networks of the specialists of the domain of inquiries (ISI, etc).. They can also participate in the comprehension of inquiries concerning household in a country, a region, etc..
3. It would be necessary to create more scientific Internet sites in the field of inquiries to household to allow the analysts to achieve available materials. Magazines in domain of households survey exist for example those produced by International Statistical Institute.

Ernestina Coast

Participant F, Many thanks for your very useful response. We'll certainly follow up on your suggestion. Your suggestion that people who use data make assumptions about the data they are using certainly reinforces our understanding of the situation.

Participant G

Some information (consumption or behaviour) can only be obtained by questioning people, individually or in groups (households), whilst knowing that the obtained data are tainted with errors (omissions, negligence...)

Historically, surveys were made with households for economy reasons: just one trip for a surveyor allows one to collect information on all the members of the household. This grouped collect leads to knot effects and to imprecisions connected to proxis. Bias tends to augment with the increase of the rates of refusal. Therefore, everytime it's possible, we turn to administrative data.

This cost constraint is less obvious when the survey is conducted by phone or by Internet. These modes of contact introduce new problems in turn (sampling, quality of the data...)