RANDOMISING WITHIN-HOUSEHOLD RESPONDENT SELECTION

Economic Development Initiatives (EDI) Limited

Johanna Choumert Nkolo*, Henry Cust, Marie Mallet, Callum Taylor, Linda Terenzi

**Corresponding author*: j.choumert.nkolo@surveybe.com

Economic Development Initiatives
EDI Group

RANDOMISING WITHIN-HOUSEHOLD RESPONDENT

Abstract

Collecting high quality data in developing countries is a crucial challenge for monitoring and assessing development policies. The increasing use of electronic surveys for data collection in recent years has led to a significant reduction in measurement errors and an improvement in the quality of data collected through surveys. In this paper, we provide a tool to randomly select respondents when reaching a household in the context of random walk sampling. We then provide an example using an 800-household survey we conducted in Tanzania. Our results show increased representativeness of sampled respondents within households.

Keywords

Sampling; Random walk; CAPI; Intra-household selection

Acknowledgements

We would like to thank the respondents who have agreed to participate in the survey carried out for this study. The authors are also grateful to EDI field teams, surveybe software developers and the Research Ream of EDI, particularly Koen Leuveld. The data used in this article was collected with the financial support of the UONGOZI Institute. The usual disclaimers apply.

Economic Development Initiatives
EDI Group

Contents

List of Figures

1. INTRODUCTION

Data is a tool for development. The data gap, data quality and, data measurement issues have been stressed in recent research and by various institutions, especially in Sub-Saharan African countries (Beegle et al. 2016; Caeyers et al., 2012; Glassman and Ezeh, 2014; Jerven and Johnston, 2015). Having reliable and accurate data is instrumental to monitoring and evaluating development policies and programmes. More resources should be allocated to high quality surveys in African countries, and high quality methodological research on how to implement these is key (e.g. Arthi et al., 2018). Where listing households' members prior to surveys is not possible, intra-household selection of respondents is an important consideration. Should interviewers interview the person who opens the door, or the household head, or any member? Restricting the probability of selecting certain household members for survey participation can lead to bias. This respondent selection bias has been widely discussed in the literature. For reviews of methods and debates, see Demombynes (2013), Gaziano (2005), Rizzio et al. (2004) and Yan et al. (2015). The recent meta-analysis on within-household respondent selection carried out by Yan et al. (2015) clearly states the problem: "*Random selection of a respondent within a sampled household is essential for maintaining the probability nature of the resulting sample and for making inference from the household sample to the general population*" and "*A good within-household respondent selection method should be able to randomly select a respondent within the household without appearing intrusive or burdensome to potential respondents*".

In addition, they summarise the absolute bias of different intra-household selection methods, defined as the percentage point difference between a sample's characteristics and the matching characteristics from the population the sample is drawn from. The paper breaks down the different methods into probability methods[1], quasi-probability methods[2], non-probability methods[3], the Rizzo-Brick-Park method[4], and convenience method.[5] They find with respect to gender there is between 2.9% and 8.5% absolute bias. Convenience methods provide the worst bias, at 8.5%, with non-probability methods providing the best at 2.9%.

However, before intra-household selection there is also the challenge of sampling. The sampling stage's rigorousness is central to the robustness and precision of any results obtained using field data (see United Nations, 2005; Lohr, 2010). Conducting full household listing is the most robust way of providing a randomised sample of the population, however it can be difficult and costly. As an alternative, researchers and survey managers can rely on simplified listing, using

[1] Probability methods require listing of all people in a sampled household and computing individual level probability of selection. Examples are the Kish Method (Kish, 1949), Age-order or Age -only methods (Denk and Hall 2000; Forsman, 1993)

[2] Bypasses full listings to reduce administrative time and intrusiveness. Examples are the next or last birthday methods (Salmon and Nichols, 1983).

[3] This streamlines the selection process by approximating population age and gender distributions. An example is the Troldahl and Carter method and its variations (Troldahl and Carter, 1964).

[4] Selects respondents based on the number of adults in a household. The method is less intrusive because only information needed is the number of adults in the household (Rizzio et al., 2004)

[5] It may be the first person to answer the phone in a phone survey; the person who replies to the mailed questionnaire, etc.

community leaders to produce lists of local households. Once the listing of households is complete, researchers can then randomly select the sample for participation in the survey.

Moreover, where listing is not possible, researchers often rely on the random walk (RW) sampling approach to select households. This approach is less expensive to implement than full household listing, or the tracking of respondents using an existing list. The drawbacks to the RW method are that it might fall into non-probability sampling methods and that interviewers' behaviour may bias the sample, i.e. respondents may not have the same probability of being selected because of interviewers' choices. Where listing is not possible intra-household selection is vital for minimising unobservable bias. Therefore, researchers must ensure that each RW protocol is tailored to their specific research project.

In this paper, we provide a fully explained tool and method for performing randomised selection of respondents within a household alongside evidence of its potential effectiveness with respect to the methods set out in Yan et al. (2015). We first present a module we developed in the computer assisted personal interviewing (CAPI) software surveybe to perform randomised within-household respondent selection. Second, we illustrate our methodology using data we collected for an 800-household survey in Tanzania in 2016. Third, we provide an analysis of the benefits our method has on sample representativeness. Finally, we conclude and provide some extensions of the use of the tools. We argue that implementing a rigorous random walk is possible when using the appropriate training tools for field teams and that quality control procedures, complemented by our randomised respondent selection method, is the optimal way to reduce bias in face-to-face survey results.

2. METHODOLOGY

2.1. THE USE OF CAPI

Surveybe is a new generation CAPI software developed by the EDI Group[6] that gives researchers and survey managers direct control of survey design, provides validated responses, data cleaning at the point of collection, time saving efficiencies in all phases of their projects and encryption-protected, analysis ready, data. Surveybe has been used for various field experiments in dozens of countries around the world (e.g. Caeyers et al., 2012).

We used surveybe to implement an 800-household survey on the perceptions of natural gas discoveries in Tanzania. The survey was conducted in November and December 2016. It consisted of a one-hour questionnaire where households were selected following a RW protocol.

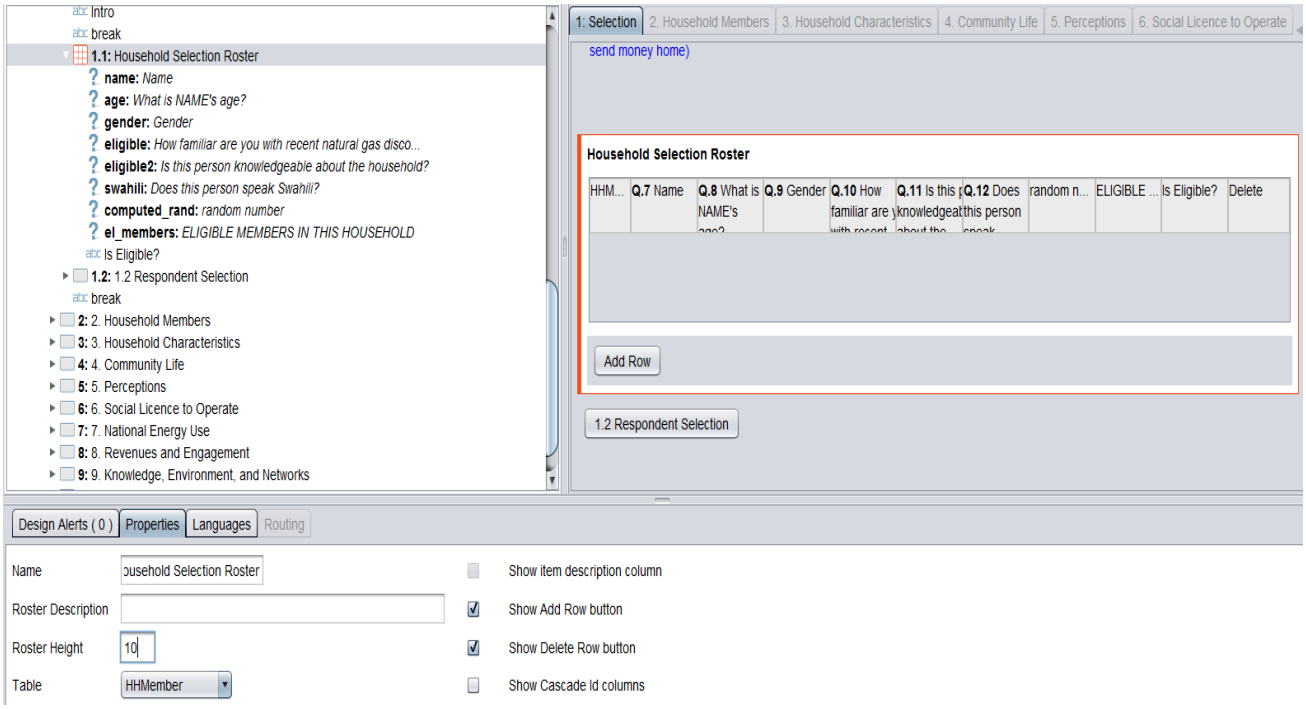
2.2. SURVEYBE METHODOLOGY

The first step required for randomised selection of respondents is to create a household roster containing questions relevant to the selection of household members (e.g. name, age, gender). This can be done quickly and easily using the intuitive surveybe designer software. The roster will then be used to collect data on each household member and generate random values which can be used for selecting household members.

For the aforementioned household survey, the household roster contained questions on name, gender, and four further eligibility criteria: age, how knowledgeable the person is about natural gas, whether or not the person is knowledgeable about the household, and whether or not the person is able to speak Swahili, to determine if they can complete the interview (See Figure 1).

[6] See more information at https://www.edi-global.com/

Figure 1. Creation of a household roster – Illustration in the surveybe designer



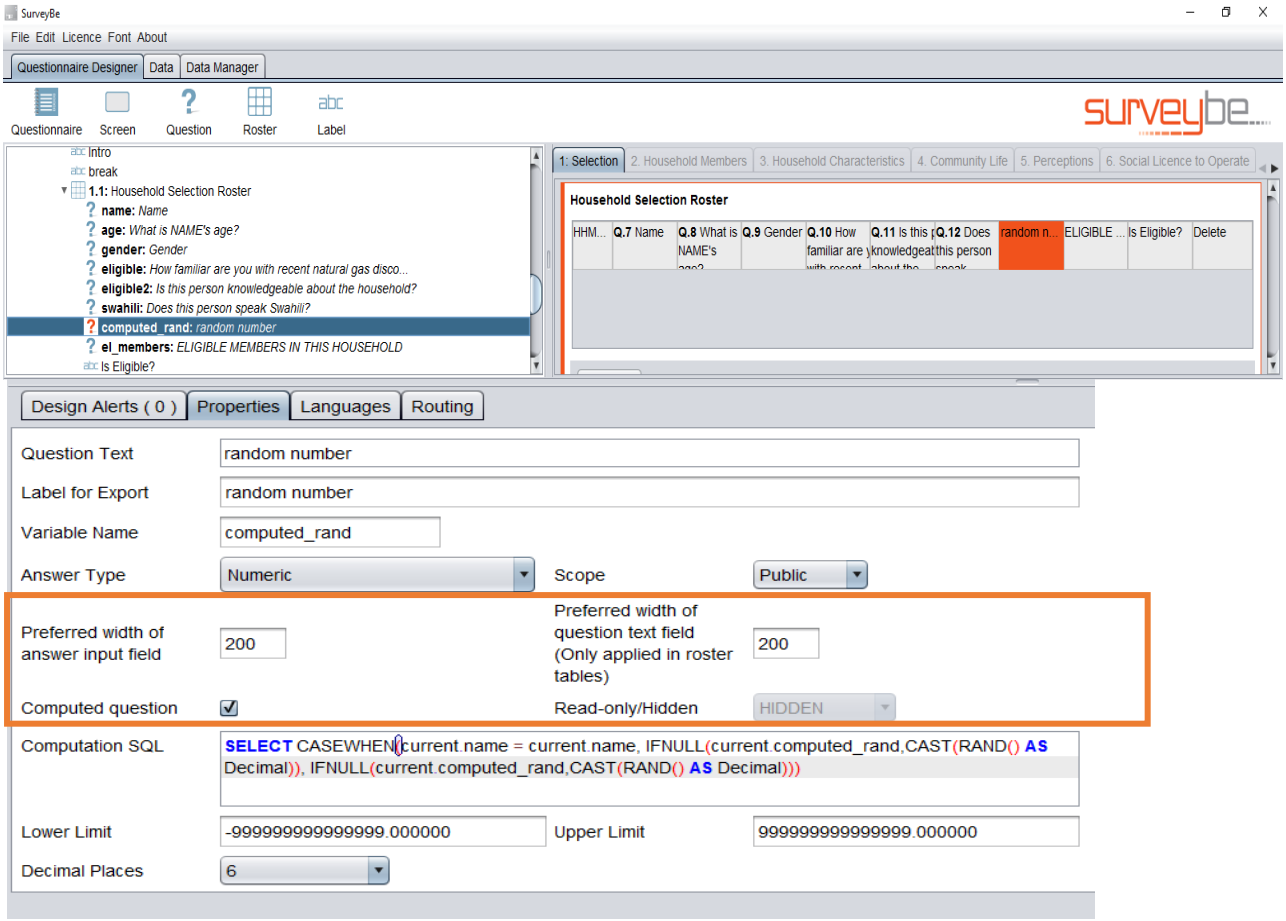
The next stage, after completing the construction of a normal household roster, is to configure the randomisation tools. First, a random number variable should be configured using surveybe's computed question functionality. Computed questions are hidden from the user during the interview but are exported to the final datasets. Typically, computed questions use SQL to calculate a data point, based on other pieces of information. They are therefore useful for calculating answer values that the interviewer or respondent should not have an impact upon (for example calculating the total consumption based on the responses from item purchases or calculating the size of the household based on the number of rows added in a household member listing roster). In this case, we will instruct the computed question to generate a single random number between zero and one, independent of any other answers in the questionnaire. We use the following SQL syntax (Figures 2 and 3):

Figure 2. Generating a random number using SQL syntax [7]

```
SELECT CASEWHEN(current.name = current.name, IFNULL(current.computed_rand,CAST(RAND() AS Decimal)), IFNULL(current.computed_rand,CAST(RAND() AS Decimal)))
```

- **CASEWHEN (current.name = current.name)**: Instructs the programme to generate the random number for each row of the table. i.e. for each household member (this statement can never be false);
- **IFNULL**: Instructs the programme to only generate a random number if a random number has not already been generated;
- **CAST((RAND() AS Decimal)**: Instructs the programme to compute a random number as a decimal, and insert it as the answer to the computed question variable.

Figure 3. Generating a random number using SQL syntax – illustration in the surveybe designer



[7] Some SQL is tailored to surveybe therefore may not appear as regular SQL syntax.

This provides the tools needed to generate random selection of household members. The next step is to instruct surveybe to randomly select the household members. This requires the use of another computed question, which references the random numbers we have just assigned to each household member. In this survey, we allowed for two potential respondents before moving to the next household. The SQL codes for randomly selecting the first and second respondent are as follows (Figures 4 and 5):

Figure 4. SQL code to randomly select the first eligible respondent

```
SELECT CASEWHEN(((SELECT COUNT(*) FROM HHMember WHERE el_members=1)>0), (SELECT TOP
1 name FROM HHMember WHERE el_members=1 ORDER BY computed_rand ASC),
'<font color=red> (HAKUNA)')
```

Figure 5. SQL code to randomly select the second eligible respondent

```
SELECT CASEWHEN(((SELECT COUNT(*) FROM HHMember WHERE el_members = 1 ) > 1), (SELECT
TOP 1 name FROM HHMember WHERE el_members=1 ORDER BY computed_rand DESC),'(HAKUNA)')
```

- **SELECT COUNT(*) from HHMember where el_members=1)>0**: Only select a household member if there is more than 0 household members who are eligible;
- **SELECT TOP 1 name**: Select the first name in the list;
- **WHERE el_members=1**: Select the first name in the list where the member is eligible. (*el_members* is a binary variable equal to 1 if the household member is eligible for the survey);
- **ORDER BY computed_rand ASC**: Order the list by the variable computed_rand, in ascending order for the first randomly selected respondent;
- **ORDER BY computed_rand DESC**: Order the list by the variable computed_rand, in descending order for the second randomly selected respondent;
- **'(HAKUNA)'**: This is the result returned if there is no eligible respondent, i.e. when the condition '**SELECT COUNT(*) from HHMember where el_members=1**' is equal to 0 ("HAKUNA" meaning "None" in Kiswahili [8]).

Because the variable *computed_rand* is a randomly generated decimal, selecting the top value from this variable gives a random selection from all the random numbers generated in that household. The final step is to create a label (see Figures 7 and 9), which uses SQL to display the names of the randomly selected household members, generated in the computed variable above. This provides a robust method for randomly selecting respondents and prevents any interviewer interference in the selection process. It should be noted the number of respondents selected can be any number one chooses.

[8] The entire questionnaire is translated into Kiswahili and any translated language can be chosen at any time within surveybe.

Figures 6 and 7 are an illustration of a household where there is only one eligible member. In Figure 6, the household is composed of three members and David is the unique member meeting the eligibility criteria represented in questions Q10, Q11 and Q12. Figure 7 represents the instructions (a dynamic label) to the interviewers according to the result of the eligibility and random selection from the household listing roster in Figure 6. As there is only one eligible respondent, secondary respondent displays “HAKUNA”, and the question about the availability of a second respondent is disabled.

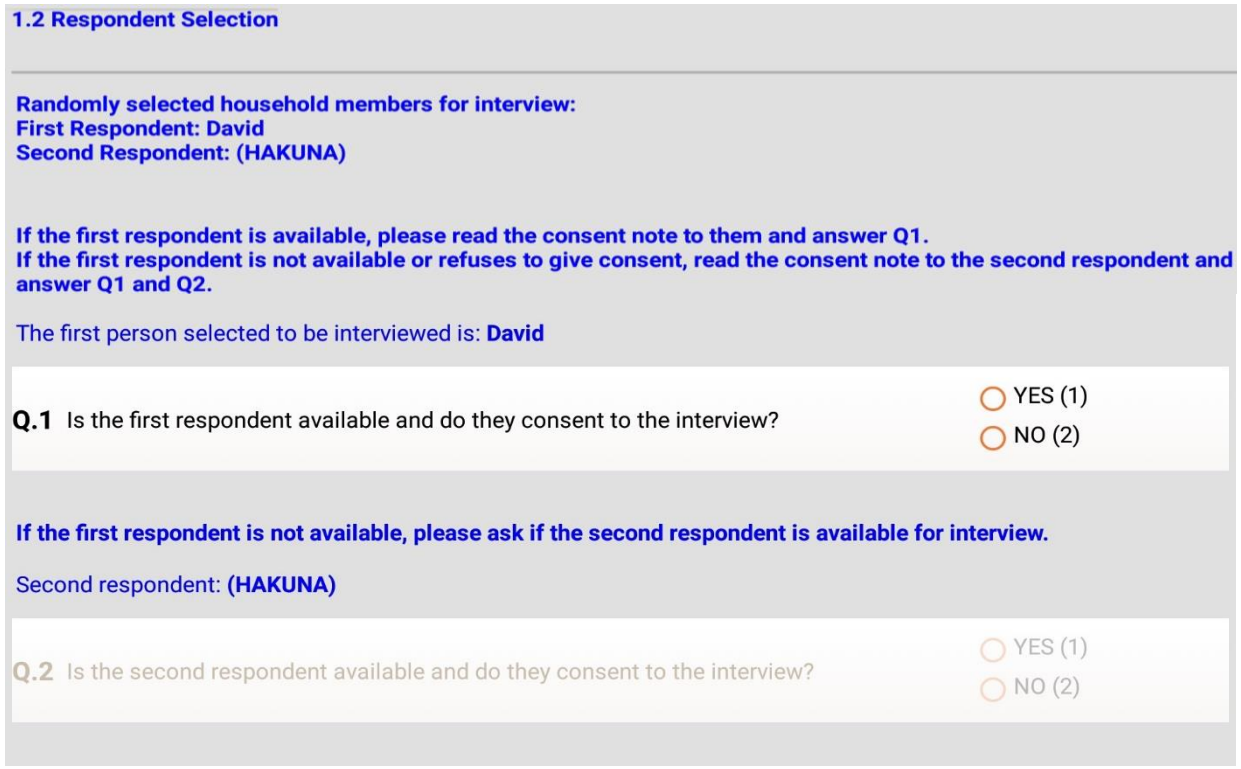
Figures 8 and 9 are an illustration of a household where there are two eligible members. The household is composed of three members but with only David and Sarah as eligible members. Figure 9 represents the instructions to the interviewers according to the result of the eligibility and random selection from the household listing roster in Figure 8. In this case, Sarah is designated as the primary respondent and David as the secondary.

Figure 6. Household member listing in android – case with one eligible respondent

Household Selection Roster

ID	Q.7 Name	Q.8 What is NAME's age?	Q.9 Gender	Q.10 How familiar are you with recent natural gas discoveries in Tanzania?	Q.11 Is this person knowledgeable about the household?	Q.12 Does this person speak Swahili?	Is Eligible?	Delete
1	David	46	Male (1)	I am fairly familiar with gas activities (4)	YES (1) / NO (2)	YES (1) / NO (2)	Eligible	X
2	Sarah	44	Female (2)	I have never heard about it (1)	YES (1) / NO (2)	YES (1) / NO (2)	Not Eligible	X
3	Adam	16	Male (1)	Please select...	YES (1) / NO (2)	YES (1) / NO (2)	Not Eligible	X

Figure 7. Respondent selection – case with one eligible respondent



1.2 Respondent Selection

Randomly selected household members for interview:
First Respondent: David
Second Respondent: (HAKUNA)

If the first respondent is available, please read the consent note to them and answer Q1.
If the first respondent is not available or refuses to give consent, read the consent note to the second respondent and answer Q1 and Q2.

The first person selected to be interviewed is: **David**

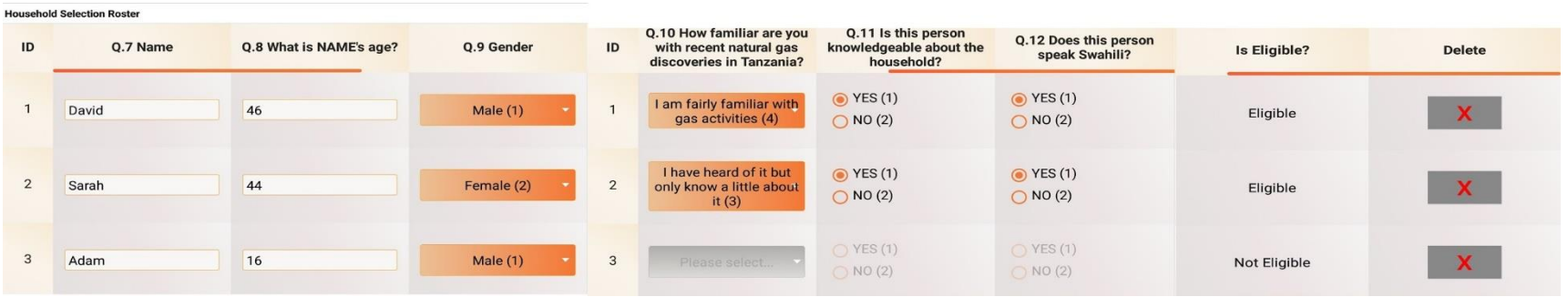
Q.1 Is the first respondent available and do they consent to the interview? ◯ YES (1) ◯ NO (2)

If the first respondent is not available, please ask if the second respondent is available for interview.

Second respondent: **(HAKUNA)**

Q.2 Is the second respondent available and do they consent to the interview? ◯ YES (1) ◯ NO (2)

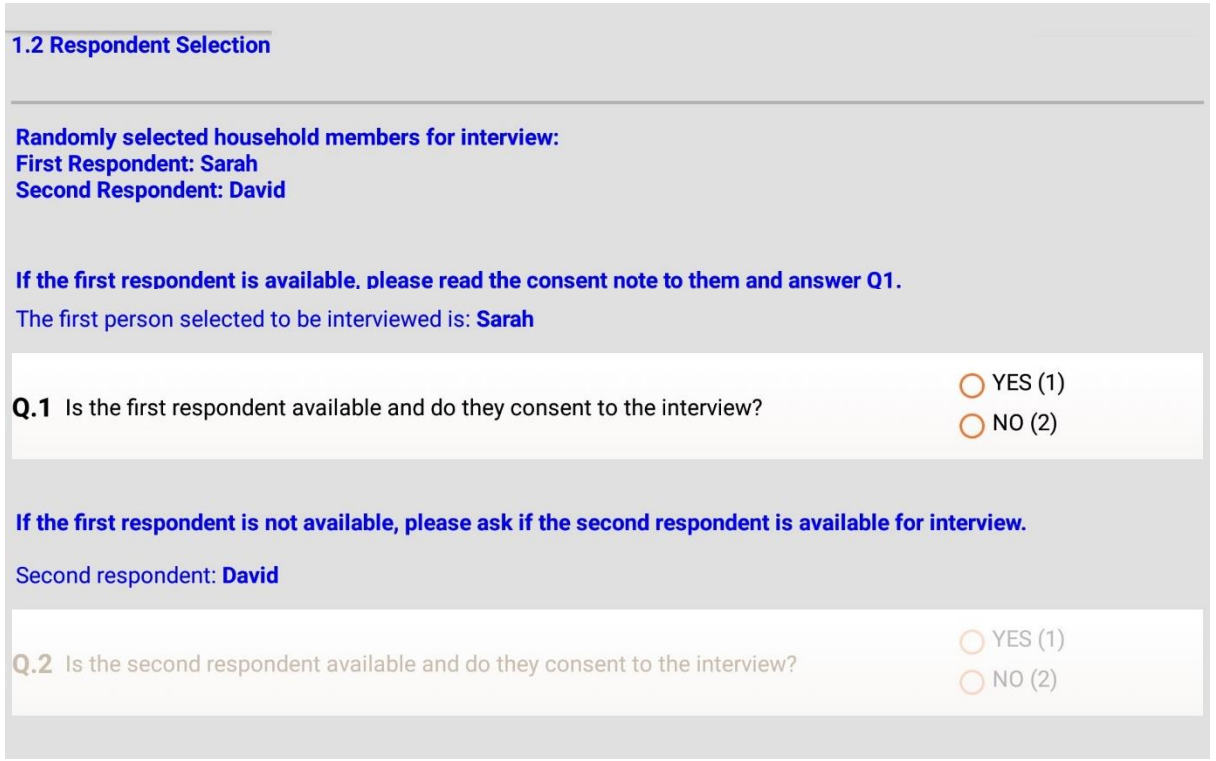
Figure 8. Household member listing in android – case with two eligible respondents



Household Selection Roster

ID	Q.7 Name	Q.8 What is NAME's age?	Q.9 Gender	ID	Q.10 How familiar are you with recent natural gas discoveries in Tanzania?	Q.11 Is this person knowledgeable about the household?	Q.12 Does this person speak Swahili?	Is Eligible?	Delete
1	David	46	Male (1)	1	I am fairly familiar with gas activities (4)	YES (1) / NO (2)	YES (1) / NO (2)	Eligible	X
2	Sarah	44	Female (2)	2	I have heard of it but only know a little about it (3)	YES (1) / NO (2)	YES (1) / NO (2)	Eligible	X
3	Adam	16	Male (1)	3	Please select...	YES (1) / NO (2)	YES (1) / NO (2)	Not Eligible	X

Figure 9. Respondent selection – case with two eligible respondents



1.2 Respondent Selection

Randomly selected household members for interview:
First Respondent: Sarah
Second Respondent: David

If the first respondent is available, please read the consent note to them and answer Q1.

The first person selected to be interviewed is: **Sarah**

Q.1 Is the first respondent available and do they consent to the interview? YES (1) / NO (2)

If the first respondent is not available, please ask if the second respondent is available for interview.

Second respondent: **David**

Q.2 Is the second respondent available and do they consent to the interview? YES (1) / NO (2)

3. EMPIRICAL ANALYSIS

3.1. CONTEXT AND RANDOM WALK PROTOCOL

We use data we collected for an 800-person survey gathering information about the perceptions of the natural gas industry in two coastal regions of southern Tanzania. The survey was sampled via a RW methodology from 20 randomly selected villages and mtaa (equivalent to villages in an urban setting) near to gas operations and infrastructure. We used teams of eight interviewers per village to perform the random walk in two sub-villages (kitongoji/sub-mtaa[9]) per village/mtaa with the target of performing five one-hour interviews per interviewer, per day, totalling 20 interviews per sub-village. The following is an abridged version of the random walk protocol[10]:

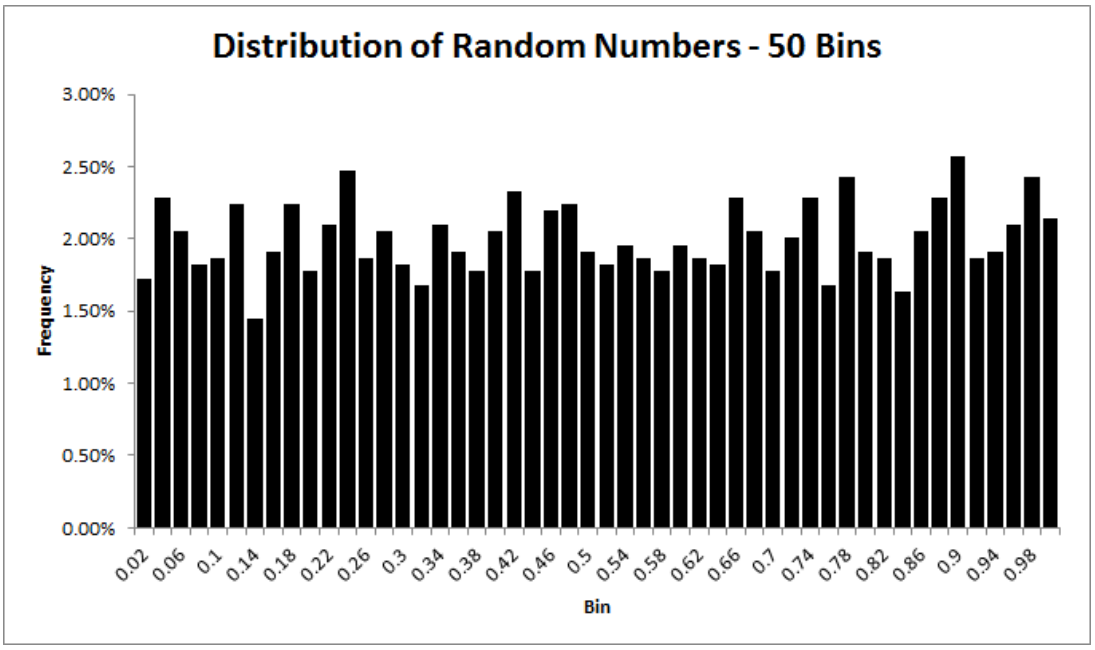
- Interviewers were allocated a random starting point in a sub-village based a number drawn from a hat that corresponds with a point on a map of the sub-village;
- Interviewers followed the same instructions counting only houses on their right and taking every 3rd right turn;
- When an interviewer came across a household and someone was available they completed the short household member listing exercise to determine the primary and secondary respondents;
- They continued along the same path once the interview was completed until they had five interviews from each sub-village.

3.2. RANDOM NUMBER GENERATOR

The listing was done by adding each member of the household to a roster (See section 2.2). Each individual in the table was assigned a hidden random number between 0 and 1. A primary and secondary respondent were chosen at random based on the highest random number allocated and the lowest from the eligible members of the household. If the primary respondent was not available to perform the interview the secondary respondent was asked. We ran a series of tests with the numbers generated from our survey to check the validity of the numbers generated. Across the 2143 random numbers generated between 0 and 1 we found a mean of 0.50625 with a range from 0.000036 to 0.999703. Figure 10 visualises the distribution of the generated numbers across 50 bins with size of 0.02 each. The expected value of each column in a uniform distribution would be 2%. Our distribution has only one column above 2.5% and one below 1.5%.

[9] Tanzanian villages are split into sub-villages called kitongoji where most villages would have one or more kitongoji. Urban areas do not always have a specific division of mtaa therefore local experts were used to divide up the mtaa when needed for clustering purposes.
[10] The full random walk protocol is available from the authors upon request.

Figure 10. List distribution of actual numbers to show the uniform distribution of numbers.



Additionally, we ran a one-sample Kolmogorov-Smirnov test to test the statistical uniformity of generated random numbers. We firmly reject the null hypothesis that continuous distribution of *computed_rand* is not uniformly distributed.

Table 1 shows the summary statistics from the first section of the survey that randomly selects the respondent from the household to interview. Only information from completed interviews (n=783) is included. The first column contains information based on the person who initially answers the door. The second contains all household members from the listing exercise, aged above 18. The third is limited to those who are eligible to be selected for the rest of the interview, which includes having at least "heard of but know nothing about natural gas", answered "yes" to "are you knowledgeable about the household" and "yes" to fluent in Swahili.[11] The fourth column is only those who completed the full interview.

3.3. Results

Our method increases the representativeness of the population by reducing the likelihood of the interview being completed by the person who answers the door. In our sample of 783 completed household interviews, the proportion of people who both answered the door and conducted the interview was 61.1%. This proportion may seem high, however 79.9% of households only had one or two eligible members meaning the person who answered the door would always be one of the selected respondents. In a household of two eligible members; if the second eligible household member, i.e. not the one who answered the door, was the second randomly selected respondent, i.e. was selected behind the person who answered the door, or was listed first and

[11] This question was added soon after field launch explaining the lower total number.

unavailable, then the first would automatically complete the interview resulting in a seemingly high number of respondents both answering the door and being the respondent for the interview.

In addition, of all those who answered the door, 54.5% were female, compared to 50.9% of total eligible members listed (including those unavailable and away from the home), showing an increased likelihood of females answering the door or being the person to greet the interviewer once they arrived at the household. After the randomisation of the respondents, 52.9% of overall respondents were female. Therefore, compared to interviewing those who answer the door, we have a sample which is more representative of the population in terms of gender. Furthermore, of those the 293 who answered the door but were not selected to perform the interview, in 30% of cases the gender changed. We can attribute the increased representativeness of the respondents taking our survey to the electronic randomisation built into the questionnaire.

Table 1. Effects of the random selection on our sample

	Person at the Door			All Household Members			All Eligible Members			Respondents Only			First Selected Respondent		
	Statistics	No.	Total[12]	Statistics	No.	Total	Statistics	No.	Total	Statistics	No.	Total	Statistics	No.	Total
Age (mean)	41.0	n/a	769	39.0	n/a	1916	38.9	n/a	1631	39.8	n/a	783	39.4	n/a	783
Males	45.5%	356	783	47.0%	901	1916	49.0%	799	1631	45.6%	357	783	48.7%	381	783
Females	54.5%	427	783	53.0%	1015	1916	51.0%	832	1631	54.4%	426	783	51.3%	402	783
HH knowledge	99.7%	767	769	96.1%	1842	1916	n/a*	n/a*	1631	n/a*	n/a*	783	n/a*	n/a*	783
Swahili**	99.9%	701	702	84.1%	1485	1765	n/a*	n/a*	1512	n/a*	n/a*	716	n/a*	n/a*	716
Eligible Members	98.2%	755	769	85.1%	1631	1916	n/a*	n/a*	1631	n/a*	n/a*	783	n/a*	n/a*	783
Never heard of natural gas	1.7%	13	769	11.7%	225	1916	n/a*	n/a*	1631	n/a*	n/a*	783	n/a*	n/a*	783
I have heard about it but know nothing about it	68.0%	523	769	68.8%	1319	1916	77.7%	1267	1631	70.5%	552	783	75.3%	590	783
I heard of it but only know a little about it	25.2%	194	769	16.6%	319	1916	19.1%	311	1631	24.8%	194	783	21.2%	166	783
I am fairly familiar with gas activity	3.5%	27	769	2.0%	39	1916	2.4%	39	1631	3.6%	28	783	2.7%	21	783
I am very familiar with gas activities	1.6%	12	769	0.7%	14	1916	0.9%	14	1631	1.1%	9	783	0.8%	6	783

* Not applicable because they are eligibility criteria

** Question added later in field work

[12] Total numbers appear low in this column due to the exclusion of children from the analysis. Gender information was taken by the interviewer for the person who answered the door, however age and eligibility criteria were only taken for those over 18. The data implies in 14 cases it was a child who answered the door.

To see where our results fit with the literature we compare them to those of Yan et al. (2015). They calculate the absolute bias of various methods by comparing their demographics to those of the population the sample is drawn from. They define the absolute bias as the percentage point difference. Of 27 papers, they find absolute biases of between 8.5% and 2.9%. We have calculated the same statistic comparing our sample statistics to the eligible population from our survey. Our survey was not intended to be nationally representative, but only representative of those who live near to gas operations for which there are no accurate population statistics. Therefore, we compare to the population we were sampling from i.e. all members of all the households we surveyed. For reference, our population is 53% female, compared to the census data for the districts we were present in which is 52.5%. As shown in Table 1 the percentage point change in our interviewed sample is 3.4% in terms of gender and a percentage change of 2.1% in terms of age. However, due to our survey selecting two people randomly from each household, there is still an availability bias. The door answerer statistics are a good indicator for the direction of this bias. If we instead look at the hypothetical selection method where only one person from each household is randomly selected, and if they are not available then the interviewer moves to the next household, we see that the percentage point change is now only 0.3% with respect to gender and a percentage change of 1.0% with respect to age.

3.4. Gender Differences

Representativeness is key because over-sampling certain demographics can lead to misleading results. For example, in our survey we found that there was a significant difference in answers to questions about what to do with revenue from natural gas production. Females were significantly more likely to choose health and education as the priority of spending needs, whereas males were more likely to focus on agricultural spending as a priority. Reduced representativeness would produce differing results and could change the conclusions of a study.

In addition, from the same question we tested the likelihood of responses based on whether the respondent was the person who initially answered the door and was chosen to answer the questionnaire compared to those who did not initially answer the door. For those who did not initially answer the door there is a significantly increased chance that they would prioritise infrastructure and on supporting the industry sector than any other option. Whilst the interpretation of this result is interesting and deserves thought, the important point it illustrates is that there is a significant difference between answers from those who answer the door and those who do not.

3.5. Is Within-Household Randomisation a Burden?

Automatic timestamps built into our survey allowed us to record precise durations for sections of the survey (See Choumert Nkolo et al., 2018). Two of these timestamps were used to measure the duration of the within-household listing and the respondent selection to measure the additional burden borne. For example, the average amount of time taken for completion of the

randomisation and respondent selection section was 5 minutes and 31 seconds, roughly 8.5% of our intended time for the full interview (one hour). This 8.5% is an upper bound of the burden considering many household surveys will already be collecting simple information about household members as part of the main survey. This section included six questions for every household member listed and seven other general questions that had to be answered only once per household. Surveybe hides the calculations and selection processes which allows us to have such an effective and simple random selection of respondents. This automised method of random selection also takes significantly less time than traditional methods, allowing more focus on the real research questions, while also reducing the data's unobservable bias.

4. Conclusion

We provide evidence that randomising respondents at the household when it is too costly to perform comprehensive listings increases representativeness and reduces bias. We find our method gives similar results to the methods supported in Yan et al. (2015) and that it has the potential, through selecting only one person per household, to improve upon these results. In addition, evidence from a real project shows the burden of such a method is very small, taking up around five minutes of a one-hour interview.

We outline clear instructions on how to use a similar method for other data collection projects which can be easily implemented in electronic surveys. In addition to the statistical advantages our method provides, there are numerous practical advantages, including a very small time burden, and use of information often collected as a part of a typical household survey. Finally, the tool offers great flexibility to randomly select people considering numerous eligibility criteria.

The tool we develop in the paper could be used in other settings. For instance, in some settings there is more than one household living in a dwelling. As such, a decision on how to sample a household here should be made using surveybe with similar programming. An example where we have adapted random selection programming is when selecting clusters within sample units (e.g. sub-villages within villages) and for different arms of surveys (e.g. male vs female only clusters).

REFERENCES

Arthi, V., Beegle, K., De Weerdt, J., Palacios-López, A., 2018. Not your average job: Measuring farm labor in Tanzania. Journal of Development Economics 130, 160–172. https://doi.org/10.1016/j.jdeveco.2017.10.005

Beegle, K., Christiaensen, L., Dabalen, A., Gaddis, I., 2016. Poverty in a Rising Africa. Washington, DC: World Bank. © World Bank.https://openknowledge.worldbank.org/handle/10986/22575 License: CC BY 3.0 IGO.

Caeyers, B., Chalmers, N., De Weerdt, J., 2012. Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment. *Journal of Development Economics* 98, 19–33. doi:10.1016/j.jdeveco.2011.12.001

Demombynes, G., 2013. What's the right way to pick the respondent for a household survey? http://blogs.worldbank.org/impactevaluations/whats-right-way-pick-respondent-household-survey (last accessed on 11/01/2017)

Choumert, J., Cust, H., Taylor, C., 2017. Using paradata to collect better survey data : Evidence from a household survey in Tanzania. Working paper

Denk, C.E., Hall, J.W., 2000. Respondent selection in RDD surveys: A randomized trial of selection performance, in: Portland, OR: Paper Presented at the Annual Meeting of the American Association for Public Opinion Research, Portland.

Gaziano, C., 2005. Comparative analysis of within-household respondent selection techniques. Public Opinion Quarterly 69, 124–157.

Glassman, A., Ezeh, A., 2014. Delivering on a data revolution in sub-Saharan Africa. Center for Global Development Brief. Retrieved from http://www.cgdev.org/sites/default/files/delivering-data-revolutionsub-saharan-africa-pdf (last accessed on 10/01/2017).

Jerven, M., Johnston, D., 2015. Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development. The Journal of Development Studies 51, 111–115. https://doi.org/10.1080/00220388.2014.968141

Kish, L., 1949. A Procedure for Objective Respondent Selection within the Household. Journal of the American Statistical Association 44, 380–387. doi:10.1080/01621459.1949.10483314

Lohr, S.L., 2010. Sampling: design and analysis, 2nd ed. ed. Brooks/Cole, Boston, Mass.

Rizzo, L., Brick, J.M., Park, I., 2004. A Minimally Intrusive Method for Sampling Persons in Random Digit Dial Surveys. Public Opinion Quarterly 68, 267–274. doi:10.1093/poq/nfh014

Troldahl, V.C., Carter Jr, R.E., 1964. Random selection of respondents within households in phone surveys. Journal of Marketing Research 71–76.

Salmon, C.T., Nichols, J.S., 1983. The Next-Birthday Method of Respondent Selection. Public Opinion Quarterly 47, 270–276. doi:10.1086/268785

United Nations, 2005. Designing household survey samples: practical guidelines. United Nations Publications, New York.

Yan, T., Tourangeau, R. McAloon, R., 2015. A Meta-analysis of Within-Household Respondent Selection Methods on Demographic Representativeness. *2015 Federal Committee on Statistical Methodology* https://fcsm.sites.usa.gov/files/2016/03/H3_Yan_2015FCSM.pdf (last accessed on 10/01/2017).