Huafeng Zhang

The 2007 Socio-economic Assessment of Ein El-Hilweh Refugee Camp

Sampling Design (Post Fieldwork Version)

Fafo

Introduction

This document describes the sample of the survey in relation to the Socio-economic Assessment of Ein El-Hilweh Refugee Camp (EHC) Project. Its main aim is to document the sampling procedures and the procedures for handling non-response in the survey.

Requirements of the sample

The design of the EHC sample was – like any other sample – subject to a number of constraints. The main design characteristics for the sample were the following:

1. The population selected for this study was all households residing in Ein El-Hilweh refugee camp and two adjacent areas.
2. The budget allowed for a sample of 1,000 households.
3. The survey contains two questionnaires: the main questionnaire and the RSI questionnaire. The main questionnaire was designed to ask one respondent in the household questions about the overall condition of the household and that would yield basic information about all household members. To complete the RSI questionnaire, in each selected household one household member should be selected randomly among all the household members aged 15+ (the Randomly Selected Individual. i.e. the RSI).

The sample frame

The sample frame for the EHC survey is based on a complete listing of all households following a housing and population census carried out by the Palestinian Central Bureau of Statistics (PCBS), the Damascus branch, in 1998 (Ugland ed. 2003:285-289). The original frame covered altogether 57 communities: 12 refugee camps and 45 so-called "gatherings". The sample frame for this survey comprises Ein El-Hilweh refugee camp and two adjacent areas or gatherings.

The census list is a complete listing of household heads in all the houses (building structures). It is organised as a file with the following variables:

Table 1: List of variables in the sample frame

Variable	Explanation
Governoret	Governorate, region
Camp	Camp (21 "Ein Al Helwi", 24 "Jabal Haleb", 26 "Baraksat")
Area	Area code in the camp
Road	Road number
Dwilling	House/structure number in the camp
Name	Name of household head

In addition to the camp itself, area 21 "Ein Al Helwi" also includes Tawareq (an areas to the north of EHC), El Sekeh (a small strip of land to the west), and Dard Seen (a tiny area south, southeast of the formal camp border). Area 24 in the frame, "Jabal Haleb" (a fairly large area to the east of the camp

– some of which is agricultural land) actually encompasses two areas: Sohoun (the northernmost part) and Jabal Haleb (the southernmost part). Area 26 in the frame, ”Baraksat”, shows up on UNRWA’s maps as a rectangular adjacent plot of land area to the north, northeast of the camp.

In total the file contains 6,426 households in 3,413 houses.

The number of households in the houses (building structures) varies from one household per house to 111 households per house. Many houses contain too few households, which makes it impossible to select enough households from them in the second sampling stage. The houses with very few households were, therefore, merged into separate strata.

Sample design

The key elements of the sampling are the following:

1. PSUs were houses in the camp.
2. PSUs/houses were explicitly stratified according to the size of PSUs/houses: all the houses with one household were assigned to stratum 1; all the houses with two households were allocated in stratum 2; all the houses with three households were allocated in stratum 3; all the houses with four to 14 households were allocated in stratum 4; one house with 15 households was stratum 5; and one house with 111 households was stratum 6.
3. Before selection, the sample frame was sorted by the size of the PSU/house.
4. The allocation of the EHC survey sample makes the sample approximately self weighting, so that the sample was proportional to the size of strata, i.e. the total number of households in each stratum.
5. All the selected PSUs/houses were re-listed.
6. From the updated list the selection of households was made with linear systematic sampling.
7. A predetermined number of household/households were selected from each selected PSUs/ houses in the strata. This number was determined before re-listing and was not changed after re-listing.
8. One household member was randomly selected from all the household members aged 15+ in each selected household to answer the RSI questionnaire.

Sample selection procedures

Selection of PSUs/houses

The selection of houses was conducted with linear systematic PPS sampling within each stratum.

Table 2: Allocation of PSUs/houses to strata

	# of houses in frame	# of households in frame	Allocation (# of houses)
House with 1 household	1,888	1,888	294
House with 2 households	760	1,520	119
House with 3 households	411	1,233	96
House with 4-14 households	352	1,674	129
House with 15 households	1	15	1
House with 111 household	1	111	1
Total	3,413	6,426	640

Re-listing of PSUs/houses

The selected PSUs/houses were re-listed. The re-listing serves two purposes: (i) to enable the selection of households, and (ii) to enable interviewers to locate the selected households. Furthermore, the sample frame is fairly old - from 1998. Although there have been restrictions on construction activities since that time, and particularly difficult to erect new houses, existing building structures may have been expanded and new households may have been added in some houses. Therefore, it is important to re-list the selected PSUs/houses in order to update the number of households currently living in each house.

All households within a selected PSU/house were listed.

Selection of households

In stratum 1, where all the PSUs/houses contains only one household in the initial estimate, only one household was selected in each PSU/house. In stratum 6, altogether 17 households were selected. In all the other strata, altogether two households were selected in each PSU/house. Linear systematic sampling was used to select households from the list of re-listed households. The sample is considered as a sample of households.

Additional households in house units

The list of households is intended to be a complete list of households in a PSU/house. In stratum 1, all the houses have only one household in the initial estimate, and only one household has been designed to be selected. Therefore, even if more than one household was re-listed afterwards, only one household was selected. Similarly, only pre-determined numbers of households should be selected in all the PSUs/houses, no matter how many households were re-listed afterwards. After re-listing, some a good number of PSUs/houses turned out to contain fewer households than the pre-determined number of households to be selected from. Therefore, the total sample became smaller than planned. Instead of 1,001 households, the sample became 904 households.

Substitution

No substitution of selected PSUs/houses or households was to take place.

Random selection of an individual aged 15 or above within the household

The interviewer is responsible for the selection of the RSI. The RSI selection is from a subset of the household members aged 15 or above and who live at least one day every week with the selected household. The random selection entails two steps. Firstly, the interviewer listed and sorted all eligible household members by sex and age, i.e. listed males first and then females, the older first and then the younger. The second step was the random selection from a pre-sorted list, with the help of a random number table, i.e. a so-called kish table, attached to the questionnaire. The kish table scheme is probably the most common way of selecting individuals at random within households. We used Kish's original set of eight tables, which is reproduced below:

		Number of eligible					
Proportions assigned	Table #	1	2	3	4	5	6+
1/6	1	1	1	1	1	1	1
1/12	2	1	1	1	1	2	2
1/12	3	1	1	1	2	2	2
1/6	4	1	1	2	2	3	3
1/6	5	1	2	2	3	4	4
1/12	6	1	2	3	3	3	5
1/12	7	1	2	3	4	5	5
1/6	8	1	2	3	4	5	6

Source: Kish 1965: 399

To use the table, the interviewer must know which table to use, and how many eligible members there are in the household. Thus, if table 4 is to be used, and there are 6 eligible members in the household, person number 3 is selected. If table 7 is to be used with 6 eligible members person number 5 would be chosen. Etc.

Furthermore, in the case of the original Kish table, the tables are allocated to the interviews in different proportions. Thus, in a sample of 1,200 households, table 1, 4, 5 and 8 should each be allocated to 200 households, while table 2, 3, 6 and 7 should be used for 100 households each. The tables were printed on the questionnaires.

Inclusion probabilities and weights

It follows from the above that the sample in general is a two-stage sample.

Notation

In order to describe the sample precisely and calculate inclusion probabilities we need to introduce some notation. This is done in Table 3. In general the notation uses subscripts to indicate the sample stage, and superscripts to indicate the source of the data used. Thus $N_{h,c}$ means the population in stratum h, cluster c.

Table 3: Notation used

Symbol	Meaning
N	Household count (initial estimate)
N^{l}	Household count as listed
N	Number of households Uppercase: Total numbers in population

Symbol	Meaning
	Lowercase: Sample numbers
$N_{h,d}^{\geq 15}$	Number of eligible household members for selection of RSI, i.e. aged 15 or older and live at least one day per week with the household
m	Sample number of PSUs /houses
p	Inclusion probability
h	Index of stratum
c	Index of PSU
f and i	Index of household (f used to indicate household in the sampling stage, i used to indicate the list of all households from 1 to n in the sample)
d and r	Index of RSI (d used to indicate RSI in the sampling stage, r used to indicate the list of all eligible household members from 1 to N in the household)

Selection of PSUs

The inclusion probability for a PSU *c* in stratum *h* is the following.

Equation 1: Inclusion probability for PSU

$$p_{h,\,c} = \frac{N_{h,c} m_h}{N_h}$$

Equation 2: Inclusion probability for household

$$p_{h,\,c,\,f} = \frac{n_{h,c}}{N_{h,c}^{l}}$$

Note that the listed number of households is used, rather than the initial estimate of households from the census. The $n_{h,c}$ is pre-determined number of households to be selected in each PSU, which is same within each stratum, but different between different strata.

The overall inclusion probability for a household then becomes:

Equation 3: Overall inclusion probability for household

$$p_i = p_{h,\,c} \cdot p_{h,\,c,\,f} = \frac{m_h N_{h,c} n_{h,c}}{N_h N_{h,c}^{l}}$$

Selection of RSIs

The inclusion probability for RSI *d* within the N adults (members 15+) of household *i* is:

Equation 4: Inclusion probability for RSI

$$p_d = \frac{1}{N_{i,d}^{\geq 15}}$$

Since only one RSI is selected, the overall inclusion probability for a random selected individual becomes:

Equation 5: Overall inclusion probability for RSI

$$p_r = p_i \cdot p_d = p_{h,c} \cdot p_{h,c,f} \cdot p_d = \frac{m_h N_{h,c} n_{h,c}}{N_h N_{h,c}^{l} N_{i,d}^{\geq 15}}$$

Sampling weights

There are two types of sampling weights. The expansion weights create estimates equivalent to real numbers in the population, while the relative weights retain the sample size and only adjust the relative contribution of each unit of analysis (household or individual). Only the expansion weights, which are the inverse of the sampling probability, were calculated in this survey.

Thus, the expansion sampling weight for household *i* is:

Equation 6

$$W_i^e = \frac{1}{p_i}$$

The expansion sampling weight for RSI *r* is:

Equation 7

$$W_r^e = \frac{1}{p_r}$$

The sampling weights as such are not used in estimation of survey results, because the sampling weights are adjusted for non-response as will be discussed below.

The sample, and the various size measures that go into it, must be documented accurately. This is necessary in order to be able to calculate the weights properly. It is practical to use a spreadsheet for this task. The sample and the various size measures that go into it is all that is needed for the

calculations of the inclusion probabilities down to the household level. A suggested variable list for the documentation file is given below, together with the source for the information.

Table 4: Sample documentation file structure

Variable name	Meaning	Source of information	Variable name in questionnaire
AC02	Number of eligible RSI	Merged from data	AC02
Governoret	Governorate	Standard coding	
Camp	Camp	Standard coding	
Area	Area	Standard coding	
Road	Road	Standard coding	
Dwilling	Dwelling number in each area	Coding	
Name	Name of Head of Household	Created by re-listing	
RESULT	Result of re-listing households	Created by re-listing	
strata	Unique stratum number	Created by sample	
Tpsu	Total number of PSUs/houses in stratum	Merged from sample frame	
cluster	Total number of households in stratum	Merged from sample frame	
household	Number of households in PSU/house (Old list)	Merged from sample frame	
psu	Number of PSUs/houses to be selected in each stratum	Created by sample	-
n	Number of households to be selected in each PSUs/houses	Created by sample	
p1	Inclusion probability of PSU	Calculated in SPSS	
		Equation 1 household*psu/cluster	
household2	Number of households in each PSU/house as listed (New list)	Field work	
p2	Inclusion probability of household	Calculated in SPSS Equation 2	
		/household2	
p_hh	Overall inclusion probability of household	Calculated in SPSS	
		Equation 3	
		P1*P2	
p3	Inclusion probability of RSI in the household	Calculated in SPSS Equation 4	
		1/AC02	
p_RSI	Overall inclusion probability of RSI	Calculated in SPSS	
		Equation 5	
		P1*P2*P3	
HHw	The expansion sampling weight for household	Calculated in SPSS Equation 6	
		1/P_hh	
RSIw	The expansion sampling weight for RSI	Calculated in SPSS Equation 7	
		1/P_RSI	

Non-response and non-response corrections

The response rate achieved during the fieldwork of a survey is crucial for the quality of the survey results. When response rates are low, one may justifiably suspect biases in the results.

In general one can distinguish between two types of non-response: unit non-response and item non-response. Unit non-response pertains to the non-response of a whole unit, such as a household. In that case almost nothing is known about that household.

Item non-response pertains to the lack of information on a specific item for a unit, for instance that a person does not answer questions about income.

Here we will only consider unit non-response.

Unit non-response: the household

The results of interviews or attempted interviews can be studied using a classification of non-response in the questionnaire, derived from Hidiroglou, Drew and Gray (1993). The response categories in the framework are given in Table 5.

The framework is built around the observation that an interview can be missing for two reasons. First, it may be that the selected household does not belong to the sampling frame. This is the case for instance for diplomats which were not considered eligible. Second, a selected household, which actually exists and is eligible, may refuse, or not be found at home. Also, the classification has to take into account that there will be some situations where the interviewer cannot determine if a household exists or not. In addition, interviewers sometimes encounter the situation where a household is available for interview, but that no useful information can be obtained because the respondent is sick or otherwise incapable of answering.

Table 5: Response categories

Category	Response type
1 Interview completed	Interview is possible (response)
2 Refusal converted by supervisor (The respondent initially refused, but co-operated after a visit by the supervisor)	Interview is possible (response)
3 Partly completed	Interview is possible (non response)
4 Status not determined (The field work team could not find out if a household was living at the address)	Not clear, usually distributed over possible and not possible interview
5 No usable information (for instance because the respondent was sick, mentally ill, not really co-operating)	Interview is possible, non-response
6 Dwelling unit did not exist	No interview possible
7 Dwelling unit was vacant	No interview possible
8 Dwelling unit is under construction	No interview possible
9 Not eligible	No interview possible
10 No contact (the household exists, but could not be found at home)	Interview is possible, non-response
11 Refusal	Interview is possible, non-response

A number of rates can be computed from the above framework as indicated by the table below.

Table 6: Calculation of non-response rates in EHC (based on first interview in each round)

Item	Calculated as	Response
Total- n	All households/dwelling units drawn in the sample	904
Resolved – n	Total minus the units with indeterminate status (code 4)	904
In scope – n	Resolved minus the not existing, not eligible or vacant units	904
Completed interview –n	Interviews with at least first visit	889
Resolved rate	Resolved/Total	100
In scope rate	In scope/Resolved	100
Non-existence rate	Non-existent units/resolved units	0
Temporary out of scope rate	(Vacant + Not eligible) / resolved	0
Response rate	Completed interviews/In scope	98.3
Refusal rate	Refusals/In scope	4.4
Refusal conversion rate	Refusals converted/(Refusals + converted)	71.4
No contact rate	(Not determined + No contact)/ (Not determined + In scope)	0
Non-response rate	(Not determined + Refusals + No contact)/(Not Determined + In scope)	71.4
Residual non-response rate	No usable information / In Scope	0.1

The framework allows for showing both non-response that is due to imperfections in the sample frame and imperfections that is due to problems during interviewing. As one can see, the response rates are very high due to the re-listing before interviewing.

References

Hidiroglou, M., J. Drew and G. Gray 1993. 'The measurement of non-response in surveys', *Survey Methodology*, Vol. 19: 81-94.

Kish, L. 1965. *Survey sampling*, New York: Wiley.

Ugland, Ole Fr. (ed.) 2003. *Difficult Past, Uncertain Future: Living Conditions among Palestinian Refugees in Camps and Gatherings in Lebanon*, Fafo-report 409, Oslo: Fafo.

Syntax used during sampling

```
get file='C:\zhang huafeng\91066 Ein Hilweh\PPS sampling\FRAME
LEBANON.sav'.
sel if camp=21 or camp=24 or camp=26.
*.
aggregate outfile * mode=addvariable
/break governoret camp area dwilling
/household=n(dwilling).
var lab household 'Number of households in the dwelling (old list)'.
sort cases by governoret camp area dwilling.
save outfile='C:\zhang huafeng\91066 Ein Hilweh\PPS sampling\FRAME Ein
Hilweh.sav'.
*.
aggregate outfile *
```

```
/break governoret camp area dwilling
/household=min(household).
*.
if household=1 strata=1.
if household=2 strata=2.
if household=3 strata=3.
if household>3 and household<15 strata=4.
if household=15 strata=5.
if household>16 strata=6.
sort cases by strata.
exe.
*.
if strata=1 psu=294.
if strata=2 psu=119.
if strata=3 psu=96.
if strata=4 psu=129.
if strata=5 psu=1.
if strata=6 psu=1.
exe.
*.

aggregate outfile * mode=addvariable
/break strata
/Tpsu=n(household)
/cluster=sum(household).
*.
var lab strata 'Number of stratum'.
var lab psu 'Number of PSUs selected in each stratum'.
var lab Tpsu 'Total number of PSUs in each stratum'.
var lab cluster 'Total number of households in each stratum'.
*.
drawpps stratv=strata mos=household psups=psu totsams=1001.
save outfile="C:\zhang huafeng\91066 Ein Hilweh\PPS
sampling\selection.sav".
get file="C:\zhang huafeng\91066 Ein Hilweh\PPS sampling\selection.sav"
/keep Governoret Camp Area Dwilling household strata cluster psu selected.
select if selected.
save outfile="C:\zhang huafeng\91066 Ein Hilweh\PPS
sampling\selection2.sav".
*.
sort cases by governoret camp area dwilling.
match files /table=*
/file='C:\zhang huafeng\91066 Ein Hilweh\PPS sampling\FRAME Ein Hilweh.sav'
/by governoret camp area dwilling.
sel if selected=1.
exe.
save outfile="C:\zhang huafeng\91066 Ein Hilweh\PPS
sampling\selection3.sav".
```