Sampling Plan
Iraqis in Jordan

Contents

Sampling Plan

Survey of Iraqis in Jordan

Introduction

This document describes the sample of the Survey of Iraqis in Jordan. The main aim of the document is to document the sampling procedures and the procedures for handling non-response in the survey.

Requirements of the sample

The Survey of Iraqis in Jordan is designed to

1. Estimate the number of Iraqis that at the time of the survey reside in Jordan.
2. Describe the main characteristics of the Iraqis

The questionnaire calls for a respondent that answers for the household. The survey will focus on presenting estimates for the whole of Jordan, rather than for governorates or for other geographic sub-domains. Nevertheless, estimates for Amman and for the rest of the country will be made.

The sampling frame

The sampling frame for the survey is the Jordanian Census of 2004. The enumeration area of the census ("Block") is a geographical area with clear borders that typically contains about 70 households, but may in some cases be considerably larger.

The 2004 Census registered 40,084 Iraqis residing in Jordan. Tabulation of the 2004 census shows that the Iraqis at that time were mainly living in Amman governorate, in fact 78.4 percent of the census blocks with Iraqis were located there, corresponding to 76.4 percent of the Iraqis. Within Amman, as well as also within other governorates the distribution of Iraqis is very uneven.

Given the higher number of Iraqis in Jordan at present than in 2004, it is likely that the distribution of Iraqis across census blocks have changed, but that the basic pattern remains, namely that the majority reside in Amman, and that they tend to cluster. Indeed, this is also suggested by the labor force surveys carried out by the Department of Statistics.

There are generally four types of places where people may live:

1. Ordinary housing (flats, independent houses etc)

2. Hotels
3. Other collective housing (QIZ – workers barracks in qualified economic zones, military barracks)
4. Non permanent residence

The sample frame includes in principle only the first category, i.e. ordinary dwellings. However, Non-permanent residence may also be included in the listing, since the enumeration area will be considered as an area frame, i.e. all households, regardless of whether they are found in permanent or non-permanent housing will be included in the listing.

Hotels will be covered separately, by asking them to provide the number of Iraqis that have stayed during a 14 day period. Note that staff living in the hotels will be included as persons any other dwelling, i.e. as a general rule employees living at their work place will

Workers staying in barracks in special economic zones are often non-Jordanian, but generally not Iraqi. However, information will be sought in order to determine if a separate sample is needed here. However, the total population of QIZs is about 35,000 CHECK people, so it is unlikely that inclusion or exclusion will make any significant difference for the estimate of total numbers of Iraqis. Military barracks will not be covered (and should also be totally unnecessary given the purpose of the survey).

Sample design

The sample design takes as its point of departure three observations. First, as suggested by the 2004 census it is likely that the population of Iraqis is clustered. Second, since the population is known to have grown substantially since 2004, it is likely that the clustering is still an important feature, but that areas where Iraqis were not present in 2004 may have Iraqis now. Third, it is likely that the pattern observed in 2004 with the majority of Iraqis residing in Amman is still the case.

As can be seen from the Table 1 below the number of Iraqis per block varies widely.

Table 1: Distribution of blocks in the 2004 census by number of persons of Iraqi origin

	Number of Iraqis per block													
Governorate	0	1	2	3	4	5-9	10-19	20-29	30-39	40-49	50-99	100-499	1000+	Total
Amman	2571	415	207	219	238	836	595	228	92	40	53	7		5501
Balqa	649	90	24	19	17	43	13	2	4					861
Zarqa	1662	154	71	57	46	120	41	5	1		1			2158
Madaba	318	33	5	12	9	18	4							399
Irbid	2221	128	42	30	29	74	26		1					2551
Mafraq	580	51	21	13	4	16	7					2	1	695
Jarash	352	16	3	3	3	9	1							387
Ajloun	308	22	6	1		2	1							340
Karak	537	22	7	7	9	24	6	4						616
Tafileh	232	11	3	2	2	3								253
Ma'an	254	25	6	8	6	18	5	1						323
Aqaba	258	17	13	8	6	24	5	1			1	1		334
Grand Total	9942	984	408	379	369	1187	704	241	98	40	55	10	1	14418

From the above observations follow that the sample frame can be divided into two main parts: the blocks (enumeration areas) that are nearly certain to contain Iraqi households, and blocks that may or may not have Iraqi households. Moreover, larger areas, such as districts, may be divided into districts with large variation with respect to numbers of Iraqis, and areas with little variation.

Three different designs have been considered for the survey:

1. A standard two stage household survey, with a large sample, so as to be able to pick up sufficient number of Iraqi households
2. An adaptive cluster sample. In this type of sample first an ordinary cluster sample is drawn, and then neighbors to a selected cluster will be included in the sample if the selected cluster have more than a given number of Iraqis present.
3. A standard cluster sample, but with adaptive allocation of strata. Adaptive allocation entails first selecting a sample across several strata, and then conduct a new phase of selection with optimal allocation of clusters based on the observed variation in the strata.

The first option was dropped, because of the large number of interviews that would be required. The two other options were chosen depending on location in Jordan, depending on what was known about the distribution of Iraqis in the various locations.

Allocation

In a stratified sample, the most efficient way to allocate the sampling units to strata is the so called Neyman allocation, where the number of sampling units selected in each stratum is allocated proportionally to the product of the standard deviation of the variable of interest and the number of sampling units in the stratum.

Table 2: Optimal allocation of PSUs based on number of Iraqis as observed in the 2004 census

Governorate	N (PSUs)	Optimal Allocation	Allocation used
Amman	4,949	723	200+200+200*
Balqa	636	31	30
Zarqa	1,746	76	70
Madadba	288	8	20
Irbid	2,153	56	40
Jerash	398	57	40
Mafraq	217	5	20
Ajloun	239	3	20
Tafileh	541	19	20
Karak	172	2	20
Ma'an	188	6	20
Aqaba	256	14	21
Total	**11,783**	**1000**	**921****

* Adaptive allocation in three phases ** The figures do not add to 1000 because adaptive addition of 79 clusters were planned for.

If the distribution of the standard deviation of the number of Iraqis remains the same in 2007, the above allocation would be the most efficient. However, despite that it appears clear that the majority of Iraqis still live in the approximate same areas as in 2004, such an allocation also exposes the survey to the risk that distributional changes since 2004 cannot be detected, since the sample in some governorates (e.g. Ajloun, Karak) would be very small. Therefore, the allocation was increased in governorates with very few clusters so that it would be possible to detect changes. Without changes in the pattern of variance of the distribution of Iraqis since 2004 this would lead to a slightly less efficient sample than the optimal, but, as noted, this design protects to some extent against changes.

Sample selection procedures

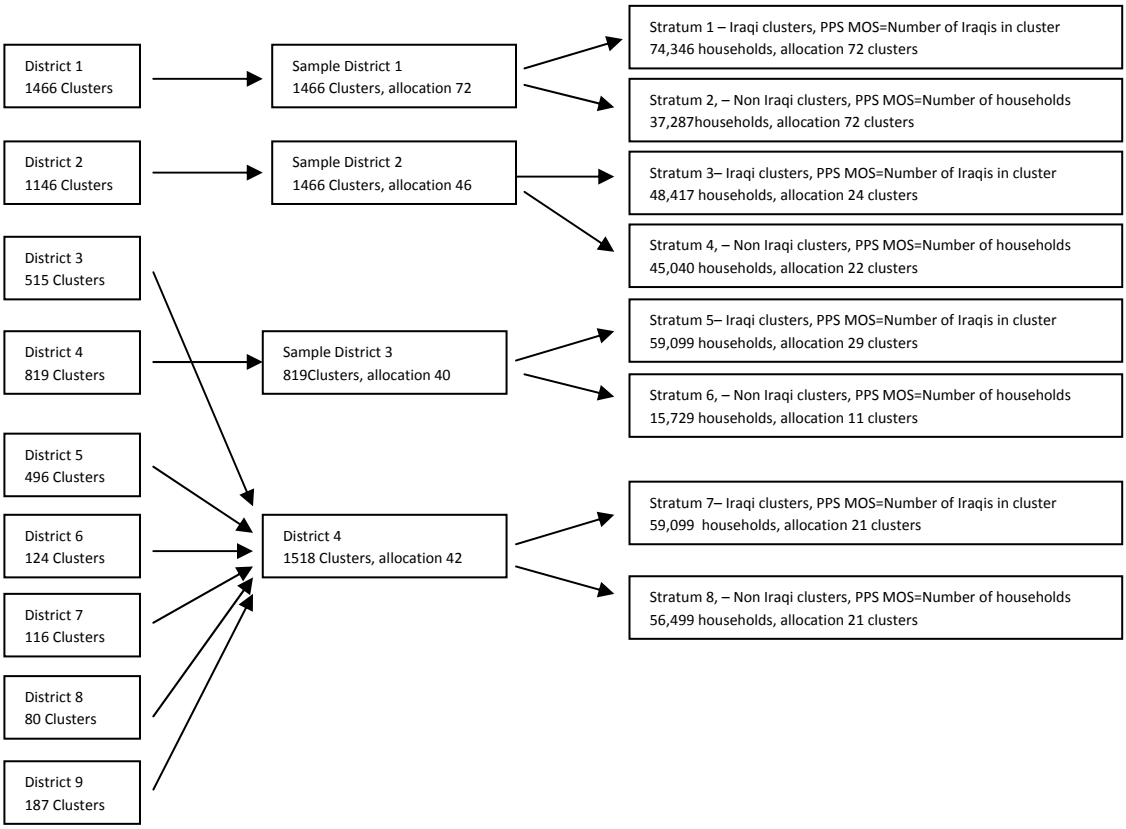
Selection of PSUs

As noted the sample consists of two main designs, an adaptively allocated stratified cluster sample in the case of Amman, and a stratified adaptive cluster sample in the rest of the country. Two different selection processes are therefore used.

Adaptively allocated stratified cluster design

Examination of the 2004 spatial distribution of Iraqis in census enumeration areas in Amman suggests that the Iraqis are unevenly distributed across the city, but that a large number of clusters have Iraqis present. To account for the distribution, 4 groups were first constructed, corresponding to administrative district 1,2,4 as group 1,2 and 3, and the remaining districts (3,5,6,7,8,9) as group 4. The groups were then each split into two parts: one containing clusters with Iraqis and the other part not and allocated proportionally to the number of households in each. Altogether the procedure means that one arrive at 8 strata (Figure 1).

Figure 1: Structure of the sample in Amman



In the case of the stratified sample (i.e. principally Amman) the selection of PSUs will be conducted with linear systematic PPS sampling within each stratum. The measure of size (MOS) will be the number of Iraqis in each cluster for the Iraqi strata, and number of households recorded in the 2004 census for the non-Iraqi strata. Several alternative ways of selecting PSUs by PPS exist, but the linear systematic one was chosen because it tends to spread out the sample geographically and because it creates implicit stratification.

The selection initial selection of the sample will be carried out with an allocation of 200 households in total. Then, the second phase will adjust the allocation based on the results from the listing of the first 200, and the third phase then will adjust the allocation again, but still use 200 clusters. Thus, a requirement for the adaptive allocation, namely that each phase have a fixed weight, is fulfilled (Thompson and Seber 1996:189).

In the second phase and third phase the 8 strata in Amman was merged into four, which correspond to the sample district 1 to 4 in the diagram above.

Adaptive cluster design

In the case of the adaptive cluster sample, the process is as follows within each stratum:

1. Draw a sample of PSUs (blocks) using linear systematic PPS.
2. For each cluster, determine the number of households of Iraqi origin. This must be done by a very thorough listing of the PSU.
3. Decide on a threshold level of households with Iraqis in a PSU. If the number is higher, select the neighboring clusters, and determine the number of Iraqi households in these. For each cluster, if the number is higher than the threshold, select neighbors. (How far the process continues from a seed cluster depends on the threshold value. If the threshold value is very high, the sample will degenerate to an ordinary cluster sample because no neighbors will be selected. If the threshold is very low, one ends up with a census.
   Arbitrarily assuming 500,000 Iraqis in Jordan, and the same proportion of these residing in Amman as was found in the census, there would be around 118,000 Iraqis living outside of Amman. If these were distributed equally across the country (excluding Amman), about 3.5 percent of the population in each block would be Iraqis, or on the average (assuming a block size of 70 households) roughly 2-3 households per block. Since the distribution of households per block is heavily right-skewed, the median number is likely less than the mean, and threshold value of around twice the mean, i.e. 6 Iraqi households may be a good initial choice. This should also be able to accommodate the situation of many more Iraqis than 500,000 altogether, and in the case of fewer Iraqis than 500,000, there would be relatively few outside Amman, and the precision of the estimate outside Amman would not be that important for the total.

Note that a *neighboring cluster* is any cluster that shares a boundary with a selected cluster. If two clusters share a corner, but have no common boundary, they are not considered neighbors.

The stratification poses potential difficulties for the field workers in conjunction with the adaptive cluster sampling in two circumstances:

1. A group of clusters at the border of a district stratum divide grows from one stratum into the other. In such cases the growth of the network of clusters will be stopped at the district boundaries.
2. A group of clusters at the border of a governorate stratum divide grows from one governorate into the other. In such cases the group should be truncated at the governorate boundary, in order to facilitate fieldwork logistics.

The adaptive cluster sample was stratified according to district and optimally allocated with respect to the distribution of Iraqis in the census of 2004, but adjusted so that no stratum received less than 2 PSUs.

Table 3: Allocation outside Amman

Governorate	District	Number of PSUs	Optimal allocation	Adjusted allocation
Balqa	1	184	6	6
	2	71	5	5
	3	67	1	2
	4	245	9	9
	5	69	9	8
	Total	636	30	30
Zarqa	1	1042	51	51
	2	609	15	15
	3	95	3	4

Governorate	District	Number of PSUs	Optimal allocation	Adjusted allocation
	Total	1746	70	70
Madadba	1	202	17	17
	2	86	3	3
	Total	288	20	20
Irbid	1	853	26	20
	2	225	5	5
	3	247	2	2
	4	140	1	2
	5	186	1	2
	6	239	3	3
	7	139	1	2
	8	70	1	2
	9	54	1	2
	Total	2153	40	40
Jerash	1	155	5	4
	2	129	4	4
	3	100	31	30
	4	14	1	2
	Total	398	40	40
Mafraq	1	217		20
	Total	217		20
Ajloun	1	175	18	18
	2	64	2	2
	Total	239	20	20
Tafileh	1	174	9	6
	2	171	7	4
	3	50	2	2
	4	65	1	2
	5	31	0	2
	6	30	0	2
	7	20	1	2
	Total	541	20	20
Karak	1	88	16	15
	2	59	3	3
	3	25	2	2
	Total	172	20	20
Ma'an	1	83	10	10
	2	59	5	5
	3	35	3	3
	4	11	1	2
	Total	188	20	20
Aqaba	1	230	16	16
	2	26	5	5
	Total	256	21	21

Mapping and listing of PSUs

Maps for selected PSUs should be updated and completely re-listed. The purposes of the mapping and re-listing are three:

1. To find the number of households with Iraqis in the PSU
2. To enable selection of households
3. To enable interviewers to locate the selected households.

The mapping and listing is described in detailed in the mapping and listing manual. It is extremely important that the mapping and listing is very thoroughly carried out, and that all locations where people can actually live are included.

Selection of households

The household listing should result in a list that distinguishes households where there is one or more Iraqi present, and households where this is not the case. The sample take would then be as shown in

Household type	Definition	Sample take
Iraqi	At least one household member is Iraqi citizen. At the time of listing, no consideration will be made of when the person arrived in Jordan.	10, or as many Iraqi households as there are if less than 10.
Non-Iraqi	No household member is Iraqi	2

There are two reasons for sampling non-Iraqi households, first, and most importantly, to verify the overall quality of listing (i.e. if households that are actually Iraqi were listed as non-Iraqi) and second, to enable the comparison between Iraqi and non-Iraqi households.

Linear systematic sampling should be used to select households from each list of households. The sample is considered as a sample of households, not dwellings.

Substitution

No substitution of selected PSUs or households is to take place. Substitution will inevitably result in bias in the estimates of the population of Iraqis, as well as of their characteristics.

Additional households in dwelling units

The list of households is intended to be a complete list of households in a PSU. Therefore, if two households are found in one dwelling unit, only the one the interviewer has been instructed to interview should be interviewed. The other household should in principle be listed separately on the household list, and therefore has an independent chance of being included in the sample.

Inclusion probabilities and weights

The sampling designs described here are both unequal probability ones, where each cluster in principle has its unique inclusion probability. Therefore, we need an estimator that takes account of the inclusion probabilities of each unit. Such an estimator is the Horwitz-Thompson estimator, which is applicable to both designs. Let us designate the variable of interest as *y*. Then an unbiased estimator of *Y* is:

$$Y = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$$

where the total Y is the sum of each y divided by its associated inclusion probability π. The problem is now finding the inclusion probability π. As we will see, however, some adaptations will have to be made in conjunction with the adaptive part of the sample.

Notation

In order to describe the sample precisely and calculate inclusion probabilities we need to introduce some notation. This is done in Table 4. 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 4: Notation used

Symbol	Meaning
	Household count (initial estimate)
N^{l}	Household count as listed
N	Number of households
	Uppercase: Total numbers in population
	Lowercase: Sample numbers
M	Number of PSUs
	Upper case: Total number
	Lower case: Sample numbers
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)

Selection of PSUs

PSU inclusion probability for PPS selection

The inclusion probabilities for a PSU *c* in stratum *h* is the following when the PSU has been selected PPS. Note that the measure of size (the Ns in the equation) can be either households (in non-Iraqi) strata, or number of Iraqi (in the Iraqi strata).

Equation 1: Inclusion probability for PSU (PPS selection)

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

In Amman, where a phased optimal allocation sample is used, an estimator for the total number of Iraqis now is:

Equation 2: Estimator for total number, Adaptive allocation

$$\hat{Y} = \sum_{k=1}^{K} w_k \hat{Y}_k$$

Where w_k represents the fixed weight associated with each phase (1/3 in this case), and $\hat{Y}_k$ the population total calculated from each phase (using the Horwitz-Thompson estimator as outlined above).

PSU Inclusion probabilities for adaptive cluster sampling

When adaptive cluster sampling has been carried out, the inclusion probabilities are different. The number of Iraqis represented in the sample obviously is not a number that can be directly expanded into the population of Iraqis, as PSUs with many Iraqis have a higher probability of being included than those with few or none.

Let us first define what a network is, namely an initial PSU and the PSUs that have been selected together with it as a result of the application of the sampling process. This is illustrated in Figure 2 which shows a grid of primary sampling units where some have at least one of the target population (i.e. an Iraqi household). If one of the PSUs with an Iraqi households were sampled initially (e.g. the one with three dots, corresponding to three households), the evenly shaded units correspond to its neighbors, while the cross-hatched units correspond to the neighbors with no Iraqi households present. The neighbors with no members of the target population present are called *edge units* (Thompson and Seber 1996:94). The network we are interested in for estimation is the network resulting from a selection of a PSU, but with edge units removed. Note that from a fieldwork point of view the edge units pose considerable mapping and listing work, but do not contribute to the estimate.

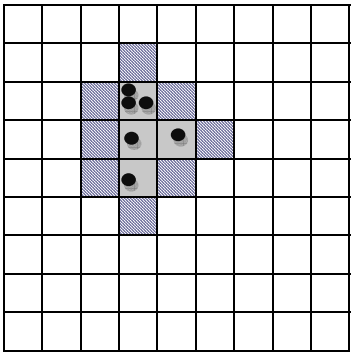


Figure 2: Network

Following Thompson and Seber (1996:94) an estimator of the total number for Iraqis that is derived from the Horvitz-Thompson estimator is:

Equation 3: Estimator for total number, Adaptive cluster sample

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i I'_i}{\pi'_i}$$

Where I' is an indicator variable that is 1 if the initial sample intersects the network i, and 0 if not. The probability π is not strictly an inclusion probability, but an intersection probability or the probability that a given cluster is used in the estimator. It can also be expressed in terms of each network, rather than in relation to individual clusters. The y's now is the total number of Iraqis in each distinct network, and k counts networks rather than clusters.

Equation 4: Estimator of total expressed with respect to networks, Adaptive cluster sample

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{K}\frac{y^*_k J'_k}{\pi'_k}$$

The probability, which is equal for networks and units, can be expressed as follows:

Equation 5: Intersection probability for networks for adaptive cluster sample

$$\alpha_k = 1 - \left[1 - \frac{N_k}{N}\right]^{n_1}$$

N_k and N now refers to the number of households in the network and in the stratum, respectively, and n_1 the number of clusters, each one corresponding to a cluster selected in the initial sample.

Note that the above formula is strictly speaking only valid when cluster inclusion is stopped at stratum boundaries (or where there is no stratification). Selection for each stratum is then independent, and strata can be treated in the usual way. (Thompson and Seber 1996: 134).

Selection of households

Within each PSU a fixed number of households is to be selected. The inclusion probability for a household *h* within a PSU *c* in stratum *s* is then the following:

Equation 6: Inclusion probability for household

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

Note the use of the superscript l, which indicates that it is the listed number of households that should be used rather than the initial estimate of households from the census. Note also the use of the subscript l, which indicates which list is used, i.e. the list of Iraqi vs the list of non-Iraqi households. The $n_{h,c}$ is constant for non-Iraqi households (2) and variable for Iraqi households. The overall inclusion probability for a household then becomes:

Equation 7: Overall inclusion probability for household

$$p_i = p_{h,\,c} \cdot p_{c,\,f}$$

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). The expansion weights are calculated as the inverse of the sampling probability, while the relative weights are calculated as the expansion weight divided by the mean of all the expansion weights. As we have noted above, sampling weights can be calculated on the basis of the inclusion probabilities for each household, regardless of the way in which the household has been sampled.

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

Equation 8

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

The relative sampling weight is then:

Equation 9

$$W_i^r = \frac{W_i}{\frac{\sum W_i^e}{n}}$$

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 5: Sample documentation file structure

Variable name	Meaning	Source of information	Variable name in questionnaire
100	First questionnaire number in PSU	Generated from sample	100
100	Final questionnaire number in PSU		100
101	Governorate	Sample frame	101
GOVNAME	Governorate name	Administrative name	-
102	District	Sample frame	102
DISNAME	Name of District	Administrative name	-
103	Sub District	Sample frame	103
104	Locality	Sample frame	104
105	Area	Sample frame	105
106	Sub area	Sample frame	106
107	Stratum number	Defined in sample	107
108	Cluster number	Sample frame	108
109	Parent cluster Serves as Network ID	Sampling	109
STYPE	Stratum type: Adaptive allocation=1 Adaptive cluster sample=2	Sample	-
PHASE	Phase in adaptive allocation (empty for adaptive cluster sample PSUs) .	Sample	-
THRESHOLD	Threshold used for selection neighbor (ACS)	Sample	-
LARGEHH_H	Number of households in stratum	Census 2004	-
LARGEI_H	Number of Iraqis in stratum	Census 2004	
LARGE_M	Number of PSUs in stratum	Census 2004	-
SMALL_M	Number of PSUs to select in stratum (initial selections for ACS)	Sample	-
LARGNPSU	Number of households in PSU	Census 2004	-
LARGIRAPSU	Number of Iraqis in PSU	Census 2004	-
MEM	Number of people in PSU 2004	Census 2004	-
SIGMAMOS	Sum of MOS in stratum	Census 2004, sample	
MOS	Measure of size used for PSU selection	Census 2004, sample	
NETWORK	Network size. Number of clusters that is part of the network starting at the parent cluster. Clusters that do not lead to the selection of neighbors have network size 1.	Listing fieldwork	
I_LISTED	Listed number of Iraqi households	Listing fieldwork	
IP_LISTED	Listed number of Iraqi persons	Listing fieldwork	
NON_I_LISTED	Listed number of non-Iraqi households	Listing fieldwork	
NON_IP_LISTED	Listed number of non-Iraqi persons	Listing fieldwork	
SMALL_N_I	Selected number of Iraqi households	Consequence of listing fieldwork: 0..10	
SMALL_N_NI	Selected number of non-Iraqi households	2	

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. Since the survey of Iraqis in Jordan aims to estimate the number of Iraqis, i.e. a total, non-response is likely to bias the estimate downwards.

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 6: Response categories

Category	Response type
1 Interview completed	Interview is possible (response)
2 Refusal converted	Interview 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, distributed equally 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 Not eligible	No interview possible
9 No contact (the household exists, but could not be found at home)	Interview is possible, non-response
10 Refusal	Interview is possible, non-response

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

Table 7: Calculation of Non-response rates in Survey of Iraqis in Jordan (based on first interview in each round)

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

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.

Non-response correction

Non-response always occurs. Nevertheless, since the extent and seriousness of non-response vary, the plan of non-response correction must be reconsidered after the fieldwork.

Weight adjustment and estimation weights

When there is unit non-response, direct use of the sampling weights will result in biased estimation. The biases generally take two forms. One is when totals are estimated with sampling expansion weights; the total will be

too small because non-response implies that units that should be added into the total are missing. The other is when estimation may be biased due to particular characteristics of the non-responding units.

One way to reduce the biases produced by unit non-response is to adjust the sampling weights. The method of correction of the weights for non-response that is used here, is the so-called “adjustment cell method”(see for instance Lehtonen and Pahkinen 1995; Little and Rubin 1987). In this approach, households considered to be fairly similar are identified and the non-response rate is calculated for each group of households, called adjustment cells. In line with the description above, when non-response rates are calculated, only the non-response of those that could have responded, but for some reason did not do so, is considered.

The inverse of the non-response rate in each adjustment cell will then be used to adjust the sampling weights (both expansion and relative) for each household. The result is the so-called estimation weights, both expansion and relative. The weighted sample size will correspond to the sample size if all households had responded. The effect includes increasing the relative contribution to the estimates of units that are similar to those missing.

In the present study the adjustment cells used are likely to consist of a number of geographically adjacent PSUs.

Using the notation in Table 8, the correction factor to the weights for non-response is given in Equation 10.

Table 8: Notation for non-response adjustment

Symbol	Explanation
C	Adjustment (Correction) factor
A	Index of adjustment cell
h^r	Responding households
h^f	Non responding households

The number of possible interviews (i.e. the denominator in the non-response rate) is the sum of categories 1,2,3,5,10 and 11 in Table 2. The number of non-respondent units is the sum of the categories 5, 10 and 11. The Status Not Determined category may be distributed over the other categories.

Equation 10

$$C_a = \frac{1}{\frac{h_a^r}{h_a^r + h_a^f}}$$

The weights are then adjusted according to the following equations:

Equation 11

$$W_i^{estimation} = C_i W_i^e$$

Equation 12

$$W_i^{r,estimation} = \frac{W_i^{e,estimation}}{\frac{\sum W_i^{e,estimation}}{n}}$$

The effect of the corrections is that the expansion weights are increased so that the sum of the estimation expansion weights corresponds to the sum of units in the sample frame (less non-existent or non-eligible units). The relative estimation weights are normalised. This means that the sum of the weights is the sum of the household records in the data file.

Use of estimation weights

Since the survey is not self-weighting, estimation weights must be used for all survey tabulations and estimation. In general, the rules for using the weights are as follows:

a) The expansion weights should be used when it is desired to obtain total population figures for subgroups, for example number of unemployed, number of children below 5 or similar.

b) The expansion weights should be used for general tabulation

Estimators for total population size

Estimators for the total population of Iraqis can be derived in several ways. The most simple estimate is simply the sum of the estimates for Amman (from the adaptive allocation, as shown in Equation 2) and for the rest of the country (from the adaptive cluster sample, as shown in Equation 4).

Corrections for listing errors

Three types of classification errors during listing may occur that directly impinges upon the estimate of the total number of Iraqis:

1. The overall number of households recorded during listing may be wrong.
2. Households classified as non-Iraqi may in fact be Iraqi.
3. Households classified as Iraqi may in fact be non-Iraqi.
4. The number of household members recorded during listing may be wrong.

The first problem will not be detected except during quality control procedures during mapping and listing (see the mapping and listing manual)

Classification of Iraqi households as non-Iraqi will to some extent be detected during interviewing because of the sample of non-Iraqi households. This sample may be expected to contain a certain number of households that are actually Iraqi.

Similarly misattribution of Iraqi status to households will also be detected during interviewing, as the Iraqi sample may in fact contain households that really are non-Iraqi.

Procedures for correcting for classification errors will be determined based on the extent of mis-classifcation detected in each direction.

Sampling errors and confidence intervals

Sampling errors for a survey with a complex design cannot be calculated using ordinary statistical procedures, because such software assumes simple random sampling.

For the total population of Iraqis, several approaches are possible. One approach consists of calculating the variances for each stratum, depending on the sampling method used.

For estimates of characteristics of the population, cross-tabulations etc explicit variance estimators are very difficult to construct. Another approach is to create a jackknife estimate of the variance.

For characteristics of the population, it is simplest to use a jackknife estimate. A jackknife estimate may be calculated by first calculating k estimators from the sample, each characterized by having one unique PSU of the k available PSUs in the sample removed. Then, the variance can be computed as the variance between these estimates Wolter (1985: 184) and Levy and Lemeshow (1999: 378-79). One should note that in the case of the adaptively added networks, a whole network counts as one PSU.

References

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

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

Lehtonen, R. and E. J. Pahkinen. 1995, *Practical methods for design and analysis of complex surveys*. Chicester: Wiley.

Little, R. and D. Rubin. 1987, *Statistical analysis with missing data*. New York: Wiley.

Wolter, K.1985, *Introduction to Variance Estimation,* New York: Springer Verlag.

Thompson, S.K. and Seber G.A.F, 1996, *Adaptive Sampling*, New York: Wiley