Iraq Household Socio-Economic Survey 2012 (IHSES-II)
Sampling design and fieldwork organization

Beirut, July 2011
(updated, November 2011)

Introduction

This technical appendix describes the IHSES-II sampling strategy and the fieldwork procedures that could be put in place for survey data collection and quality assurance.

Sample size and strata

The IHSES-II intends to provide estimators of comparable quality for each of Iraq's 118 gadahs (districts). This implies that the sample should be explicitly stratified by gadah, with a similar sample size allocated to each gadah, regardless of its size. A sample size of 216 households per gadah is proposed, equivalent to a total sample of 25,488 households for the country.

Sampling stages

Within each gadah, the sample will be selected in two stages, as follows:

- First, using Census Enumeration Areas (EAs) as Primary Sampling Units (PSUs), select 24 EAs with Probability Proportional to Size (PPS), using the number of households as a Measure of Size (MoS), and with implicit stratification by urban/rural and the subsequent geographical codes (nahya, mahala, village, mukataa and census block).
- Second, using households as secondary Sampling Units (SSUs), select a cluster of 9 households by systematic, equal probability sampling (SEPS) in each of the selected EAs.

The sample frames for both stages can be developed from the 2010 Census enumeration, with no updating of the household lists.

In some of the smallest gadahs, the standard PPS procedure may result in the selection of fewer than 24 EAs, with some of the larger EAs selected more than once. In those cases, two or more clusters will be taken in the EA, as needed.

2,832 EAs were selected in total. 33 of them had less than the 9 households nominally required in the second stage and were merged ex-post with neighboring EAs.

Selection probabilities and sampling weights

With the proposed sampling strategy, the probability p_{hij} of selecting household *hij* in PSU *hi* of Gadah *h* is given by

$$p_{hij} = \frac{24n_{hi}}{N_h} \times \frac{9}{n_{hi}}$$

where n_{hi} is the number of households in the PSU and N_h is the number of households in the gadah. The two factors on the right-hand side respectively represent the probability of selecting the PSU, and the conditional probability of selecting the household within the PSU.

This formula simplifies to p_{hji} = 216 / N_h, which is a constant in each gadah. In other words, the proposed strategy provides an equal-probability sample of 216 households in each gadah. This is advantageous from the standpoint of precision, and it also simplifies analyses because the sample will be self-weighed for gadah-level analyses –although national and governorate-level analyses will require gadah-wise expansion factors.

Fieldwork organization

The IHSES-II fieldworkers will be organized into teams of three interviewers, headed by a supervisor. Each team is responsible for two gadahs (48 clusters) throughout the full 12-month period of data collection.

The team's work plan requires visiting four clusters per month –two from each gadah, as show in Figure 1 below. The month is divided into two *waves*. In Wave 1 (days 1 to 14), the team visits two clusters from one of the gadahs, and in Wave 2 (days 15 to 29)[1], the two clusters from the other gadah.

In each wave, the team will move between clusters (but not between gadahs) on a daily basis, visiting one of the clusters on odd-numbered days, and the other cluster on even-numbered days.

Each interviewer will be responsible for three households, and will visit each of them every other day five times, with the following task schedule:

- In the first visit, the interviewer will complete sections 1 to 3 and 24, deliver the food consumption diaries and explain their use.
- In the second visit, s/he will transfer the data from the first day of diary-keeping to Section 12, and complete sections 4 to 8.
- In the third visit, s/he will transfer the data from the second and third day of diary-keeping to Section 12, and complete sections 9 to 11.
- On the fourth visit, s/he will transfer the data from the fourth and fifth day of diary-keeping to Section 12, and complete sections 13 to 16.
- On the fifth visit, s/he will complete sections 17 to 23.

---

[1] If 2012 were not a leap year, the second wave of February would have had to be brought forward one day.

After the last scheduled visit, the interviewer will conduct as many additional check-up visits as needed, to correct any doubts or inconsistencies in the data that might have been detected by the IHSES data entry program in any of the previous visits.

Each interviewer will use a dedicated laptop computer to enter the data from his three households on a daily basis, meaning that the correction of doubts and inconsistencies won't need to be postponed till the final days of the wave in many cases. In other words, error correction will also be a complement of the data-collection tasks scheduled for the second to fifth visits.

Figure 1

Day	Monthly activities of a IHSES-II Field Team	Day
1	**Gadah A - Cluster 1** / Sections 1 to 3 + 24 — **Gadah A - Cluster 2**	1
2	Diary - Day 1 — Sections 1 to 3 + 24	2
3	Diary - Day 2 Sections 4 to 8 + 12 — Diary - Day 1	3
4	Diary - Day 3 Sections 4 to 8 + 12 Diary - Day 2	4
5	Diary - Day 4 Sections 9 to 11 + 12 — Diary - Day 3	5
6	Diary - Day 5 Sections 9 to 11 + 12 Diary - Day 4	6
7	Diary - Day 6 Sections 13 to 16 + 12 — Diary - Day 5	7
8	Diary - Day 7 Sections 13 to 16 + 12 Diary - Day 6	8
9	Sections 17 to 23 + 12 Diary - Day 7	9
10	Sections 17 to 23 + 12	10
11	Check-up	11
12	Check-up Check-up	12
13	Check-up	13
14		14
15		15
16	**Gadah B - Cluster 1** / Sections 1 to 3 + 24 — **Gadah B - Cluster 2**	16
17	Diary - Day 1 — Sections 1 to 3 + 24	17
18	Diary - Day 2 Sections 4 to 8 + 12 — Diary - Day 1	18
19	Diary - Day 3 Sections 4 to 8 + 12 Diary - Day 2	19
20	Diary - Day 4 Sections 9 to 11 + 12 — Diary - Day 3	20
21	Diary - Day 5 Sections 9 to 11 + 12 Diary - Day 4	21
22	Diary - Day 6 Sections 13 to 16 + 12 — Diary - Day 5	22
23	Diary - Day 7 Sections 13 to 16 + 12 Diary - Day 6	23
24	Sections 17 to 23 + 12 Diary - Day 7	24
25	Sections 17 to 23 + 12	25
26	Check-up	26
27	Check-up Check-up	27
28	Check-up	28
29		29
30		30

(Rows 1–14 = Wave 1; rows 15/16–30 = Wave 2, on both Day columns.)

Quality assurance

The IHSES-II will put in operation the same quality assurance methods successfully implemented by the IHSES-I five years ago: visual scrutiny of all questionnaires, visual observation of some interviews, random check-up visits of the supervisor, and computer-based quality controls. However, having each interviewer enter his/own data is expected to enhance the impact of computer-based quality controls, and to also have positive externalities on the check-up visits of the supervisor, by way of supporting the effective random selection of the households, as well as the effective random selection of the questions to be asked in these visits (in the IHSES-I —as in most surveys worldwide— the check-up supervision form

had a fixed format with which interviewers were very familiar). These enhancements will require that the supervisor has also access to a computer and a printer.

Human and material resources

Since each team will be responsible for two gadhas, The survey could in principle be fielded with only 59 teams (118 / 2). However, this would require having some of the teams work in different governorates, which would be administratively difficult. To avoid this, more teams than strictly needed will have to be deployed in the governorates with an odd number of gadahs. Around 65 teams will thus be needed in total, which translates in a workforce of 18 regional coordinators, 65 supervisors, and 195 interviewers with laptops.[2]

Subsamples

Most questionnaire modules will be administered to all households. However, in order to reduce some of the fieldwork effort, three of the modules will be administered in subsamples, each composed of three of the nine households visited in each Gadha. The three modules are:

- Anthropometrics (Section 7),
- Time use (Section 21), and
- Food consumption by recall (Section 24).

In order to evenly balance the effort of the three interviewers in each team, the nine households in each Gadha will be allocated into interviewers and subsamples as shown in Figure 2 below:

Figure 2
Allocation of the nine households in each Gadha into interviewers and subsamples

	Anthropometrics (Section 7)	Time use (Section 21)	Food consumption by recall (Section 24)
Interviewer 1	Household 1	Household 2	Household 3
Interviewer 2	Household 4	Household 5	Household 6
Interviewer 3	Household 7	Household 8	Household 9

Regarding anthropometrics, it should be pointed out that the analytic value of recording the weight and height of children younger than 60 months would be substantial (because these measures are essential to assess the incidence of malnutrition), and that the savings resulting from restricting these measures to a subsample would be very marginal. This suggests that it may be better to administer Section 7 to children from all households in the cluster. On the other hand, measuring and weighing older children and adults entails substantial difficulties (because this requires different tools, and because adults are less available than young children), and the analytic value of adult anthropometrics is much more

---

[2] The pilot survey was fielded in five governorates during one wave, shortly after Ramadan, 12 teams were deployed: 4 in Baghdad, 4 in Erbil, 4 in Basra, 4 in Najaf and 4 in Ninewa. The operation mobilized 5 regional coordinators, 12 supervisors, and 36 interviewers.

dubious, suggesting that older children and adults should indeed be measured in a subsample of households, or possibly not measured at all.

An important consideration regarding the recording of food consumption by recall is that, despite its physical location at the end of the questionnaire, Section 24 should be administered <u>in the first visit to the household</u>, before the recording of food consumption by diaries. Asking these questions afterwards (when both the respondent and the interviewer will know the diary records) would defeat the purpose of this module, which is to compare the results obtained from the two instruments, to assess the possibility of applying in future surveys the recall method instead of diaries.