

# **Basic Information Document**

## **Mongolia Socio-Economic Survey 2007-2008**

**Mongolia National Statistical Office  
The World Bank  
September 2007**

## Acronyms

HIES	Household Income and Expenditure Survey
LSMS	Living Standards Measurement Survey
NSO	National Statistical Office
SES	Socio-Economic Survey

## Contents

Introduction.....	1
Survey Instruments .....	1
Sample Design .....	4
Training.....	5
Field Work .....	5
Data Entry .....	5
Appendix A.....	6
Appendix B.....	8

## Introduction

The Mongolian NSO has implemented a series of household surveys designed to measure the socio-economic situation in the country, estimate the main poverty indicators, and measure progress towards the MDGs. These surveys were fielded in 1995, 1998, and 2002-03.

The most recent survey was fielded in 2002-03. This survey integrated the Household Income and Expenditure Survey (HIES) with Living Standards Measurement Study (LSMS) features to provide more thorough and complete analyses. The LSMS portion of the survey was administered to a sub-sample of all of the households that participated in the HIES. The total sample of the HIES was 11,232 households equally distributed over four quarters of one year (February 2002 through January 2003). The LSMS portion was administered to one-third of the households from the HIES. Households interviewed for the HIES in the first and second quarters were re-interviewed for the LSMS in March and April 2003, and households interviewed for the HIES in the third and fourth quarters were re-interviewed for the LSMS in May through July 2003. In total, 3,308 households were included in the LSMS subsample.

The 2007 Socio-Economic Survey (SES07) was designed to be comparable to the HIES-LSMS 2002-03 and to be a template for future efforts. In the future, the annual HIES will no longer be implemented, but rather will be replaced with the periodic HIES-LSMS survey.

## Survey Instruments

Four questionnaires were used to collect information for the SES07: a household questionnaire, a diary questionnaire for recording household food consumption in urban households, a community questionnaire and a price questionnaire. Table 1 provides information on the contents of the survey instruments.

Table 1. Contents of SES07 Survey Questionnaires

Section	Level	Description
<b>HOUSEHOLD QUESTIONNAIRE</b>		
0. General Questions	Individual	Information on the location of the household (cluster number, household number, aimag/capital, soum/district, and location) and information on the enumerator, supervisor, and data entry operator. This section also collects information on the dates on which visits to the household occurred and whether or not the survey was completed.
1. Household Roster	Individual	This module contains the roster of individuals living in the household, their relationship to the household, gender, year of birth, age, and marital status. It also allows the identification of spouses, and parents of household members.

2. Education	Individual	For all household members 6 years of age and older, information is collected on highest education level achieved (for those not currently in school), reasons for drop outs, and for those members who never attended school, the reasons why. For those members currently in school, information is collected on current grade level, location of the school, how the student travels to and from school, tuition paid, receipt of tuition assistance, and expenditures on education for the last 12 months.
3. Health	Individual	The Health section collects information on insurance coverage, chronic illnesses, disabilities, health complaints in the last month, treatment sought for health complaints in the last month, expenditures for health treatments, hospital stays in Mongolia or abroad, and infectious diseases in the past 12 months.
4. Reproductive Health	Individual	For women 15-49 years of age, information is collected on abortions, and contraceptive use.
5. Migration	Individual	For all members 15 years of age and older, information is collected on birth location, whether or not the member has ever migrated, reasons for moving to present location, and if the member lived in the present location in 2003.
6. Jobs in Last 12 Months	Individual	All members 10 years of age and older are asked to describe all of the jobs that they have worked in the last 12 months. They are asked what kind of work it was (unpaid, wage, self employment, own enterprise), and in which of the last 12 months they worked that job. In addition, they are asked the average number of hours worked per week in each job.
7. Wage Jobs	Individual	For all wage jobs listed in Section 6, individuals are asked detailed questions about what the job was, who the employer was, transportation used to go to work, wages received, and if the employer paid for medical insurance or social insurance.
8. Job Search	Individual	For all household members 10 years of age and older who did not work in the last 12 months, information is collected on the reason why they didn't work, whether or not they looked for a job, if so, what methods were used, and whether or not the member registered at the Employment and Social Welfare Office.
9. Agricultural Production	Household	The section is divided into 4 parts: A-1. Crops A-2. Crop expenditures and agricultural equipment B-1. Herding B-2. Production and consumption of herding byproducts
10. Non-Farm Household Enterprises	Household	This section collects information on up to 3 household enterprises operated during the last 12 months. Information includes location of the enterprise, expenditures for the enterprise and receipts for the enterprise in an average month, in a bad month, or in a good month.

11. Other Income	Individual	Part A of this section collects information on members that receive income from non-employment sources such as pensions, benefits, social welfare, rent, sale of assets, and interest and dividends among others. Part B collects on information on income from remittances and aid.
12. Savings and Loan	Household	Information is collected on reasons for loans to be taken, source of the loan, possibility to repay the loan, and if the household has been refused loans in the past.
13. Housing and Energy	Household	This section collects information on the materials that the dwelling is made from, how the dwelling was acquired, source of heating, source of electricity, source of water, method of waste disposal, and use of internet.
14. Durable Goods	Household	For a list of 42 items, the household indicates if the item is owned, how old the item is, and its current value.
15. Non-Food Expenditure and Consumption	Household	For a comprehensive list of non-food items, taken from the HIES data, information is collected on purchases in the last month, and the last 12 months, and receipt of the items for free during the past 12 months.
16. Purchases of Food During the Past Month (Recall)	Urban Households	On the first visit to the urban household, information is collected on whether or not the household has consumed a list of food items during the past 12 months. If yes, the amount of the item currently in stock is collected. On the last visit to the urban household, approximately 30 days later, for each item, the household is asked if the item was purchased during the past month, how much of the item was purchased, and the unit price for the purchase. Information is also collected on how much of each item was, in the past month: produced by the household; received free; sold; given to others; and the current stock.
17. Consumption of Food and other Frequently Purchased Commodities (From Diary)	Urban Households	Three diaries are left at the household and collected approximately every 10 days. At the end of the 10-day period, the enumerator collects the diaries and records the information on the amount of food consumed during the 10 day period and the source of the food that was consumed.
18. Purchases of Food During the Past Months (By Recall)	Rural Households	Information is collected on whether or not the rural household has consumed a list of food items during the past 12 months. The household is asked if the item was purchased during the past month, how much of the item was purchased, and the unit price for the purchase. Information is also collected on how much of each item was, in the past month: in stock at the beginning of the month, produced by the household; received free; sold; given to others; and the current stock.
19. Consumption of Food During the Past 7 Days (By Recall)	Rural Households	Rural households are asked how much food was consumed in the past 7 days, and the source of that consumed food (purchases, received for free, from own production).
COMMUNITY QUESTIONNAIRE		
A. Population and Households	Community	Demographics of the community.
B. Economy and Infrastructure	Community	General characteristics of the community in terms of economic activities, local infrastructure, and the working population.

C. Education	Community	Availability of education services in the soum.
D. Health	Community	Availability of health services in the soum.
E. Agriculture and Livestock	Community	Agricultural activities common in the community
F. Non-Agricultural Business	Community	Non-agricultural activities in the soum.
PRICE QUESTIONNAIRE		
Prices	Community	Prices were collected at three different locations in each area selected for inclusion in the survey.

Food consumption data was collected differently in urban and rural households. In urban households, on the first visit to the household, respondents were asked about food items consumed during the last 12 months. For any items consumed during the last 12 months, the amount of that item in stock on the date of the interview was collected. The first of three diaries was then left at the household to record food consumed over a 10 day period. Over the course of 30 days, urban households fill in the diaries with the food items, quantity consumed and how the household acquired the food (purchase, received free, produced by the household). This information was later transcribed by the interviewers into Section 17 of the household questionnaire. On the last visit to the household, the respondent was asked to provide recall information on purchases of food during the last month.

In rural households, on the first visit to the household, respondents were asked about food items consumed during the last 12 months. For any items consumed during the last 12 months, the amount of that item in stock on the date of the interview was collected. The respondent was also asked to recall the amount of food consumption, by food item, the household had during the last 7 days.

The diary method was used in urban areas in order to maintain comparability with the earlier HIES surveys which use a diary to collect information. The recall method was used in rural areas because there was less possibility of being able to visit rural households three times in one month. Rural households are, in large part, nomadic households and the possibility of following the households as they move across the country would have added to the costs of the survey.

## Sample Design<sup>1</sup>

A total sample of 11,232 households was allocated into three major strata as follows:

- Ulaanbaatar .....3,600 households
- Aimag Centers .....2,640 households
- Rural Area .....4,992 households

---

<sup>1</sup> For a complete description of the sample, see Appendix B.

The sample was implicitly allocated by districts and *horoos*<sup>2</sup> in Ulaanbaatar, and by aimags in the Rural Area. Each aimag center was an explicit sub-stratum, with 240 households allocated to Darhan-Uul and Orhon, and 120 households to each of the other aimag centers. The Govisumber aimag was an explicitly excluded stratum.

## Training

Training for trainers was held during 11-16 June 2007. At this time, the trainers who would be training the enumerators were given extensive instructions in the details of the questionnaires. During this time, the questionnaires were reviewed on a question by question basis to familiarize the trainers with all aspects of the questionnaires and to prepare them for potential questions from the enumerators during the next phase of training.

Training for the supervisors, enumerators and data entry clerks was held from 18-29 June 2007. Facilities at a local university were used. The 101 enumerators were divided into four classrooms with two trainers per classroom. Training was done through power point presentations and practical examples.

## Field Work

Field work for the SES 07 began on 2 July 2007 and went through XX 2008. Field work teams consisted of one supervisor, three enumerators and one driver. In total there were XX teams. The supervisors were responsible for making sure that the interviewers had the materials they needed and for making sure that any problems that came up in the field were resolved. Each team was assigned, on average, XX segments to interview.

## Data Entry

The data entry program was designed using CSPro, a data entry package developed by the US Census Bureau. This software allows programs to be developed to perform three types of data checks: (a) range checks; (b) intra-record checks to verify inconsistencies pertinent to particular module of the questionnaire; and (c) inter-record checks to determine inconsistencies between the different modules of the questionnaire.

The data were key entered at the NSO headquarters in Ulaan Baatar starting XX 2007.

---

<sup>2</sup> Mongolia is divided into 22 *aimags*. The largest of them – Ulaanbaatar – is subdivided into 9 districts, 121 *horoos* and 1,035 *khesegs*. Each *kheseg* has approximately 200 households. The rest of the country is divided into *soums* and *bags*. One of the soums in each aimag is normatively considered as the *Aimag Center* and the others as the *Rural Area*.



## **Appendix A**

### **HOW TO OBTAIN COPIES OF THE DOCUMENTATION AND DATA**

Copies of the documentation for the 2007 Socio-Economic Survey can be downloaded from the LSMS Web Site:

<http://www.worldbank.org/lsms/lsmshome.html>

or obtained by mail through the LSMS Office (see address below). It is recommended that individuals who are interested in using the data for analyses read the documentation prior to requesting copies of the data to ensure that the data will be useful for their analyses.

LSMS Database Manager  
Development Research Group  
The World Bank  
1818 H Street, NW  
MSN MC3-306  
Washington, DC 20433  
Tel: (202) 473-9041  
Fax: (202) 522-1153  
e-mail: [lsms@worldbank.org](mailto:lsms@worldbank.org)

Government permission is required prior to the World Bank being allowed to distribute the data from the 2007 Socio-Economic Survey. Users who are interested in using the data should contact:

Mr. P. Byambatseren  
Chairman  
National Statistical Office of Mongolia  
Government Building III  
Baga Toiruu-44  
Ulaan Baatar-20A  
Mongolia  
e-mail: [nso@magicnet.mn](mailto:nso@magicnet.mn)

Users should explain that they would like to use the data and give a brief description of their research proposal. Please copy this e-mail to LSMS office ([lsms@worldbank.org](mailto:lsms@worldbank.org)). The National Statistical Office will authorize the data release and provide instructions on how to provide the National Statistical Office with a processing fee.

In addition to the fee for the National Statistical Office, the World Bank charges a processing fee to distribute the data. Please check the LSMS Web Site or contact the LSMS Office for the most current information.

Once permission to use the data has been received from the National Statistical Office, users should send:

- \* a check, made out to the World Bank, for the World Bank processing fee;
- \* a brief description of the research proposal; and
- \* a copy of the permission from the National Statistical Office.

Individuals who receive copies of the data agree to:

- \* give recognition to the National Statistical Office as the source of the data in all publications, conference papers and manuscripts;
- \* make copies of all reports and documents resulting from the research on the data available to the National Statistical Office and to the Poverty and Human Resources Division of the World Bank; and
- \* not pass the data to any third parties for any reasons.

Researchers found to be in violation of these agreements will not be able to receive copies of other data sets from the LSMS Office in the future.

## **Appendix B**

### **Mongolia Socioeconomic Survey – 2007**

#### **Sampling design and implementation**

**J. Muñoz<sup>3</sup>**

#### **Introduction**

This appendix describes the sampling strategy adopted by the 2007 Mongolia Socioeconomic Survey (SES) and the details of its practical implementation until immediately before the survey was fielded, on July 1<sup>st</sup> 2007. It supersedes another two documents on sampling prepared in the course of the past six months<sup>4</sup>, reproducing the parts of them that do not deserve amendments as a result of recent circumstances.

### **Outline of the design**

#### ***Sample sizes and strata***

A total sample of 11,232 households was allocated into three major strata as follows:

- Ulaanbaatar .....3,600 households
- Aimag Centers.....2,640 households
- Rural Area .....4,992 households

The sample was implicitly allocated by districts and *horoos*<sup>5</sup> in Ulaanbaatar, and by aimags in the Rural Area. Each aimag center was an explicit sub-stratum, with 240 households allocated to Darhan-Uul and Orhon, and 120 households to each of the other aimag centers. The Govisumber aimag was an explicitly excluded stratum.

#### ***Sampling stages***

The selection strategy was different in each of the three major strata:

In Ulaanbaatar the sample was selected in two stages:

- First, 360 *khesegs*.
- Second, 10 households in each *kheseg*.

In The Aimag Capitals, in two stages:

- First, 12 or 24 *bags* in each aimag center.
- Second, 10 households in each *bag*.

---

<sup>3</sup> This appendix was prepared by Juan Muñoz (Consultant).

<sup>4</sup> Muñoz, J. *Sampling design of the Mongolia Integrated Household Income and Expenditure Survey and Living Standards Measurement Survey*. Mission reports (December 2006 and April 2007.)

<sup>5</sup> Mongolia is divided into 22 *aimags*. The largest of them – Ulaanbaatar – is subdivided into 9 districts, 121 *horoos* and 1,035 *khesegs*. Each *kheseg* has approximately 200 households. The rest of the country is divided into *soums* and *bags*. One of the *soums* in each aimag is normatively considered as the *Aimag Center* and the others as the *Rural Area*.

In the Rural Areas, in three stages:

- First, 52 soums.<sup>6</sup>
- Second, 12 bags in each soum.
- Third, 8 households in each bag.

In the first sampling stage, area units (khesegs, bags or soums) were selected with probability proportional to size (PPS,) using as a measure of size the number of households at the time of the 2005 administrative registration.

Subsequent sampling stages used as sample frame updated lists of households compiled and computerized by the NSO aimag or district offices. In rural soums, administrative records generally sort the *suurin* (sedentary population) and the *malchid* (nomadic herders,) into separate bags, but even when that was not the case (that is, when a bag contained both *suurin* and *malchid*,) each category was still considered separately for sampling purposes (in other words, each mixed bag was conceptually split into a *suurin* sub-bag and a *malchid* sub-bag.) This was also done in the few aimag centers that also reported the presence of herders.

The household lists contain the address, size, and name of the head of each household, and whether any of its members is known to work as a non-agricultural self-employed. The group of 8 (or 10) households to be visited by the survey in each kheseg or bag (hereafter referred to as a *cluster*) was selected from these lists as follows:

- In non-herder units, by systematic, unequal probability sampling, giving the households assumed to contain self-employed twice as many chances of being selected than the rest of the households in the unit.
- In herder bags, by equal probability, circular cluster sampling.

### ***Reserve households***

In addition to the 8 (or 10) households targeted by the survey, three extra households were selected for each cluster, with the intention of being used as a reserve for eventual nonresponse among the target households. This was done in practice by selecting 11 (or 13)-household clusters first, and then three among them by systematic equal probability sampling to make the reserve.

### ***Multiple clusters in large area units***

In large khesegs or bags that were selected more than once by the standard PPS procedure, the corresponding number of clusters was selected in the final stage. (For instance, if a unit was selected twice in the first stage, two clusters were selected in the unit in the second stage.) This happened seldom in Ulaanbaatar but often in the rest of the country. The selection procedure was as follows:

- In non-herder units, all households in all required clusters were selected in a single step – with unequal probabilities as described above – and then allocated to specific clusters by systematic, equal probability sampling.

---

<sup>6</sup> This resulted in the allocation of 1 to 4 soums per aimag.

- In herder bags, the starting points of the required clusters were equally spaced in the list of households.

### Allocation of the sample in time

The 360 clusters selected in Ulaanbaatar, and the 12 or 24 clusters selected in each aimag center or rural soum were randomly allocated into the 12 months of survey fieldwork. The survey will thus visit a random sub-sample of 104 clusters (936 households) each month.

### Description of the sample

Table 1 below gives the total number of households in the population, the number of area units and the number of households of the proposed design, by stratum and by aimag.

**Table 1: Mongolia Socioeconomic Survey 2007**  
Number of households in the population,  
and number of soums, clusters and households in the sample  
by major stratum and by aimag.

Aimag	Total No. of households				Selected rural soums	Selected clusters				Selected households			
	Ulaan baatar	Aimag Centers	Rural Areas	Total		Ulaan baatar	Aimag Centers	Rural Areas	Total	Ulaan baatar	Aimag Centers	Rural Areas	Total
1 Arhangai		4,399	19,877	24,276	4		12	48	60		120	384	504
2 Bayan-Ulgii		6,289	15,039	21,328	3		12	36	48		120	288	408
3 Bayanhongor		6,433	14,502	20,935	3		12	36	48		120	288	408
4 Bulgan		3,081	11,935	15,016	2		12	24	36		120	192	312
5 Govi-Altai		4,666	10,807	15,473	2		12	24	36		120	192	312
6 Dornogovi		5,121	8,847	13,968	2		12	24	36		120	192	312
7 Dornod		9,603	8,484	18,087	2		12	24	36		120	192	312
8 Dundgovi		3,429	9,199	12,628	2		12	24	36		120	192	312
9 Zavhan		4,074	15,855	19,929	3		12	36	48		120	288	408
10 Uvurhangai		5,566	23,227	28,793	5		12	60	72		120	480	600
11 Umnugovi		4,358	8,440	12,798	2		12	24	36		120	192	312
12 Suhbaatar		3,510	9,829	13,339	2		12	24	36		120	192	312
13 Selenge		4,691	17,502	22,193	3		12	36	48		120	288	408
14 Tuv		3,657	19,652	23,309	4		12	48	60		120	384	504
15 Uvs		6,345	13,455	19,800	3		12	36	48		120	288	408
16 Hovd		6,675	12,803	19,478	2		12	24	36		120	192	312
17 Hovsgul		8,672	22,111	30,783	4		12	48	60		120	384	504
18 Hentii		4,545	13,396	17,941	3		12	36	48		120	288	408
19 Darhan-Uul		18,666	4,075	22,741	1		24	12	36		240	96	336
20 Ulaanbaatar	216,342			216,342		360			360	3,600			3,600
21 Orhon		20,059	811	20,870	0		24	0	24		240	0	240
22 Govisumber		2,301	944	3,245	0		0	0	0		0	0	0
<b>Total</b>	<b>216,342</b>	<b>136,140</b>	<b>260,790</b>	<b>613,272</b>	<b>52</b>	<b>360</b>	<b>264</b>	<b>624</b>	<b>1,248</b>	<b>3,600</b>	<b>2,640</b>	<b>4,992</b>	<b>11,232</b>

### Justification of the sampling design

The SES sampling strategy updates the design adopted by the Household Income and Expenditure Survey (HIES) in 2001, on the basis of the 2000 census.<sup>7</sup> The amendments take three major factors into consideration:

- The availability of a more recent sample frame, developed by the NSO on the basis of 2005 population figures from the local registration agencies. These figures show major changes in the population distribution over the past five years (Ulaanbaatar, for instance, has grown from 755,000 to 912,000 in the period,) revealing that progressively larger parts of the population were escaping from the scope of the surveys based on the 2001 census.

<sup>7</sup> For a description of the HIES design, see Levinson, A. *Mongolia Income and Expenditure Survey*. Ulaanbaatar, October 2000; and Muñoz, J. *Mongolia Household Survey System: Sampling Implementation and Survey Integration*. Ulaanbaatar, February 2001.

- A raising concern for measuring and understanding the economic importance of the self-employed (often but wrongly referred to as “the informal sector.”) The SES design responds to this challenge by selecting households known to contain self-employed preferentially in the final sampling stage, and by marginally increasing the portion of the sample allocated to Ulaanbaatar, where most of the self-employment activities are known to be concentrated.
- The fact that the Socioeconomic Survey will need to be fielded with an operational strategy that is significantly different from the one used by the HIES so far. The proposed design recognizes that the integrated survey will need well-trained interviewers, organized into teams and devoted to the survey on a full-time basis, and that it will also need a decentralized data management component that integrates computer-based quality controls to fieldwork. Several teams will operate in Ulaanbaatar and one or two teams in each of the other aimags.

The SES sample conserves many features of the HIES, particularly a total sample size of 11,232 households per year. However, while the HIES recognized only four explicit strata ([1] Ulaanbaatar, [2] Aimag Capitals and Small Towns, [3a] Soum Centers and [3b] Countryside,) the SES collapses the small towns with the rural areas, and explicitly recognizes many more aimag-level strata. These enhancements are justified by various reasons:

- Neither the HIES nor the SES will be able to produce reliable aimag-level estimations. The only way of achieving this would be to significantly increase the total sample size, to around a thousand households per aimag – something that cannot be recommended at this moment. However, the allocation of at least a minimum sample to each aimag and the decentralization of fieldwork and data management are significant first steps in the direction of this ambitious goal.
- Small towns are indeed very small (most of them have less than 2,000 households) and collectively represent only about 10 percent of the population. Under these conditions their presence as partners of aimag capitals in an explicit stratum was not compatible with a minimum urban size in each aimag, and with the deployment of at least one interviewer in each aimag capital. It is important to underline that the SES design does not exclude small towns from the sample – they are just transferred to another stratum. This may require some care at the analytic stage, but does not engage the comparability of the SES with the HIES series.

Another deviation of the SES relative to HIES is that, although the total size of the rural sample (4,992 households) remains unchanged, it will now consist of 96 households (12 bags) in each of 52 soums, rather than of 64 households (8 bags) in each of 78 soums. This probably will bring about slightly higher cluster effects, but it will also improve the capture of seasonal variations, and will keep rural interviewers busy year round at a marginal cost.

A final difference between the SES and HIES samples is in the selection technique adopted for the herder communities: whereas the HIES selected both herder and sedentary clusters by systematic sampling within their respective bags, the SES will try to facilitate fieldwork selecting herders by circular cluster sampling instead – the underlying assumption being that households that appear close to each other in the lists are also likely to be neighbors in the field. Cluster sampling may bring about larger sampling errors than systematic sampling, but it is also expected to be less vulnerable to nonresponse and to the selection biases that have affected the recent rounds of the HIES.

## Implementation of the sampling stages

### *Initial sampling stages*

In December 2006, the 2005 population registration figures were organized into adequate sample frames for the three major strata, and the first sampling stage was conducted in Ulaanbaatar and the rural areas. This resulted in the selection of 357 kheseqs in the capital (three of them selected twice,) and 52 rural soums (allocated into aimags as shown in Table 1.) A household listing operation was conducted in all of these kheseqs and soums.

The first sampling stage was also conducted in the aimag centers, but the relatively small total number of bags per center resulted in almost all of them being chosen.<sup>8</sup> Under these conditions, it would have been easier and better to ignore this preliminary exercise and simply conduct the household listing operation in all urban centers, leaving the actual allocation of clusters into bags for a later moment, when updated figures would be available for all of them. This was actually done in all but four of the aimags: Bayan-Ulgii, Govi-Altai, Tuv and Uts. In each of these aimags, the listing operation took place in all but one of the bags, thus making the December exercise a *de facto* zero-th sampling stage that will require a very small adjustment in one of the estimation formulas presented below (Formula 2.)

### *Household listing operation*

The household listing operation was conducted in May and June of 2007 in all selected kheseqs of Ulaanbaatar, all 52 selected rural soums, and almost all aimag centers, as explained above. The final database contains over 275,000 households in total and, beyond its immediate utilization for the SES, has a high potential value as a master sample frame for other household surveys conducted by the NSO in future years.

A sudden rescheduling of the survey launching date prevented the implementation of uniform practices for the computerization of the household lists, and resulted in a wide variety of heterogeneous spreadsheets demanding considerable data management effort in order to build from them a reliable frame for the subsequent sampling stages. One of the aimags (Hentii) delivered its files only days before the launching of the survey, and in some of its bags the lists consisted of simple sequences of serial numbers, without household addresses or names.

The operation was also confused in one of the Ulaanbaatar districts (Baianzurj,) where the maps of four horoos had been redrawn and local staff had trouble identifying the boundaries of the twenty “old” kheseqs that had been selected in these horoos. This was solved by taking a sample of twenty “new” kheseqs in the affected horoos, using the most recent population figures for the PPS selection. The December 2006 exercise can also be considered in this case as a *de facto* zero-th sampling stage demanding a small adjustment in Formula 1 below.

Some aimags or districts delivered lists of a few extra kheseqs or bags, in addition to those they had been asked for. These additional units were ignored in the subsequent sampling stages.

---

<sup>8</sup> There are 168 bags in total in all aimag centers (Ulaanbaatar and Govisumber excluded.) All but seven of them were selected in December, and half of those selected were so more than once.

### ***Selection of clusters***

The final allocation of clusters into bags and the selection of the target and reserve households in each cluster were conducted with a special-purpose program (an Excel macro.) Additional programs were responsible for the production of the two associated survey instruments:

- The ***Cluster Tracking Sheets*** will be used by fieldworkers to identify the selected households. They give the name of the household head, the household size and the address of the 8 (or 10) target households and 3 reserve households in a cluster.
- The ***Household Listing Printouts*** reproduce the list of all households in the area, flagging the selected households among them. These lists will facilitate the identification of the selected households when precise addresses are not available.

Figures 1 and 2 below reproduce one Cluster Tracking Sheet and a page of the corresponding Household Listing Printout. Notice that although in sedentary areas households known to contain self-employed are preferentially chosen, neither of these instruments reveal this information to the fieldworkers. They have been instructed to probe for all self-employment activities of all members in all households, not just in some of them.

**Figure 1: A Cluster Tracking Sheet  
(50 percent of actual size)**

[illegible]



АА	ОА	АА	АА	АА	ОААА	ОАААА
1		4	П.Борис	Я-37-152		
2		4	М.Цолмон	Я-37-151		
3	1143	01	М.Жамбаатар	Я-37-151		
4		3	Д.Нацаг	Я-37-150		
5		3	Н.Цэцэгбал	Я-37-149		
6		4	Л.Наранц	Я-37-147		
7		5	Д.Равиш	Я-37-147		
8		4	О.Авирмад	Я-37-146		
9		3	С.Наранцгарал	Я-37-146		
10		3	В.Могойтвал	Я-37-146		
11		4	Б.Ирээбат	Я-37-145		
12		6	П.Сергей	Я-37-143		
13		3	П.Валеритина	Я-37-143		
14		8	Б.Батбаяр	Я-37-142		
15		3	О.Мундраалер	Я-37-142		
16	1143	02	С.Луган	Я-37-142		
17		7	Н.Дамбаас	Я-37-141		
18		2	Т.Мансал	Я-37-140		
19		5	Т.Ганболд	Я-37-139		
20		3	Ч.Гармаа	Я-37-139		
21		4	Б.Ганбаатар	Я-37-139		
22		4	Б.Ганболд	Я-37-139		
23		5	Отгон	Я-37-139		
24		5	Ч.Ганболд	Я-37-558		
25		4	Х.Энхбаатар	Я-37-558		
26		3	Ц.Омид	Я-37-559		
27		2	Д.Баяраа	Я-37-560		
28		4	К.Коля	Я-37-563		
29	1143	11	Г.Лоренцолгор	Я-37-563		
30		6	Г.Аюузраа	Я-37-563		
31		3	Ш.Цэрэннадмид	Я-37-563		
32		3	Ц.Бандарья	Я-37-564		
33		6	Н.Жамъянбаатар	Я-37-565		
34		3	Х.Энхбаяр	Я-37-566		
35		6	Ч.Могойвар	Я-37-567		
36		7	О.Халиукай	Я-37-568		
37		6	Г.Эрдэнэцогт	Я-37-569		
38		3	Я.Янцгармаа	Я-37-571		
39		4	Д.Тогтох	Я-37-572		
40		1	Л.Шарх	Я-37-574		
41	1143	03	Д.Энх-Амгалан	Я-37-573		
42		3	С.Эрдэнэс	Я-37-575		
43		2	Ц.Энхтува	Я-37-575		
44		3	Б.Батсайхан	Я-37-576		
45		4	Ж.Цулунням	Я-37-571		
46		3	Н.Бат-Эрдэнэ	Я-37-572		
47		2	Ч.Индэрлалмаа	Я-37-03		
48		3	Д.Наранццогт	Я-37-06		
49		8	У.Оюунчимэг	Я-37-07		
50		2	У.Оюунчимэг	Я-37-07		
51		1	Д.Понкид	Я-38-578		
52		2	Ш.Нанзад	Я-38-579		
53		4	Н.Дашням	Я-38-579		
54		3	Д.Өнөржаргал	Я-38-579		
55	1143	04	С.Аюузраа	Я-38-580		

***Ulaanbaatar***

$$p_{ij} = \frac{360 n_i}{216,342} \times \frac{10\mu_{ij}}{\sum_{\alpha=1}^{n'_i} \mu_{i\alpha}} = \frac{3,600 n_i \mu_{ij}}{216,342 \sum_{\alpha=1}^{n'_i} \mu_{i\alpha}} \dots\dots\dots (1)$$

360 is the number of clusters selected in Ulaanbaatar:

216,342 is the number of households in Ulaanbaatar, as recorded in the first stage sample frame (see Table 1:)

10 is the number of households per cluster;

$n'_i$  is the number of households in khesege  $i$ , as reported by the 2007 household listing operation;

$\mu_{i\alpha}$  is the measure of size assigned to household  $i\alpha$  in the second sampling stage:  $\mu_{i\alpha} = 2$  for households assumed to contain self-employed and  $\mu_{i\alpha} = 1$  for the rest of the households.

### **Aimag Centers**

In the aimag centers, the selection probability of household  $hij$  in bag (or sub-bag)  $hi$  of aimag  $h$  is given by

$$p_{hij} = \frac{b_h n_{hi}}{N_h} \times \frac{10 \mu_{hij}}{\sum_{\alpha=1}^{n_{hi}} \mu_{hi\alpha}} \dots\dots\dots (2)$$

where

$b_h$  is the number of clusters selected in the aimag center:  $b_h = 24$  in Darhan-Uul and Orhon and  $b_h = 12$  in the other aimag centers;

$n_{hi}$  is the number of households in bag  $hi$ , as recorded in the 2007 household listing operation;

$N_h$  is the number of households in the aimag center, as reported by the household listing operation;

10 is the number of households per cluster;

$\mu_{i\alpha}$  is the measure of size assigned to household  $i\alpha$  in the second sampling stage:  $\mu_{i\alpha} = 2$  for households assumed to contain self-employed and  $\mu_{i\alpha} = 1$  for the rest of the households.

### **Rural Areas**

In rural areas, the selection probability of household  $hij$  in bag (or sub-bag)  $hi$  of soum  $h$  is given by

$$p_{hij} = \frac{52 N_h}{260,790} \times \frac{12 n'_{hi}}{N'_h} \times \frac{8}{n'_{hi}} = \frac{4,992}{260,790} \times \frac{N_h}{N'_h} \dots\dots\dots (3)$$

where

52 is the total number of rural soums selected;

$N_h$  is the number of households in soum  $h$ , as recorded in the 2005 administrative registration files used as a sample frame for the first sampling stage;

260,790 is the total number of rural households, as recorded in the first stage sample frame (see Table 1;)

12 is the number of clusters selected in soum  $h$ ;

- $n'_{hi}$  is the number of households in bag  $hi$ , as recorded in the 2007 household listing operation;
- $N'_h$  is the number of households in soum  $h$ , as reported by the household listing operation; and
- 8 is the number of households per cluster.