BASIC INFORMATION DOCUMENT

Bosnia-Herzegovina
Living Standards Measurement Study Survey
2001

Table of Contents

Chapters

TABLE OF CONTENTS, cont.

Tables

1. Introduction

In 1992, Bosnia-Herzegovina, one of the six republics in former Yugoslavia, became an independent nation. A civil war started soon thereafter, lasting until 1995 and causing widespread destruction and losses of lives. Following the Dayton accord, Bosnia-Herzegovina (BiH) emerged as an independent state comprised of two entities, namely, the Federation of Bosnia-Herzegovina (FBiH) and the Republika Srpska (RS), and the district of Brcko. In addition to the destruction caused to the physical infrastructure, there was considerable social disruption and decline in living standards for a large section of the population. Along side these events, a period of economic transition to a market economy was occurring. The distributive impacts of this transition, both positive and negative, are unknown. In short, while it is clear that welfare levels have changed, there is very little information on poverty and social indicators on which to base policies and programs.

In the post-war process of rebuilding the economic and social base of the country, the government has faced the problems created by having little relevant data at the household level. The three statistical organizations in the country (State Agency for Statistics for BiH – BHAS, the RS Institute of Statistics-RSIS, and the FBiH Institute of Statistics-FIS)[1] have been active in working to improve the data available to policy makers: both at the macro and the household level. One facet of their activities is to design and implement a series of household series. The first of these surveys is the Living Standards Measurement Study survey (LSMS). Later surveys will include the Household Budget Survey (an Income and Expenditure Survey) and a Labor Force Survey. A subset of the LSMS households will be re-interviewed in the two years following the LSMS to create a panel data set.

The three statistical organizations began work on the design of the Living Standards Measurement Study Survey (LSMS) in 1999. The purpose of the survey was to collect data needed for assessing the living standards of the population and for providing the key indicators needed for social and economic policy formulation. The survey was to provide data at the country and the entity level and to allow valid comparisons between entities to be made.

The LSMS survey was carried out in the Fall of 2001 by the three statistical organizations with financial and technical support from the Department for International Development of the British Government (DfID), United Nations Development Program (UNDP), the Japanese Government, and the World Bank (WB). The creation of a Master Sample for the survey was supported by the Swedish Government through SIDA, the European Commission, the Department for International Development of the British Government and the World Bank.

The overall management of the project was carried out by the Steering Board, comprised of the Directors of the RS and FBiH Statistical Institutes, the Management Board of the State Agency for Statistics and representatives from DfID, UNDP and the WB. The day-to-day project activities were carried out by the Survey Mangement Team, made up of two professionals from each of the three statistical organizations.

[1] In principal, the BHAS, is the country level statistical office responsible for collating information from the two entity level statistical institutes (FIS and the RSIS) and for setting country-wide standards in the field of statistics. The two entity-level statistical offices are responsible for data collection and collation within their respective entities. At the time of the survey, the political status of Brcko was still under discussion and did not yet have a separate statistical office as it now does.

The present document is designed to provide data users with the information they need to understand the LSMS data set and to use the data appropriately. The next section provides a summary of the survey instruments. Section 3 outlines the sample design used and the weighting needed when using the LSMS data. Sections 4 and 5 discuss the pilot survey that was done and the organization of the actual survey in terms of field work. The final section provides a description of the LSMS data set, starting with the data entry system and then providing detailed information on the structure of the data sets.

2. Survey Instruments

The LSMS in Bosnia-Herzegovina is a multi-topic household survey covering a wide range of topics that affect welfare: housing, education, health, labor, migration, credit, vouchers, social assistance, consumption, agricultural and non-agricultural activities. The LSMS was designed to collect the information required for an assessment of living standards and to provide the key indicators required for social and economic planning. *Inter alia*, the LSMS in BiH was designed to measure welfare in both monetary and non-monetary terms. Detailed information was collected on household consumption (expenditures, home production, use value of housing and durables), on social assistance such as old age pensions, war veterans pensions, assistance received by orphans, widows, and on sources of income. Non-monetary measures include detailed information on housing, and access to, and the use of, public services such as education and health.

In addition to the household questionnaire, a price questionnaire was also administered to identify the variations in price levels of key food products in the different municipalities covered by the survey.

2.1 Household Questionnaire

The overall content of the household questionnaire and the individual questions included in it were designed to address the specific situation of the country and the data needs of policy-makers. In addition, several sections of the questionnaire were based on draft questionnaires for future surveys (the HBS and the LFS) and/or older surveys and thus will be helpful in allowing some tracking of indicators over time. The process of designing the questionnaire was lengthy and involved an inter-institutional team from the three statistical organizations of the country—the Survey Management Team. Although efforts to create a formal data users' group of line ministries and other users were not successful, several ministries did provide detailed comments and suggestions on the modules relevant to their ministries.

The complete list of modules included in the household questionnaire can be found in Table 1. It is worth noting the importance of several of these in the BiH context. First, the migration module collected information on present status: given the dislocation of the population by the war and the legal ramifications of present status this module was considered to be of great importance. Second, a module on non-agricultural household businesses was used as the existing administrative data in the country cannot provide any information for assessing the prevalence or size of this sector. Third, in the health module questions pertaining to depression were added to determine how prevalent this ailment was given the post-conflict situation. Fourth, a module on anthropometric measurement of

children was not included: a recent Multiple Indicators Cluster Survey (MICs) done by UNICEF had shown that malnutrition was negligible in the country.[2]

Table 1: Contents of BiH-LSMS Household Questionnaire

Module	Description	Respondent
Round I		
Roster	Basic demographic information on the household. The module was used to list all the members of the household, their relationship to the head of household and other household members, their age, sex, and marital status. Information was also collected about individuals absent from the household.	Head of Household
Housing	Information on the housing in which the household lives as well as utilities used. The module has four sections: **A.** Description of Primary Residence: Type and condition of dwelling, number of rooms, living area, and presence of utilities such as electricity, water, sewerage, and telephone. **B.** Legal status of ownership of dwelling unit: Legal status as well as expenditures on housing and related services, **B2.** Ownership and Purpose of Secondary Residence **C.** Durable goods: Ownership, date of purchase and present value of such goods	Head of Household
Education	Data on levels of schooling, attendance and characteristics of schooling, including:	
	A. Child Care and Kindergarten: Attendance and monthly expenditures for child care or pre-school. The section was administered for children from 0 through 6 years of age.	Parent or Guardian of child
	B. General Education: Literacy status, educational qualifications and specialization, type of schools attended, formal and informal education expenditures, source of financial assistance during the academic year 2000-2001, distance of the school from home etc. The section was administered to all persons 7 years and older and for children less than 7 years who attended school.	Each, individual household member age 15 or older. For children under 15 years of age the parent or guardian responded for the child.
Health	Data on health status and use of health services including:	
	A. Utilization of Health Care Services: Use of different levels/types of health services , self medication and all health expenditures. Questions were also included on the prevalence of chronic ailments and the availability of health insurance. The section was administered for all household members, regardless of age.	Each household member age 15 or older. The parent or guardian responded for children under 15.
	B. Health Status: This section elicited information on individual's self-reported health status as well as the screening questions for clinical depression. The section was only administered to adults 17 years and older	Each household member aged 17 or older. No indirect informants were allowed in this section

[2] See the UNICEF website for more information on the MICs survey: http://www.unicef.org/research/mics.html.

Table 1: Contents of BiH-LSMS Household Questionnaire, cont.

Module	Description	Respondent
Labor	This module elicited information on labor market activity status during the reference week preceding the survey. For employed persons, information on their occupation, sector of employment, type of employment, place of work, previous employment, number of hours worked in the week and monthly earnings were asked. For unemployed persons, questions were asked on the duration of unemployment, previous employment (sector, occupation), method of seeking work, and whether or not they were registered as unemployed with the Employment Bureau. For inactive persons, questions on present status, previous employment as well as registration at the Employment Bureau were asked. The entire module was administered to persons 15 years and older.	Each individual household member age 15 or older.
Credit	Information was gathered on the number of times the person had borrowed from different sources, amount borrowed during the last 12 months, and the amount presently owed, as well as the month and year of the last borrowing, reasons for borrowing and refusals of loans. The entire module was administered to persons 15 years and older.	Each individual household member age 15 or older.
Privatization Vouchers and Certificates	This module included questions on a person's eligibility for a voucher or certificate, the value of the vouchers or certificates received, transactions made with them, sale value of vouchers or certificates sold, and the nominal value of the vouchers or certificates in their possession. The module was administered for all household members even though the certificates in the Federation were not given to children. But the RS vouchers were and, given that people with rights in one entity can live in another, information was needed for all household members.	Each individual household member age 15 or older. For children under 15 years of age the parent or guardian responded for the child.
Migration	Information collected on the person's (i) current residence, (ii) municipality of birth, (iii) residence prior to the war (April 1992), (iv) reason for migration and (v) current residential status (categories based on migration history not simply present place of residence). The module was administered to persons 15 years and older.	Each individual household member age 15 or older
Social Assistance	This module included questions on (i) the individual's eligibility for old age pension, disability pensions, survivors pensions, and/or war veteran's pension, (ii) monthly pension received, and (iii) the allowances and services received during the preceding 12 months.	Each household member age 15 or older. Parents or guardians responded for children under 15 years of age .
End of First Visit	This module is intended to identify households to be covered by Module 12 (non-agricultural activities) and Module 13 (agricultural activities). It also includes questions on efforts to start a household business (whether this effort was or was not successful) and key problems faced.	Household Head
Round II		
Household Consumption	Each of the following sections elicited information on the quantity and value of purchased items, own production and the value of items received as gifts.	
	A. Daily Expenses: Purchases in the last 7 days of frequently purchased items such as tobacco, cigarettes and meals/snacks eaten outside the home.	Best informed member of the household.
	B. Food Consumption: Average monthly expenditures on items of food consumption such as bread and cereals, meat, fish, edible oil and fat, sugar, and confectionary, other commonly consumed items like salt, vinegar etc, soft drinks, and alcoholic drinks, and, seasonal products such as fruits and vegetables.	Best informed member of the household.

Table 1: Contents of BiH-LSMS Household Questionnaire, cont

Module	Description	Respondent
Household Consumption, cont.	**C**. Non-Food: Monthly Expenditures on such non-food products as transport, cosmetics, fuel , and cleaning products. Annual Expenditures on such non-food products such as clothing and footwear, furniture and fixtures, personal transport, recreation equipments and services, personal care services, financial services, other miscellaneous expenses such as gifts, losses from lottery, thefts etc and expenditures on weddings and other ceremonies	Best informed member of the household
Non-agricultural Household Businesses	This module elicited information from households engaged in non-agricultural business activities: **A**. Identification of enterprises or household businesses: nature of the activity pursued, persons engaged in such activities and the number of such activities. **B.** General Information on enterprise or household business: Length of time the enterprise has been in operation, location, ownership, number of days in a week operated, number of persons engaged. **C.** Labor in Enterprise or Household Business: The number of persons engaged in the business, both household member and non-member, the number receiving wages in cash or in kind. **D.** Revenues and Inputs: The number of months the business operated, gross earnings in an average month, expenses on inputs in an average month **E**. Capital Assets: The value of fixed capital such as land, buildings, equipment and machines, furniture, small and large tools, big vehicles, small vehicles, other fixed assets, value additions to total assets during the past 12 months and main problems faced by the establishment	Best informed member of the household.
Agricultural Activities	This module collected information on farming operations with special focus on farm management, inputs and earnings. **A1.** Land Used: Area of land used by type of use, irrigation, present value of the land, ownership, lease value during 2000-01. **A2**. Unused Land Owned by Household: The type of land, how obtained, present value, time since last used, type of use contract, lease amount received during 2000-01 etc. **B1**. Use of Forest Land: Age of forest, whether the forest was harvested, value of products sold, value of products used by household **B2**. Crop Production and Use: Area of land used by crop, amount harvested, sold, lost to pests, used as wages, used as animal feed, processed, consumed by the household and given away as gifts. **C1**. Inputs and Investments: The quantity of seeds or seedlings used, amount purchased, cost, used from own production, whether obtained as gift and from whom. **C2**. Inputs and Investments—Fertilizers: Quantity used, purchased, cost, obtained as gift and from whom. **C3**. Inputs and Investments--Fuel and Energy: Amount used, purchased, cost, obtained as gift and from whom. **C4**. Inputs and Investments—Labor: Then number of paid workers by job type (soil preparation, sowing and planting, input preparation, weeding, spraying, watering, harvesting, mowing and other), number of paid work days, average daily wage, whether payment was made in-kind.	Best informed member of the household. .

Table 1: Contents of BiH-LSMS Household Questionnaire, cont

Module	Description	Respondent
Agricultural Activities, cont.	**C5**. Inputs and Investments—Machinery: Whether machinery was hired for ploughing, harrowing, other cultivation, sowing and planting, harvesting, mowing, transport or other activities. The source of hire, number of machine hours hired, amount paid per hour and whether payments were made in-kind. **D1**. Livestock: Quantity of livestock and their value. Number sold, consumed, lost, given away, or bought during the past 12 months. Number of new born, number received as gifts, whether any livestock product was sold and its value. **D2**. Animal Feed: Quantity of animal feed used during past 12 months, quantity and value of purchases, own-produced and received as gifts, and source. **E**. Farm Capital Assets: Type of capital assets, their market value, age of the assets, whether the asset is rented out, earnings during 2000-01 from renting out the capital assets.	Best informed member of the household. .

2.2 The Price Questionnaire

A price questionnaire was administered in each group of enumeration areas covered by the survey. Three locations where food is sold (market, shop, etc.) were visited in each area and prices were collected for 39 commonly consumed food items. Limited information on the point of sale was also collected.

It should be noted that a community questionnaire, usually standard in an LSMS survey to collect data on the presence of services and social infrastructure in the areas in which households selected for the survey are situated, was not done in the BiH LSMS.

3. Sample Design and Weighting of Resulting Data

A total sample of 5,400 households was determined to be adequate for the needs of the survey: with 2,400 in the Republika Srpska and 3,000 in the Federation of BiH. The difficulty was in selecting a probability sample that would be representative of the country's population. The sample design for any survey depends upon the availability of information on the universe of households and individuals in the country. Usually this comes from a census or administrative records. In the case of BiH the most recent census was done in 1991. The data from this census were rendered obsolete due to both the simple passage of time but, more importantly, due to the massive population displacements that occurred during the war.

At the initial stages of this project it was decided that a master sample should be constructed. Experts from Statistics Sweden developed the plan for the master sample and provided the procedures for its construction. From this master sample, the households for the LSMS were selected.

3.1 Master Sample[3]

The master sample is based on a selection of municipalities and a full enumeration of the selected municipalities. Optimally, one would prefer smaller units (geographic or administrative) than municipalities. However, while it was considered that the population

[3] This section is based on Peter Lynn's note "LSMS Sample Design and Weighting – Summary". April, 2002. Essex University, commissioned by DfID.

estimates of municipalities were reasonably accurate, this was not the case for smaller geographic or administrative areas. To avoid the error involved in sampling smaller areas with very uncertain population estimates, municipalities were used as the base unit for the master sample.

The Statistics Sweden team proposed two options based on this same method, with the only difference being in the number of municipalities included and enumerated. For reasons of funding, the smaller option proposed by the team was used, or Option B.

3.1.1 Stratification of Municipalities

The first step in creating the Master Sample was to group the 146 municipalities in the country into three strata- Urban, Rural and Mixed – within each of the two entities. Urban municipalities are those where 65 percent or more of the households are considered to be urban, and rural municipalities are those where the proportion of urban households is below 35 percent. The remaining municipalities were classified as Mixed (Urban and Rural) Municipalities. Brcko was excluded from the sampling frame.

Urban, Rural and Mixed Municipalities: It is worth noting that the urban-rural definitions used in BiH are unusual with such large administrative units as municipalities classified as if they were completely homogeneous. Their classification into urban, rural, mixed comes from the 1991 Census which used the predominant type of income of households in the municipality to define the municipality. This definition is imperfect in two ways. First, the distribution of income sources may have changed dramatically from the pre-war times: populations have shifted, large industries have closed and much agricultural land remains unusable due to the presence of land mines. Second, the definition is not comparable to other countries' where villages, towns and cities are classified by population size into rural or urban or by types of services and infrastructure available. Clearly, the types of communities within a municipality vary substantially in terms of both population and infrastructure. However, these imperfections are not detrimental to the sample design (the urban/rural definition may not be very useful for analysis purposes, but that is a separate issue[4]). The classification is used simply for stratification. The stratification is likely to have some small impact on the variance of survey estimates, but it does not introduce any bias.

3.1.2 Selection of Municipalities

Option B of the Master Sample involved sampling municipalities independently from each of the six strata described in the previous section. Municipalities were selected with probability proportional to estimated population size (PPES) within each stratum, so as to select approximately 50% of the mostly urban municipalities, 20% of the mixed and 10% of the mostly rural ones. Overall, 25 municipalities were selected (out of 146) with 14 in the FbiH and 11 in the RS. The distribution of selected municipalities over the sampling strata is shown in Table 2.

[4] It may be noted that the percent of LSMS households in each stratum reporting using agricultural land or having livestock is highest in the "rural" municipalities and lowest in the "urban" municipalities. However, the concentration of agricultural households is higher in RS, so the municipality types are not comparable across entities. The percent reporting no land or livestock in RS was 74.7% in "urban" municipalities, 43.4% in "mixed" municipalities and 31.2% in "rural" municipalities. Respective figures for FbiH were 88.7%, 60.4% and 40.0%.

Table 2: Selection of Municipalities

Stratum i	Total municipalities M_i	Sampled municipalities m_i
1. Federation, mostly urban	10	5
2. Federation, mostly mixed	26	4
3. Federation, mostly rural	48	5
4. RS, mostly urban	4	2
5. RS, mostly mixed	29	5
6. RS, mostly rural	29	4

Note: M_i is the total number of municipalities in stratum i (i=1, … , 6);

m_i is the number of municipalities selected from stratum i;

As the selection of the specific municipalities in the Master Sample was made PPES within strata, for each municipality, the probability of selection was:

$$P_j = m_i \times \frac{N_{ij}}{N_{i*}}$$

Where:

M_i is the total number of municipalities in stratum i (i=1, … , 6);
m_i is the number of municipalities selected from stratum i;
N_{ij} is the estimated number of households in municipality j in stratum i (j = 1, …, M_i);
N_{i*} is the estimated total number of households in stratum i.

These selection probabilities are shown in Table 3 for the selected municipalities.

Table 3: Probability of Selection for the Selected Municipalities

j	Municipality	Entity	Strata	Initial Estimates of the Number of Households N_{ij}	Total Number of Households in Stratum N_{i*}	Probability of Selection P_j
1	Banja Luka	RS	Urban	65420	82071	1.594[†]
2	Srpska Ilidza	RS	Urban	4888	82071	0.119
3	Cajnice	RS	Mixed	1487	182543	0.041
4	Modrica	RS	Mixed	8266	182543	0.226
5	Novi Grad	RS	Mixed	8961	182543	0.245
6	Prijedor	RS	Mixed	28339	182543	0.776
7	Visegrad	RS	Mixed	5581	182543	0.153
8	Knezevo	RS	Rural	3564	154170	0.092
9	Samac	RS	Rural	6746	154170	0.175
10	Srbac	RS	Rural	7215	154170	0.187
11	Zvornik	RS	Rural	14623	154170	0.379
12	Centar	FBIH	Urban	18870	202307	0.466
13	Nov Sarajevo	FBIH	Urban	19839	202307	0.490
14	Novi Grad	FBIH	Urban	31453	202307	0.777
15	Tuzla	FBIH	Urban	38537	202307	0.952
16	Zenica	FBIH	Urban	36447	202307	0.901
17	Breza	FBIH	Mixed	3900	146688	0.106
18	Travnik	FBIH	Mixed	14375	146688	0.392
19	Visoko	FBIH	Mixed	11312	146688	0.308
20	Vogosca	FBIH	Mixed	5371	146688	0.146
21	Gradacac	FBIH	Rural	13182	296691	0.222
22	Grude	FBIH	Rural	3823	296691	0.064
23	Kakanj	FBIH	Rural	12365	296691	0.208
24	Posusje	FBIH	Rural	4564	296691	0.077
25	Zavidavici	FBIH	Rural	10758	296691	0.181

[†] The expected number of times that Banja Luka would be sampled under this design is 1.59. In other words, it would be certain to be sampled at least once. There is a 0.41 probability that it would be selected once and a 0.59 probability that it would be selected twice. Normal practice might be to treat such units as a separate stratum with P=1.0. It is not clear what practice was adopted in this case. It is assumed here that Banja Luka was left on the list to be sampled PPES, and that if it were selected twice, this was ignored (and the number of EAs to select was calculated in the same way as if it had only been sampled once). This is equivalent to just giving a selection probability of 1.0, so this is what has been assumed in subsequent calculations.

3.1.3 Listing Operation

In each of the selected municipalities a full listing of households (”microcensus”) was carried out. The work was carried out in a decentralized approach, wherein the FIS and the RSIS were responsible for carrying out the fieldwork under the general guidance of the BHAS. The municipalities cooperated by providing temporary office and storage space and recruitment of enumerators and controllers for the survey. The fieldwork was supervised by the staff of the two entity institutes, and these were trained in their respective institutes. This involved three phases:

Preparatory Phase: The tasks carried out during this phase included updating of maps with respect to street names, street numbers and buildings, defining the boundaries of the municipalities, and the enumeration areas within them. This was done by the geodesic institutes of the two entities. The next step was identifying enumerators, controllers and supervisors, training them and assigning them to specific areas. The other tasks during this phase were the printing of questionnaires and instructions, defining the codes to be used and informing the municipalities about their specific responsibilities. While the controllers were selected by municipalities, the supervisors were provided by the entity institutes.

Listing Phase: Enumerators were provided maps of their areas and the questionnaires and instruction manuals They collected information on the households in their assigned areas using a short questionnaire which gathered information on the identify of the head of household, address, and the number of members in the household by sex and age. If no one was home, the household was visited again to record the information. If, after three such visits, no one was home, the information was obtained from the neighbors. The controllers supervised the fieldwork, checked the filled-in schedules and completed a report form on the fieldwork. They also assisted the interviewers whenever there were difficulties. The supervisors of the entity institutes conducted spot checks and ensured completeness and accuracy of data collection and the transfer of all the filled schedules to the entity institutes..

Data-entry Phase The data entry was performed at the entity institutes using a custom data entry system based on ACCESS software. Forty data entry operators (18 in the RS and 22 in the Federation) were selected and trained by the institute staff. The data entry was performed in two shifts and was supervised by two programmers of the entity institutes. The data were checked for logic and coding errors and tabulated to provide the essential information such as number of enumeration areas covered, number of households covered, number of members in the households by sex, number of refusals, number of households whose members were absent even after three visits etc. These tabulations were made by municipality and enumeration areas and formed the basis for the second stage sampling.

3.2 LSMS Sample[5]

3.2.1 Selection of EAs

The municipalities are divided into geographic areas called enumeration areas (EAs). In theory, each enumeration area consists of the number of households that can be interviewed in a census by an enumerator in one day. The EAs in BiH are based on the 1991 Census. But, at the time the Master Sample listing operation was carried out, many of the enumeration areas actually contained many fewer households (in some cases, zero). As enumeration areas were to be the primary sampling unit for the LSMS survey, the first step was to combine contiguous EAs until a new enumeration area with a minimum of 50 households was formed. These newly constructed EAs were called groups of enumeration areas (GNDs) and replaced the original small EAs. Thus the primary sampling units (PSUs) were actually a mix of the original EAs of sufficient size and the new constructed GNDs. For simplicity, the remaining discussion will use the term EA to refer to both.

Based on the population figures from the Master Sample microcensus, 250 EAs were selected with PPS from the municipalities in the FBiH and, and 200 EAs were selected with PPS in the municipalities of the RS.

In the FBiH, the number of EAs to select in municipality j was calculated as follows:

$$e_{ij} = 250 \times \frac{NA_{ij}}{\sum_{j=1}^{14} NA_{ij}}.$$

Where, NA_{ij} is the enumerated number of households in municipality j in stratum i (not to be confused with N_{ij}, the prior estimate of the number of households in the municipality)

Within each municipality, the e_{ij} EAs are selected PPS, so the probability of selecting EA k in municipality j (conditional upon having selected municipality j) is:

$$P_{k|j} = e_{ij} \frac{NA_{ijk}}{NA_{ij*}} = \frac{250 \times NA_{ijk}}{\sum_{j=1}^{14} NA_{ij*}}.$$

Where NA_{ijk} is the enumerated number of households in EA k in municipality j in stratum i

Similarly, in the RS, the probability of selecting EA k in municipality j (conditional upon having selected municipality j) is:

[5] This section is based largely on a Peter Lynn note "LSMS Sample Design and Weighting—Summary" April 2002. Essex University, commissioned by DfID.

$$P_{k|j} = \frac{200 \times NA_{ijk}}{\sum_{j=15}^{25} NA_{ij*}}$$

Note that:

$\sum_{j=1}^{14} NA_{ij*} = 224{,}796$ and $\sum_{j=15}^{25} NA_{ij*} = 155{,}090$,

so for the Federation: $P_{k|j} = \frac{250 \times NA_{ijk}}{224{,}796}$

and for the RS $P_{k|j} = \frac{200 \times NA_{ijk}}{155{,}090}$

3.2.2 Selection of Households

Within each of the 450 selected EAs, 12 households were selected systematically. Thus the probability of selecting household l in EA k in municipality j (conditional upon having selected EA k in municipality j is:

$$P_{l|jk} = {}^{12}\!/\!_{NA_{ijk}}$$

where:

NA_{ijk} is the enumerated number of households in EA k in municipality j in stratum i.

3.2.3 Overall Selection Probabilities

The overall probability of selection for household l in EA k in municipality j in stratum i is the product of the three conditional probabilities:

$P_l = P_j \times P_{k|j} \times P_{l|jk}$, i.e.

in FBiH : $P_l = \frac{m_i \times N_{ij}}{N_{i*}} \times \frac{3000}{224{,}796}$

In the RS: $P_l = \frac{m_i \times N_{ij}}{N_{i*}} \times \frac{2400}{155{,}090}$

The probability therefore has two components. The first component reflects differential probabilities between municipalities. These arise because different sampling fractions were used in each of the three strata within each entity (reflected in the term ${}^{m_i}\!/\!_{N_{i*}}$) and because municipalities were selected within strata PPS (reflected in the term N_{ij}) – an imbalance that was not corrected at the subsequent stage. The second component reflects the (small) difference between entities in the conditional selection probabilities of households.

Thus, we can write:

$P_l = P_j \times P_{l|j}$, where $P_{l|j} = 12 \times K_h$.

(where K_1 = 250/224,796 for FbiH and K_2=200/155,090 for RS.)

These probabilities are shown in table 4 below for each municipality. It can be seen that there is a very large range of household selection probabilities, from around 0.0006 in Cajnice to 0.0155 in Banja Luka (so, households in Banja Luka had 26 times the chance of being selected of households in Cajnice).

Table 4: Overall Selection Probabilities of Households

Municipality		Probability for municipality (from **Error! Reference source not found.**)	Probability for EAs	Probability for households	Overall probability
j		P_j	$P_{k\|j}$	$P_{l\|jk}$	P_l
1	Novi Grad	0.77736	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.01037
2	Centar	0.46637	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00622
3	Novo Sarajevo	0.49032	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00654
4	Zenica	0.90078	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.01202
5	Tuzla	0.95244	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.01271
6	Vogošća	0.14646	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00195
7	Travnik	0.39199	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00523
8	Visoko	0.30846	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00412
9	Breza	0.10635	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00142
10	Zavidovići	0.18130	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00242
11	Gradačac	0.22215	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00296
12	Posušje	0.07692	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00103
13	Kakanj	0.20838	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00278
14	Grude	0.06443	$K_1 NA_{ijk}$	$12 / NA_{ijk}$	0.00086
15	Srpska Ilidža	0.11912	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00184
16	Banja Luka	1.00000	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.01547
17	Čajniče	0.04073	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00063
18	Novi Grad	0.24545	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00380
19	Prijedor	0.77623	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.01201
20	Modriča	0.22641	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00350
21	Višegrad	0.15287	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00237
22	Kneževo	0.09247	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00143
23	Šamac	0.17503	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00271
24	Zvornik	0.37940	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00587
25	Srbac	0.18720	$K_2 NA_{ijk}$	$12 / NA_{ijk}$	0.00290

Note: $K_1 = {}^{250}\!/_{224,796}$ and $K_2 = {}^{200}\!/_{155,090}$.

3.3 Weights

To produce unbiased estimates for LSMS, each sample household should be weighted by the inverse of its selection probability, viz:

$W_l = \frac{1}{P_l}$,

In the Federation: $W_l = \frac{224{,}796 N_{i*}}{3000 \times m_i \times N_{ij}}$ and;

in the RS: $W_l = \frac{155{,}090 N_{i*}}{2400 \times m_i \times N_{ij}}$.

These weights are shown in Table 5 along with the impact they have on the sample distribution across municipalities.

Table 5: Weights and Impact on Sample Distribution by Municipalities

Municipality j		Weight for each household w_j	Sample households n_j	Sample proportion $n_j / \sum_{j=1}^{25} n_j$	Weighted sample households $w_j n_j$	Weighted sample proportion $w_j n_j / \sum_{j=1}^{25} w_j n_j$
1	Banja Luka	64.621	936	0.173	60485	0.055
2	Srpska Ilidza	542.502	84	0.016	45570	0.041
3	Cajnice	1586.561	36	0.007	57116	0.051
4	Mordica	285.412	132	0.024	37674	0.034
5	Novi Grad	263.276	156	0.029	41071	0.037
6	Prijedor	83.250	432	0.080	35964	0.032
7	Visegrad	422.723	84	0.016	35509	0.032
8	Knezevo	698.835	60	0.011	41930	0.038
9	Samac	369.204	108	0.020	39874	0.036
10	Srbac	345.204	120	0.022	41425	0.037
11	Zvornik	170.324	252	0.047	42922	0.039
12	Centar	160.671	276	0.051	44345	0.040
13	Nov Sarajevo	152.823	288	0.053	44013	0.040
14	Novi Grad	96.393	432	0.080	41642	0.038
15	Tuzla	78.674	528	0.098	41540	0.037
16	Zenica	83.185	468	0.087	38931	0.035
17	Breza	704.591	60	0.011	42275	0.038
18	Travnik	191.159	192	0.036	36702	0.033
19	Visoko	242.920	144	0.027	34980	0.032
20	Vogosca	511.619	84	0.016	42976	0.039
21	Gradacac	337.303	144	0.027	48572	0.044
22	Grude	1163.047	48	0.009	55826	0.050
23	Kakanj	359.590	144	0.027	51781	0.047
24	Posusje	974.218	48	0.009	46762	0.042
25	Zavidavici	413.305	144	0.027	59516	0.054

The impact on the distribution across strata is shown in Table 6. It can be seen that the weighted sample distribution across the six strata is much closer to the population distribution than the unweighted sample distribution.

In fact, using the results of the master sample microcensus we can obtain better estimates of the population stratum sizes using a ratio estimation approach. An unbiased estimate of the actual stratum size can be obtained as follows:

$$\hat{N}A_{i*} = \frac{\sum_{j=1}^{m_i} \frac{NA_{ij}}{P_j}}{\sum_{j=1}^{m_i} \frac{N_{ij}}{P_j}} N_{i*}$$

These revised population estimates were then compared with the design-based sample estimates and a post-stratification weight to correct the remaining imbalance was also applied. This weight, w_i^{PS}, was calculated as follows:

$$w_i^{PS} = \frac{\hat{N}A_{i*}}{\sum_{j=1}^{m_i} w_{ij} n_{ij}};$$

the overall weight to be used with LSMS survey data is $W^*_{ij} = W_{ij} \times w_i^{PS}$.

Table 6: Impact of Weights on Sample Distribution by Strata

Stratum i	Estimated Population households *Ni*	Population proportion $N_{i*} \Big/ \sum_{i=1}^{6} N_{i*}$	Sample households n_{i*}	Sample proportion $n_{i*} \Big/ \sum_{i=1}^{6} n_{i*}$	Weighted sample households $\sum_{j=1}^{m_i} w_{ij} n_{ij}$	Weighted sample proportion $\sum_{j=1}^{m_i} w_{ij} n_{ij} \Big/ \sum_{i=1}^{6} \sum_{j=1}^{m_j} w_{ij} n_{ij}$
1 FBiH: Urban	202307	.189	1992	0.369	210471	0.190
2 FBiH: Mixed	146688	.137	480	0.089	156935	0.141
3 FBiH: Rural	296691	.277	528	0.098	262458	0.237
4 RS: Urban	82071	.077	1020	0.189	106056	0.096
5 RS: Mixed	182543	.171	840	0.156	207335	0.187
6 RS: Rural	154170	.149	540	0.100	166151	0.150

It should be noted that the numerator of the post-stratification weights is calculated in a way that takes into account the "actual" (microcensus) values for household counts, rather than just the prior estimates. As these are only known for sampled municipalities, the actual count for strata is estimated by the ratio estimator above. The numeric values of the weight are presented in Table 7. Note that it is very important that the overall weight has been calculated as the product of the design weight and the post-stratification weight, *not* just as the ratio of population size to sample size within strata. Though this latter approach too would give the correct distribution across strata, it would *not* give the correct distribution within strata and would result in a sample that is still biased towards larger municipalities (within strata).

An important point about the LSMS weights is that they have considerable variability, as can be already seen in the column w_j in Table 5. This will tend to increase the variance (standard errors) of survey estimates. This is the price to be paid for removing bias. Estimates of the design effect due to weighting (for a few key estimates) produced an increase of the standard error by 4.5-5.5 times (which effectively means that the precision of some estimates obtained is equivalent to a true random sample of just 1000 households).

Table 7: Post-Stratification Weights

Stratum i	Actual (Listed) Households	Post-stratification weight w_i^{PS}	Post-stratified sample proportion
1 Fed: Urban	213,802	1.024	0.206
2 Fed: Mixed	159,518	0.991	0.154
3 Fed: Rural	272,010	1.004	0.262
4 RS: Urban	72,261	1.006	0.070
5 RS: Mixed	177,387	0.989	0.171
6 RS: Rural	143,351	0.979	0.138

4. Pilot Survey

A draft questionnaire was prepared comprised of the following 11 modules: Roster, Housing, Education, Health, Labor, Credit, Voucher, Migration, Consumption, Non-Agricultural activities and Agricultural activities. This was piloted (tested) during the period June 25-July 20, 2001 in the two entities. For the Pilot survey 9 interviewers were selected in each entity and were trained in the concepts and methodology of the survey. Each interviewer was required to interview 12 households in specified areas. Both the areas and the field staff were selected by the entity institutes. Training for the Pilot Survey was carried out from June 18 to 22, 2001 by the Survey Management Team with participation of experts from UNDP, World Bank and DfID. The training covered the concepts and approaches used in the survey modules, question and answer sessions and practice sessions. Two data entry operators from each entity institute were also trained in the use of a specialized data entry software: CS-Pro.

The actual pilot survey was carried out over a four week period. In the first week, the interviewers visited their 12 households and administered the first 9 modules of the questionnaire: essentially the basic household data and the individual data sections. The Survey Management Team served as supervisors for the Pilot survey.

A workshop was then held in Laktasi (July 3-6, 2001) with all interviewers and members of the survey management team to review the experience and discuss any issues that had arisen. In parallel, data collected from the first week of interviews was entered into the data entry program so that this was also tested during the pilot survey.

During the third week the interviewers returned to their 12 selected households and finalized the interview by completing modules 10-13. The final week was used for a second workshop (in Zenica, July 17-20, 2001) to discuss the final modules and field experiences.[6] Again, data from the third week interviews were entered and the resulting problems identified with the data entry phase discussed.

The main conclusions from the two workshops are summarized below:

1. The questionnaire, particularly modules on health and labor were considered to be too long as it took, on average, 2 hours to complete the questionnaire (interviewing all household members) which was considered to be too long.
2. The Roster sheet that folds out during the interview needed to be made of thicker paper since it was frequently opened and closed and had a tendency to tear.
3. Changes were needed in the wording of some questions (particularly about housing, health and migration).
4. Concerns were raised about the housing module and the accuracy of responses. Unlike other countries, the housing module in BiH is one of the most sensitive. Housing tenure for many people is extremely uncertain (due to the war) many persons were facing eviction if the owner of the dwelling where they live returned.
5. There was a need to do more publicity about the survey prior to beginning the actual field work.
6. Non-response rates overall were low but were high in specific areas where war-related activities had had the hardest impact.
7. There was a discussion of 'rewarding households' as some respondents had asked what the interviewer would give them for answering the interview. Various suggestions about providing some small gift, like chocolates or gum be given to the respondents at the end of the interview. This issue was debated extensively as there was concern that paying households might bias results and/or create precedents that could not be followed in future surveys.
8. Some interviewers expressed concerns about the accuracy of responses to personal questions on credit, ownership of housing, and durable goods.
9. The need for inclusion of an additional module on social assistance was brought out if the survey was to capture actual welfare.
10. There was concern that households only answered questions on household business activities if these businesses were legally registered thus omitting the informal sector or gray economy.
11. The five days of training for the pilot was considered to be inadequate.

The household questionnaires were revised incorporating the suggestions received in the Laktasi and Zenica workshops. Two additional modules were added- the End of First Round Module and Social Assistance Module. The End of First Round Module was intended to identify households where the agricultural and non-agricultural business modules needed to be administered in the second visit to the household. This section was designed to minimize

[6] Reports of the workshops are available from the statistical institutes.

any loss of information on household businesses and agricultural activities. The Social Assistance module was included to obtain information on the various social welfare benefits received by individuals such as old age pension, family pension, disability pensions, etc. The health and labor modules were cut back substantially. The credit module, given the concern about responses was also cut back. The non-agricultural enterprise module was also reduced substantially. The refusal to provide information on non-registered businesses lowered the value of the module. It was felt that a reduced, less invasive module could elicit better responses, although it would not provide the detailed data required to analyze the sector.

The concern about the need for more training was taken into account and a three-week training course for the survey was developed. Finally, it was decided not to pay households for participating in the survey.

5. Fieldwork

5.1 Organization of Data Collection

The field work for the LSMS survey was carried out in the following manner. Mobile teams of interviewers were formed with three interviewers each plus one supervisor. A data entry operation with a computer was assigned to each pair of teams. The team was provided with a car and driver to ensure that time was not wasted in transportation.

Each interviewer was assigned, per month, two clusters of households. (Each cluster was 12 households in an enumeration area or group of enumeration areas). In week 1, the interviewer carried out the first half of the interview (modules 1-10) with the 12 households in Cluster A. In week 2, the interviewer carried out the first half of the interview with the 12 households in Cluster B. While the interviewer was working in Cluster B, all of the questionnaires from Cluster A were entered electronically by the data entry operator and lists of errors, inconsistencies and missing data were produced. In the third week, the interviewer returned to Cluster A to finish the interview (modules 11-13) with the 12 households and clarify with the households any problems found from the first visit and fill in any missing information.

While the interviewer was in Cluster A for the second time, the data from Cluster B were entered, and lists of errors created. In week four, the interviewer returned to Cluster B to finalize the interview and to make any necessary corrections.

Often, the interviewers visited each household more than two times. All information was collected from direct informants, except in the case of children under 15 whose parents were asked to provide the information. Otherwise, the interviewer carried out a series of interviews in the household, one for each member. In order to find and interview each member of the household, it was often necessary to return to the household multiple times. For this reason, the work load of 12 households in a two-week period was considered sufficient.

5.2 Recruitment and Training of Field Staff

Interviewers and supervisors were recruited through the entity Employment Bureaus. The responsibility for recruiting the field staff was vested with the entity institutes. The Institutes

contacted the Employment Bureaus and obtained lists of unemployed persons who were on their roster and selected those with at least high school certificates and some prior work experience.

The Survey Management Team was responsible for conducting the training for the field staff. Four training sessions running parallel were held in Zenica (FBiH) and Teslic (RS). Each training session had a mixture of interviewers from both entities to ensure that the implementation of the survey did not vary between entities. In each training session, the trainers also represented a mixture of staff from the three statistical organizations. Details on the training outline can be found in Box 1.

Box 1: Training for Interviewers

The organization of the training included the following elements:

1. Introduction to LSMS and general survey procedures;
2. Explanation of the questionnaire structure and contents and concepts and definitions;
3. Description of each Module 1-10 (round 1) followed by at least two interviews by a pair of interviewers
4. Discussion of the experiences of completing Modules 1-10;
5. Discussion of data entry programme reports;
6. Discussion of Modules 11-13 (round 2) followed by at least two interviews by a pair of interviewers.
7. Discussion of experiences of completing Modules 11-13;
8. Discussion of control procedures, map reading etc;
9. Test of the interviewers to assess their knowledge
10. The supervisors of the Pilot survey from the two entities were asked to speak about their experience during the pilot survey.

5.2.1 Training Course

Each training course was three weeks long and had a practical orientation. The morning sessions were usually devoted to discussing the individual modules, and in the afternoons the interviewers and supervisors completed the different modules by interviewing each other-one playing the role of interviewer and the other playing the role of respondent by turn. The completed questionnaires were then discussed and mistakes were pointed out and corrected. These completed questionnaires were later used for training data entry operators. In the Zenica courses, the interviewers also carried out 1-2 actual interviews with households. In Teslic, this was not feasible: instead interviewers carried out a full interview on another member of the training session. Two days of training were devoted to learning about the control procedures--four control forms were provided to monitor the flow of questionnaires from the time when they are given to the interviewers until they are received finally after data entry—as well as map reading and other administrative and control details. Following the training a test was conducted to determine each person's level of knowledge of the questionnaire and instructions. The candidates who performed best were selected as supervisors. Note that most of the supervisors were those people who had been interviewers during the pilot test.

5.3 Fieldwork

Each survey team was comprised of 3 interviewers, one supervisor and one driver with a car. Since the interviewers were recruited from the same municipalities where they were to work, they knew the area well. In addition they were provided with maps of the area assigned to them. The supervisors provided logistic support, and helped solve difficulties. The fieldwork started on 26 September 2001 and ended on 23 November 2001. The timing of the fieldwork was limited by the need to finalize all interviews before the start of Ramadan since household consumption patterns were expected to change during the fasting month. On average, interviewers took 1.5 hours per household to collect the data. Only in the case of households with over 5 members did the interview take longer. The interval between the two rounds benefited the survey in the following ways. First, it shortened the time spent in a household for a given visit, thus reducing the risk of respondent fatigue And, second, this structure allowed sufficient time for entering the data and listing the errors for field verification

One data entry operator was provided for every two teams. The data were entered soon after the questionnaire was completed, and the customized data entry programme was used to produce a list of errors (missing data, inconsistencies and the like) in the data. This enabled the interviewers and supervisors to review each questionnaire, resolve any small difficulties and/or decide that the questionnaire needed to be sent back to the household for clarification.

The interviewing was conducted at the convenience of the respondents which meant interviews were conducted both during the day and during the evenings and throughout the week, including weekends. The supervisors were responsible for planning each day's work for their teams. He or she also planned the activity of the driver to ensure that the questionnaires are collected each day and delivered to the data entry operators, and, once entered and an error list produced, returned to the interviewers for correction. In many municipalities, temporary office accommodation was provided where the interviewers could meet and store the questionnaires. Where such accommodation was not available, the car served as a temporary office and in some cases the supervisor's home served as office. Finally, a member of the Survey Management Team visited the different municipalities and conducted spot checks of the fieldwork throughout the interviewing period.

Each entity provided badges and letters of introduction to the interviewers and supervisors. Communication with field staff was improved by recruiting interviewers and drivers who had cell-phones

Overall, the response rate in the survey was 82 percent. For each enumeration area, four replacement households were selected prior to the field work. Using these replacement households as needed (a total of 938 households), the final sample size was 5,402 households interviewed.

5.4 Data Entry

An integrated approach to data entry and fieldwork was adopted in Bosnia and Herzegovina. Data entry proceeded side by side with data gathering to ensure verification and correction in the field. Data entry stations were located in the regional offices of the entity institutes and

were equipped with computers, modem and a dedicated telephone line. The completed questionnaires were delivered to these stations each day for data entry.

Twenty data entry operators (10 from Federation and 10 from RS) were trained in two training sessions held for a week each in Sarajevo and Banja Luka. The trainers were the staff of the two entity institutes who had undergone training in the CSPro software earlier and had participated in the workshops of the Pilot survey. Prior to the training, laptop computers were provided to the entity institutes, and the CSPro software was installed in them. The training for the data entry operators covered the following elements:

- Introduction to the LSMS Survey questionnaire; Introduction to the personal computers/ lap top computers; Copying data on diskette and printing of output;
- The Data entry programme (CSPro). Understanding of the Round 1 data entry screens (Modules 1-10);
- Practice of Round 1 (data entry trainees enter questionnaires completed by interviewer trainees during practice interviews);
- Understanding of Round 2 Data entry screen (Modules 11-13)
- Practice of Round 2 Data entry screens (data entry trainees entered the questionnaires completed by interviewer trainees)
- Control Procedures; Copying data on diskette and printing lists of errors; Transfer of the data through email to the institutes.

The data entry programme was fine-tuned during the training. Some unexpected responses during the interviews had to be accommodated and a few skip patterns fixed. The training emphasized the role of the data entry operator as a member of the survey team, and how the outputs of the programme (error lists) were to be provided to the supervisors and interviewers for necessary correction.

The goal was to produce high quality data. Several of the key features of this were:

1. Pre-coded verbatim questionnaires ;
2. Error detection at the time of data entry;
3. Data entry that was concurrent with fieldwork;
4. Correction of suspected errors in the field.

The following checks were incorporated in the data entry software:

1. Value Range: The program checked to ensure that the values entered were within the valid range for each variable;
2. Reference tables: Where appropriate, the entered data were checked against reference values (e.g. the price of a kilo of tomato could not exceed 10 KM[7]);
3. Skip checks: The program checked that all appropriate skips were followed, both within and between different units of observation;
4. Checks for consistency between different responses: The program checked for internal consistency. For example, whether the age of a person was sufficient for the

[7] The Convertible Mark (KM) was equivalent to the Deutsch Mark at the time of the survey, or approximately US$1.9.

education level attained, if a filter question for agriculture had a positive response that the module had all relevant information entered and the like;

After the data entry was completed in the field, the data were transferred through email to the central offices in Sarajevo and Banja Luka with the help of PCAnywhere Software.

The data entry programme was designed to detect many of the errors even at the stage of data entry, thereby minimizing the need for ex-post facto data editing. Once all data was compiled in the entity offices, a check was made to ensure the structural consistency of data files, i.e. that no records were duplicated or omitted.

When the RS and FBiH data files were merged it became apparent that a last minute decision on the treatment of decimal places in several modules had been different in the two entities. Thus the two data bases were not compatible. A correction was made and data from these modules were re-entered. Once this was done, the data sets were compatible and a country-wide data set was created. During this process some additional double entry was carried out to correct any data entry operator errors that had occurred.

6. The Data Set

6.1 Data Cleaning

It is important to note what is meant by 'data cleaning' in terms of the BiH-LSMS data set. In the sense that the data set is a faithful reflection of the responses of all interviewees the data set can be considered 'cleaned'. Every effort was made to ensure that the information provided during the interviews was correctly entered in electronic format. As in any survey, this does not mean that the data set is perfect. As participation in the survey is voluntary, informants had the option to refuse to answer specific questions, and may have provided information that is not always consistent. The interviewers resolved as many inconsistencies as possible with the informants but there are, of course, limits.

However, given the widely differing needs of the range of analysts who will use the BiH-LSMS data, nothing further has been done to the original data. While some data sets are processed so that all missing values are imputed, all outliers revalued and all inconsistencies fixed based on some set of assumptions, this has not been done here. The reason being that there is no correct way to resolve the problems of missing data, outliers and inconsistencies. Each person will need to make his or her own decision on how to treat such data problems based on the type of analysis being carried out. For some analyses, the information in outlier values is key while for others, such outliers would distort findings and would need to be dropped or provided an imputed value. The same for missing values. Some analysts will chose to drop cases with missing values for the variables of interest to them while others will impute such values, using medians, mean or complex multi-variate techniques. In order to ensure the usefulness of the data set for all users, no attempt has been made to impute missing values, reconcile inconsistencies, re-value outliers, or in any way alter the responses provided by the respondents.

6.2 Basic Data Files

The data from the data entry program was converted into the formats acceptable for use with SPSS, SAS and STATA. The Description of the data files (regardless of software used) is given in Table 8.

Table 8: Description of Data Files

Name of File	Topic	Unit of Observation	RS No. of Cases	FBiH No. of Cases	BiH No. of Cases
M1.*	Roster, Demographics	Individual	7,751	9,225	16,976
M2_AB.*	Housing	Household	2,400	3,002	5,402
M2_C.*	Durable Goods	Household: Items	18,567	24,359	42,926
M3.*	Education	Individual	7,862	9,265	17,127
M4_A.*	Health, Use of Services	Individual	7,862	9,265	17,127
M4_B.*	Health, Status	Individual	7,862	9,265	17,127
M5.*	Labor	Individual	7,862	9,265	17,127
M6.*	Credit	Individual	7,862	9,265	17,127
M7.*	Vouchers	Individual	7,862	9,265	17,127
M8.*	Migration	Individual	7,862	9,265	17,127
M9.*	Social Assistance	Individual	7,862	9,265	17,127
M10.*	Filter and Non-Agri. Businesses	Household			
M11_A1.*	Consumption: Daily	Household: Items	11,812	14,882	26,694
M11_A2.*	Consumption: Meals	Household: Items	9,440	11,904	21,344
M11_B1.*	Consumption: Food	Household: Items	103,838	130,944	234,782
M11_B2.*	Consumption: Seasonal Food	Household: Items	51,887	65,472	117,337
M11_C1.*	Consumption: Non-food, monthly	Household: Items	9,432	11,904	21,336
M11_C2.*	Consumption: Non-food, annual	Household: Items	113,184	142,848	255,984
M12_FIL.*	Non-agricultural business, first question	Household	2,400	2,983	5,383
M12.*	Non-agricultural business	Household Business	197	173	370
M13_FIL.*	Agriculture, first question from each section of Module	Household	2,400	3,002	5,402
M13_A1.*	Agriculture: Land Used	Household: Land	2,448	1,396	3,844
M13_A2.*	Agriculture: Land rented out	Household: Land	758	382	1,140
M13_B1.*	Agriculture: Forest Land	Household Land	409	159	568
M13_B2.*	Agriculture: Crop Production	Household Crops	52,780	40,832	93,612
M13_C1.*	Agriculture: Inputs, Seeds	Household Inputs	10,529	6,523	17,052
M13_C2,*	Agriculture: Inputs,	Household Inputs	6,632	4,824	11,456

	Fertilizers				
M13_C3.*	Agriculture: Inputs, Fuel	Household Inputs	868	404	1,272
M13_C4.*	Agriculture: Inputs, Labor	Household Inputs	1,683	540	2,223
M13_C5.*	Agriculture: Inputs, Machinery	Household Inputs	5,080	3,560	8,640
M13_D1.*	Agriculture: Livestock	Household Livestock	13,712	8,064	21,776
M13_D2.*	Agriculture: Animal feed	Household livestock feed	6,380	3,752	10,132
M13_E.*	Agriculture: Capital Assets	Household Assets	18,567	24,359	9,241

The data files contain the country-wide data. For an analyst interested only in one entity or the other, the Entity level data sets are easily constructed. This is done by selecting within each file, using the variable ENTITY, the entity wanted. If the analyst wishes to create data files containing only Republika Srpska data, select all cases where ENTITY equals 1. For the FBiH data, select if ENTITY equals 2.

6.2.1 Naming Conventions

The questionnaire is the basic guide to the data set. Each module in the questionnaire is numbered and the data sets reflect this structure. Thus, file M1.* contains the data from Module 1 of the questionnaire (roster and demographic information) and M12.* is Module 12 (non-agricultural household enterprises). When the module has more than one part and the resulting data set needed to be split, the parts (or sections) are also found in the file name. For example, Module 4 (health) is split into two parts, A and B. The file name reflects this: M4_A.* and M4_B.*. There are two exceptions to this rule. M12_FIL.* and M13_FIL.*. These two files contain the filter questions from modules 12 and 13 respectively. In module 12 this is the first question and was used to determine whether the module should be administered in the household. M13_FIL.* contains the filter questions from each of the 12 parts of the Agricultural Module that were used to determine if the entire part was to be administered.

Within each of the data files, the numbering of the questions is also based on the questionnaire. Thus, question number two in the roster module will have the name: m1_q02 (for Module 1, question number 2). In the case that a question has more than one part, the question name reflects this. For example, question 4 in the first module asks the day, month and year of birth. The question numbers respectively for these are, respectively: m1_q04a, m1_q04b, m1_q04c.

The exception to this rule is for modules 10,11, 12 and 13. For variable names within these modules, the modules were given a letter such that module 10 is signified by the letter ‘A’, module 11 by the letter ‘B,’ module 12 by the letter ‘C ‘and module 13 by the letter ‘D’. Thus, mcc1_q1a refers to module 12, part C1, question 1a and mdd1_q01 refers to question 1, section d1 of Module 13.

6.2.2 Merging Files

Merging files together requires that each unit of observation have a unique identifying code. Four variables are required to construct a unique household identity code. These are:

→ **MUNCODE** (municipality code, see Table 9 for the corresponding names for each code),
→ **GND** (groups of enumeration areas),
→ **NUMIST** (enumeration area number), and
→ **HID** (household identification number within the enumeration area).

To create a unique identification code for each individual in the individual level data sets, the variable PID must be added on to the previous four variables.

6.3 Constructed Variables: Weights and Welfare

One additional file is added here, called POVERTY1.*. This file contains the constructed welfare variables and poverty indicators. The variables are:

ADJYRCON→ total household yearly consumption, adjusted for cost of living
PCAYRCON→ per capita yearly consumption, adjusted for cost of living
EXTPLINE → value of the extreme or food poverty line
GENPLINE → value of the general poverty line (includes allowance for non-food consumption).
POOR → dichotomous variable taking on the value of 1 if the individual's per capita annual consumption (adjusted for the cost of living) is below the general poverty line.
WHHD → this is the variable which must be used to weight the household level data files to get accurate results from analysis
WPOP → this is the variable which must be used to weight the individual level data files to get accurate results from the analysis.

For information on how the welfare measures and poverty lines were constructed, please refer to "Welfare in Bosnia and Herzegovina, 2001: Measurement and Findings" done jointly by the three statistical organizations and the WB.

Table 9: Municipality Codes

Strata	Municipality Name	Municipality Code
Federation of BiH		
Urban	Novigrad	10871
	Centar	10839
	Novi Sarajevo	10880
	Zenica	11185
	Tuzla	11088
Mixed	Vogosca	10928
	Travnik	11061
	Visoko	11126
	Breza	10189
Rural	Zavidovici	11177
	Gradacac	10391
	Posusje	10731
	Kakanj	10448
	Grude	10405
Republika Srpska		
Urban	Sprska Illidza	10855
	Banja Luka	10022
Mixed	Cajnice	10235
	Novi Grad	10111
	Prijedor	10740
	Modrica	10642
	Vicegrad	11134
Rural	Knezevo	10936
	Samac	10138
	Zvornik	11193
	Srbac	10952

References Cited

The World Bank, A Manual for Planning and Implementation of the Living Standards Measurement Study by Margaret Grosh and Juan Munoz, LSMS, Working Paper No 126, Washington D.C, 1996.

The World Bank, Designing Household Survey Questionnaires for Developing Countries : Lessons from 15 years of LSMS by Margaret Grosh and Paul Glewwe, 3 Volumes, Washington D.C. 2000.

Statistics Sweden, Advice on Master Sample for Bosnia and Herzegovina (BiH), Part II by Jette Bodin and Claes Cassel, June 2000

IBF, Advice on Master Sample for the two entities of Bosnia-Herzegovina, Report from IBH Mission to BIH, 23 March, 2001.

Peter Lynn, "LSMS-BiH Sample Design and Weighting", Personal Communication, March 2002.

LSMS Project, BiH, Report of the Workshops held in Laktasi and Zenica, July 2001 prepared by K.E.Vaidyanathan, C.T.A

UNDP, World Bank and DFID. "Building a Basis for a Statistical System in BiH, Project Document , Dec 2000.

Vaidyanathan, K.E. "Notes on the Selection on Sample Households"

BHAS, RSIS, FIS, WB, "Welfare in Bosnia and Herzegovina, 2001: Measurement and Findings", October 2002.

Appendix 1
How to Obtain Copies of Documentation and Data

The data and documentation of LSMS-BiH can be downloaded from the websites of the State Agency for Statistics (BHAS), the FBiH Web Site and the RSIS website (shortly). In addition the data and documentation can be downloaded from the World Bank website. The website addresses are as follows:

State Agency for Statistics of BiH: http://www.bhas.ba

Federation Statistical Institute: http://www.fzs.ba

Republika Srpska Statistical Institute: http://www.rzs.rs.ba

World Bank: http://www.worldbank.org/lsms/lsmshome.html

The documentation and data sets can also be obtained on CD rom if requested from the following institutions:

At the State level:
Agency for Statistics of Bosnia and Herzegovina
TRG Bosne i Hercegovine1
71000, Sarajevo,
Bosnia-Herzegovina
Email: bhas@bih.net.ba

For the Federation of BiH:
Federal Office of Statistics
Zelenih Beretki 26
71000, Sarajevo
Bosnia-Herzegovina
Email: bhstat@bih.net.ba

For Republika Serpska:
The Republika Srpska Institute for Statistics
Veljka Mlađenovića bb
Banja Luka,
78000 Banja Luka
Bosnia-Herzegovina
Email:rs_stat@inecco.net

Data can also be obtained from the World Bank by addressing an email to:
LSMS@worldbank.org

Individuals who receive the datasets should agree to (i) Cite the three statistical organizations of BiH as the authors of the datasets in their publications; (ii) provide copies of their publications to the institutions from where they obtained their datasets; (iii) not to pass on their datasets to other persons for any reason.

It is recommended that the user of the datasets read the documentation before using the data for their analysis.

Appendix 2
Documents Available for LSMS in BiH

1. Questionnaires of the BiH-LSMS in Local Language[8] (Latin and Cyrillic alphabets) and English. (separately)

2. Manual of Interviewer Instructions for BiH-LSMS in Local Language (Latin and Cyrillic alphabets) and English (separately)

3. Notes on the Selection of Sample Households for BiH-LSMS in English and Local Language (Latin alphabet)

4. Abstracts of LSMS-BiH: in Local Language (Latin and Cyrillic alphabets) and English

5. "Welfare in Bosnia and Herzegovina, 2001: Measurement and Findings". Methodology of calculation of consumption aggregates, and the results obtained in Local Language (Latin and Cyrillic alphabets) and English

6. Syntax Files used to construct Welfare Measures

7. Reports of the Workshops following the Pilot Survey held at Laktasi and Zenica, July 2001 in English and Local Language (Latin alphabet)

[8] Local Language is the used as an umbrella term for Bosnia, Croatian and Serbian in BiH.

Appendix 3
Codes Not Included in the Questionnaires

4. 1: CODES FOR OCCUPATIONS

ROD 1: LEGISLATIVE OFFICIALS, GOVERNMENT ADMINISTRATORS AND GOVERNMENT EXECUTIVES; MANAGERS

11 Legislative officials, Government Administrators, Government executives other than government administrators
12 General Managers of big companies
13 General Managers of small companies

ROD 2: SCIENTISTS AND RESEARCHERS

21 Physical scientists, Chemist, Physical science technicians
22 Medical doctors, Biologists, and Life sciences technicians
23 Teachers
24 Other scientists and researches (social scientists and related workers)

ROD 3: TECHNICAL AND OTHER PROFESSIONAL OCCUPATIONS

31 Technical and related workers
32 Technical and assistants in medical, veterinary, biology, biotechnology and related workers
33 Supervisors, Teachers and related workers
34 Commercial and administrative workers

ROD 4: CLERICAL WORKERS AND CASH DESK CASHIERS

41 Clerical workers
42 Cash desk cashiers

ROD 5: SERVICE AND SALES WORKERS

51 Service workers
52 Salesman and demonstrators

ROD 6: AGRICULTURE, ANIMALHUSBANDRY AND FORESTRY WORKERS, FISHERMEN AND HUNTERS

61 Agriculture, animal husbandry and forestry workers, fishermen

ROD 7: NON-INDUSTRIAL OCCUPATION

71 Mining and civil engineering occupation
72 Machinery Fitters, Machine Assemblers Electrical Fitters, and Mechanics
73 Precision-Instrument Makers, Glass Formers, Printers
74 Non-industrial food processors, wood preparation workers, textile and leather makers

ROD 8: OPERATORS OF MACHINES AND VEHICLES

81 Machine operators
82 Operators of equipment in processing industry and products assemblers
83 Drivers and operators of motor vehicles, movable machines and ship crew

ROD 9: SIMPLE OCCUPATIUONS

91 Selling and services workers
92 Agriculture, forestry and fishmen workers
93 Mining, civil engineering, processing and transport workers
94 Simple occupation not else classified

ROD 0. MILITARY OCCUPATIONS

01. Military occupation

4.2. Industrial Classification (NACE rev.2)

A. Agriculture, hunting and forestry
01 Agriculture, hunting and relates service activities
02 Forestry, Logging and Related Service Activities
B. Fishing
05 Fishing, Fish hatcheries and service activities related to fishing
C. Mining and Quarrying
10 Mining of Coal and Lignite; extraction of peat
11 Extraction of crude petroleum and natural gas
12 Mining of uranium and thorium ores
13 Mining of metal ores
14 Other mining and quarrying
D. Manufacturing
15 Manufacture of food products and beverages
16 Manufacture of tobacco products.
17 Manufacture of Textiles
18 Manufacture of wearing apparel, dressing and dyeing of fur
19 Tanning and dressing of leather; manufacture of luggage, handbags, footwear etc.

20 Manufacture of wood and wood products;
21 Manufacture of pulp, paper and paper products
22 Publishing, Printing and reproduction of recorded media
23 Manufacture of coke, refined petroleum products and nuclear fuel
24 Manufacture of chemicals and chemical products
25 Manufacture of rubber and plastic products

26 Manufacture of other non-metallic mineral products
27 Manufacture of basic metals
28 Manufacture of fabricated metal products, except machinery and equipment
29 Manufacture of machinery and equipment n.e.c

30 Manufacture of office machinery and computers
31 Manufacture of electrical machinery and apparatus n.e.c
32 Manufacture of radio, television and communication equipment and apparatus
33 Manufacture of medical, precision and optical instruments, watches and clocks
34 Manufacture of motor vehicles, trailers and semi-trailers
35 Manufacture of other transport equipment
36 Manufacture of furniture, manufacturing n.e.c
37 Recycling

E. Electricity, gas and water supply

40 Electricity, gas, steam and water supply
41 Collection, purification and distribution of water

F. Construction.

45 Construction

G. Wholesale and retail trade, repair of motor vehicles, motor cycles and personal and household goods

50 Sale, maintenance and repair of motor vehicles and motor cycles, retail sale of automotive fuel
51 Wholesale trade and commission trade, except of motor vehicles and motor cycles
52 Retail trade, except of motor vehicles and motorcycles, repair of personal and household goods

H. Hotels and Restaurants

55 Hotels and restaurants

I. Transport, storage and communications

60 Land transport, transport via pipelines
61 Water transport
62 Air transport
63 Supporting and auxiliary transport activities, activities of travel agencies
64 Post and telecommunications

J. Financial Intermediation

65 Financial intermediation, except insurance and pension funding
66 Insurance and pension funding, except compulsory social security
67 Activities auxiliary to financial intermediation

K. Real estate, renting and business activities

70 Real estate activities
71 Renting of machinery and equipment without operator and of personal and household goods
72 Computer and related activities
73 Research and development
74 Other business activities

L. Public Administration and Defense; Compulsory Social security

Public administration and defense; compulsory social security

M. Education

80 Education

N. Health and Social Work

85 Health and social work

O. Other Community, Social and Personal Services

90 Sewage and refuse disposal, sanitation, and similar activities

91 Activities of membership organizations n.e.c

92 Recreational, cultural and sporting activities

P. Private Households with Employed Persons

95 Private households with employed persons

Q. Extra-territorial organizations and bodies

99 Extra-territorial organizations and bodies