

The views expressed in this paper are the views of the authors and do not necessarily reflect the views or policies of the Asian Development Bank (ADB), or its Board of Directors or the governments they represent. ADB makes no representation concerning and does not guarantee the source, originality, accuracy, completeness or reliability of any statement, information, data, finding, interpretation, advice, opinion, or view presented.

Enhancing Social and Gender Statistics Workshop

Bangkok, Thailand

May 28-31, 2002

social statistics in vietnam

Country Paper

presented by Nguyen Van Tien

Deputy Director-General

General Statistics Office

Vietnam

A. Organization of Statistical Services

General Statistics Office (GSO) of Viet nam is large, long established (since 1946) agency in the Prime Minister Office. In 1956 GSO became independent Agency of the Government. The GSO is centralized Agency consisting of Departments and Bureaus in the Headquarters and local Offices in each province and city. General Statistics Office of Vietnam is independent Agency of the Government. Director-General of the GSO is appointed by the Prime Minister. Director of GSO's departments and local provincial Office are appointed by the Director-General. But The Head of Statistics units in line ministries is appointed by the Minister of this Ministry.

As a Government agency responsible for state management of statistical activities and compiling all kind of statistics in Vietnam, the GSO is leading Agency also in compiling social statistics. Other Ministries, such as Ministry of Labour and Social Affairs, Ministry of Health, Ministry of Education, Ministry of Culture and information, National Committee for population and Family Planning ... are involved in compiling social statistics by providing the GSO with information got from their administrative records or their own surveys. In Vietnam as in other countries, the social statistics are based on a combination of administrative records and specialized censuses and surveys.

Social statistics in Vietnam usually are posted in the following publications: Statistical Yearbook of the GSO, Results of Household and other specialized surveys, publications of Population Census, Yearbook of Ministry of Health, Yearbook of Ministry of Education. Yearbook of Ministry of Culture, Results of Labour force and employment survey of the Ministry of Labour and Social Affairs. The GSO also posted special publications on social statistics, such as Social Indicators in Vietnam for 1986-2000, Social statistics for 1990-1995.

Within the system of the GSO, Department of social and environment statistics is responsible for collecting and compiling social statistics in Vietnam. The personnel of the Department is 25. In each provincial Statistics Office there is Division of population and social statistics, consisting of about 3-4 statisticians involved in social statistics. In related line ministries all staff of statistical units are involved in Social statistics., in general they are about 4-5 people.

However, currently the GSO is facing some difficulties and challengers in compiling social statistics. It is very weak collaboration of Ministries and Agencies in compiling social statistics. In some cases there is no consistency in methodology, coverage of

indicators. Statistics provided from line ministries to the GSO are not on time, usually with big late. Quality of statistics provided is also a topic for discussion.

B. Historical Background

The General Statistical Office of Vietnam collected social statistics since date of it's establishment. The first social statistics series compiled in 1956, but only for major indicators related to living standard survey, separately for rural and urban population. From 1996, the GSO compiled much more social data and almost indicator with gender breakdown. Usually the GSO compiled data in annual base. In monthly and quarterly statistical reports of the GSO included only statistics related to very special social event. Currently, the GSO compiles statistics by the following groups of indicators:

- Population and family planning;
- Labour force, employment and unemployment;
- Health care and child nutrition;
- Clean water supply and environment sanitary;
- Education;
- Culture and entertainment;
- Child and mother care;
- Poverty alleviation and social development;
- ...

Annual social statistics series are available in Statistical Yearbook of the GSO, usually releases in April. The main users of social statistics in Vietnam are government agencies and international organizations, research institutes and universities as well as researchers and students.

C. Data Sources

Social statistics are fully dependent on population statistics and should be seen as integrated elements in the system of population statistics. In Vietnam social statistics are compiled in the base In Vietnam as in other countries, the social statistics are based on a combination of administrative records and specialized surveys. The number of teachers and pupils at various levels come from the school system, the number of doctors, nurses, attendants ... come from employment records or from administrative procedures of licensing such professionals. Number of patients treated for various illnesses come from the records from doctors and hospitals. Number of crimes of various types come from the police or court records.

To be more meaningful such statistics based on administrative records need reliable population statistics to provide a base for calculation of rates and proportions for social indicators like number of doctors, nurses, hospital beds per 100, 000 population.

Population statistics is generally based on censuses, surveys and vital registration. The last population census was conducted in Vietnam in 1999. Besides canvassing a larger population, the 1999 census also included in a 3 percent sample questions on fertility and death in the household in the last 12 months.

Some social indicators such as living conditions, level of social services in rural areas are also compiled from Rural and Agricultural Census conducted every 5 years, in the year ending by 1 and 6.

Very important resource of social statistics is Living standard survey that the GSO conducts every 2 years. Statistics on household income, expenditure and consumption is very important outputs of national official statistics. The needs are strong for such data in a social perspective as basis for evaluating living conditions, prevalence of poverty, consumption pattern and nutrition.

The various social sectors are also served by statistics from specialized household surveys, such as labour force surveys, literacy surveys, The most important arguments for these specialized surveys are that meaningful interpretations of change in statistical series based on administrative records alone often can not be made.

Social statistics can be compiled also on the base of vital registration. In Vietnam vital registration is implemented by the system of Ministry of Justice. A system for vital statistics covers variables like Birth, Death, Residence and Migration of the population.

D. Estimation Methodology

The Population and Social Statistics Departments of the GSO comprise the subject matter programs that collect data from and about households and individuals. Estimates are coordinated by the use of a common sampling frame derived from the census register of households. Surveys samples are post-stratified so as to conform to the updated population numbers and the age/sex structure by main regions. All household surveys collect a set of common data (sex, age, occupation, education and

household size and composition, that are used to report the variables by sub groupings.

It is needed here to emphasize very important role of international assistance in development of statistical methodology in Vietnam. Concerning social statistics, with assistance of UNDP, WB, Sida (Sweden) and ADB, the GSO has successfully implemented projects in improvement of social statistics. Two Living standard surveys in 1992 and 1998 were conducted with funds and methodological assistance of UNDP, WB and Sida (Sweden). ADB has assisted a project named "Financing for social services".

The results of international assistance is fact that the GSO could improve its system of social statistics and released some data publications as well as methodological manual "Guide to social statistics".

(See attached annex 1)

E. Social Statistics Compilation Program

The General Statistics Office of Vietnam has submitted to the Government "Programme for Development of Vietnam Statistics for the period 2001-2010". The most important part of the Programme is to improve legislative environment for statistical activities. In this concern, draft Statistical Law has been approved by the Government and will be submitted to the Parliament for consideration and approval.

The new Statistical Law will be a basis for coordination of all statistical activities in the country. The Law will also help the GSO in strengthening its legal mandate in implementation of statistical works as well as improving administrative records system.

In this programme the system of social statistics for Vietnam will continue to be based on surveys integrated with the system of population statistics. The GSO will continue to develop its present infrastructure for household surveys in terms of the household sampling frame, efficient sampling design, methods of estimation and system of maps to be able to coordinate the system and estimates from different surveys. The GSO will also develop a general household surveys scheme.

In this direction, the Project jointly assisted by the World Bank, UNDP and Sida (Sweden) calling “Strategy of Household surveys in Vietnam” was prepared and will be signed this year.

Attachment: Annex 1: Household surveys in Vietnam

II. Household Surveys in Vietnam

The GSO and other government organizations have administered a large number of nation-wide household surveys in Vietnam since 1990. In this section we review all nation-wide surveys conducted since 1990, focusing on those that collect economic and social data. We give particular attention to the VLSS and MPHS surveys conducted in the 1990s. In addition, we also review current plans for future surveys in Vietnam.

1. Nation-wide Household Surveys in Vietnam in the 1990s.

In the 1970s and 1980s, very few nationwide household surveys were conducted in Vietnam. This reflects the wartime conditions in the early 1970s and the lack of government resources throughout this entire period. This situation began to change in the 1990s. The adoption of the renovation (*doi moi*) economic policies that began in the late 1980s brought about many changes in the social and economic conditions of Vietnamese households, which led to a need for new information to measure those changes. At the same time, donor interest in and support for Vietnam increased in the 1990s, which provided increased resources for conducting household surveys during that decade. These resources, along with budgetary allocations by the Vietnamese government, were put to good use in the form of a wide variety of household surveys in Vietnam. The following paragraphs summarize the kinds of surveys conducted in Vietnam in the 1990s.

a) Socioeconomic surveys.¹ The two main surveys that collected a wide range of data on socioeconomic characteristics of Vietnamese households in the 1990s were the 1992-93 and 1997-98 Vietnam Living Standards Surveys (VLSS) and the Multi-Purpose Household Surveys (MPHS) conducted in 1994-97 and 1999.² Both of these surveys were implemented by the Department of Social and Environmental Statistics in the GSO. These two surveys are the most important for this report, and as such they will be discussed in more detail below. Another survey was the 1993 Rich and Poor Survey. This survey was conducted only once, in 1993, and was replaced by the MPHS in 1994. A final survey to note is the 1996 **Social Sector Financing Survey**, which reinterviewed about 3000 households that participated in the 1996 MPHS survey to collect more detailed information on education and health topics. For our purposes this can be regarded as an extension of the MPHS survey.

b) Demographic surveys. The main demographic survey conducted in Vietnam is the **Population Dynamics Survey**, which has been conducted annually by the GSO's Department of Population and Labor Statistics since 1993. The sample sizes for these surveys are very large, often around 200,000 households. The purpose of this survey, which is supported by the U.N. Population

¹ In addition to the surveys discussed in this paragraph, there was also a "Government Staff and Farmer Expenditure Survey" that was last conducted by the GSO in [the period between 1980s – 1990s](#) and a "[survey on poverty-agricultural rural](#)" jointly conducted by GSO and the Ministry of Agriculture and Rural Development in 1990. Little information is available on these surveys because they were conducted so long ago.

² The MPHS is sometimes referred to in English as the "Multi-Objective Household Survey (MOHS)". This is simply an alternative translation of the Vietnamese name for the survey.

Fund (UNFPA), is to collect annual information on fertility, mortality and migration. The name and exact content of this survey has changed over the years. For example, it was known as the “Multi-Round Survey” when it was conducted from January 1994 to October 1998. A second demographic survey is the **Multi-Indicator Cluster Survey (MICS)**, which is supported by UNICEF. It has been implemented in 1996 and 2000 by the GSO’s Department of Social and Environmental Statistics. This survey covered 6000 households in 1996 and 7600 in 2000. It was designed to collect indicators of child health and well-being. Third, in 1994 the Department of Population and Labor Statistics of the GSO completed the **Intercensal Demographic Survey**, which was designed to collect demographic and health information (like DHS) in the midpoint of the 10 year period between the 1989 and 1999 Population Censuses. This survey had a sample size of 13,200 households and is supported by the UNFPA. Finally, a **Demographic and Health Survey (DHS)** was done in 1997 by the Population and Labor Statistics Department of the GSO. This survey had a sample size of about 5,600 households and was funded out of a World Bank loan for a family health project.

c) Labor Force Surveys. Up to now (the year 2000) all labor force surveys in Vietnam have been conducted by the Ministry of Labor, Invalids and Social Affairs (MOLISA) in collaboration with the Population and Labor Statistics Department of GSO. The main labor force survey conducted in Vietnam is the annual **Labor Force Survey**, the first of which was undertaken in 1997 (this survey is called the “Labour and Employment”, but we will refer to it as the “Labor Force Survey”). These surveys have very large sample sizes, of around 105,000 households. They focus only on the working age population (age 15 and older). This survey builds on earlier “one-shot” surveys done in 1994 (urban areas only) and 1996 (all Vietnam).³ All other surveys done by MOLISA focus on special populations (e.g. university graduates, returnees from Czechoslovakia or Germany, street children, disabled people, women working in the informal sector) or are not nationwide in scope (e.g. migration in Daklak province and migrants from rural areas to Hanoi).

2. The VLSS and MPHS Surveys in the 1990s.

Most of this report will concentrate on the VLSS and MPHS surveys, since they are very similar to each other and are the only general socioeconomic surveys conducted on a nationwide scale in Vietnam. Thus it is worth describing these surveys in detail.

a) The VLSS Surveys. The 1992-93 and 1997-98 VLSS surveys were implemented by the GSO⁴ with financial support from the United Nations Development Programme (UNDP) and the Swedish International Development Agency (SIDA), and with technical support from the World Bank. The design of the surveys closely follows the standard design of the World Bank’s Living Standards Measurement Study (LSMS) surveys. The questionnaires are very detailed and cover a very wide range of topics, specifically income, expenditures, housing, education, health (including height, weight and arm circumference of all household members), employment, migration, agricultural activities, household business activities, fertility, and savings and credit. The sample size is much smaller than most other surveys in Vietnam, 4800 households in 1997-98 and 6000

³ We have also been told of a Labor and Social Issues survey conducted by MOLISA from 1992 to 1997, but no one in the GSO was familiar with this survey, and we were unable to discuss it with MOLISA.

⁴ The 1992-93 VLSS was part of a larger UNDP project with the State Planning Committee (now the Ministry of Planning and Investment, MPI). MPI organised and managed the first VLSS with participation from the GSO.

households in 1997-98 (note: the sample size for 1997-98 had originally been planned for 8000 households but a shift in funds from the survey to training activities reduced the final sample size to 6000). There are several characteristics of the two VLSS surveys that are important for understanding their success and for comparison with the MPHS surveys:

- Both VLSS surveys were implemented over a 12 month period, which allowed the GSO to train a small number of interviewers (30 in 1992-98 and 38 in 1997-98) who then worked full-time for 12 months. All of these interviewers were GSO staff, as opposed to individuals hired to by GSO for a short time.
- Both VLSS surveys had a high rate of supervision, with one supervisor for every two interviewers.
- Data entry was done in the field using a system that allowed interviewers to go back to households to check inconsistencies found in the data. This was done using a customized software program for data entry that detected data inconsistencies.
- The VLSS questionnaire writes out the exact questions to be used by the interviewers for all questions. This reduces variation across interviewers in the questions actually posed to the households.

Both government agencies and international donors have expressed general satisfaction with the reports and data generated by the VLSS surveys. For example, both surveys were a major input into the report “Vietnam: Attacking Poverty” that was written in 1999 by the Government-Donor-NGO Poverty Working Group led by the World Bank. Another example, is the comments received from different government agencies in the meeting of government data users convened during this mission on August 10, 2000. In that meeting several participants said that they used the VLSS results in their policy work, including representatives from the Ministry of Health, Ministry of Agriculture and Rural Development, Ministry of Education and Training, and the Ministry of Planning and Investment. However, three major complaints have been raised about the VLSS surveys:

- The most common complaint is that the sample size was too small. In particular, many ministries would like the sample size to be large enough to provide province level estimates of various indicators.
- Second, some ministries would like the survey to be implemented more often, rather than only once every five years.
- Finally, the cost of the VLSS surveys per household interviewed is high. For example, according to GSO the costs of 1997-98 survey were about \$977,000, which implies a cost of about \$163 per household⁵ If the same survey was done in the future, the cost would probably come down to under \$750,000 (i.e., less than \$125 per household) because much of the equipment bought for the first two surveys (especially vehicles and computers) would not need be bought again and fewer international consultants would need to be hired..

⁵ This total cost of the VIE/95/043 project was \$1,687,000. The remaining funds were used for training activities, both in Vietnam and abroad, and are not costs that are directly related to implementing the survey.

b) The MPHS Surveys. The MPHS was implemented on a regular basis from 1994 to 1997, and then again in 1999. Much of the motivation behind it was to overcome the first two complaints about main limitations of the VLSS surveys, namely their small sample size and the long period of time between the two VLSS surveys. Thus the 1994-97 MPHS surveys had sample sizes ranging from 45,000 to 47,000 households, and the data were collected for every year except 1998.

The fieldwork in the 1994-97 MPHS surveys was rather complicated, so it is worth explaining in some detail. First, 45,000 households were interviewed for the first time in July, 1995, at which time they were asked to provide, by recall, information on all of 1994 and information on the first half of 1995. The information on all of 1994 is the basis for all figures from the MPHS that refer to 1994. Second, these same households were reinterviewed in January, 1996, at which time information was collected on the second half of 1995. Much of the income and expenditure data collected for the second half of 1995 is based on diaries that were left with the households in July of 1995. Households were visited once every month to check on how they were filling out the diaries. These monthly visits proved expensive, and the use of diaries had other problems, so this form of data collection was discontinued in 1996. Data for the year 1995 from the MPHS are based on the information for the first six months of 1995 collected in the July 1995 interview, and the information on the second six months of 1996 (part of which is based on household diaries) that was collected in January 1996. Third, the same households were interviewed in July 1996, with the reference period the first 6 months of 1996, and again in January 1997, with a reference period of the last 6 months of 1996. Information from these last two interviews was used to generate the MPHS figures for 1996. Finally, the households were interviewed in July, 1997, at which time the reference period was the first six months of 1997, and again in October 1997, at which time the only questions asked concerned employment and labor force status. Thus, the data on income and expenditure in 1997 are not for the whole year, and the MPHS data for 1997 are used only for estimates on education, health and labor force activities, not for estimating income or consumption.

While the 1994-97 MPHS survey had two advantages over the VLSS surveys, namely a larger sample size and data collection every year, there were several problems in conducting the survey:

- First, the combination of large sample size and short time to conduct the interviews (one month) required the GSO to use approximately 1500 interviewers (compared to only 30 to 38 interviewers in the two VLSS surveys), which, in turn, forced it to hire temporary workers in some provinces. The training of these interviewers was probably insufficient, which appears to have led to problems with the quality of some of the data.
- Second, there were not enough experienced supervisors to check carefully the work of these 1500 interviewers, which is also likely to have led to problems of data quality and delays.

Third, all of the data entry was done in Hanoi, which slowed down data entry and made it extremely difficult, when data errors were caught, to return to the households to correct them. This resulted in delays in analyzing the data and publishing the results. Summary results for 1994,

which were based on the July 1995 interview, were published in May 1996, with the full 1994 results published in April 1997. Summary results from the 1995, 1996 and 1997 MPHS were published in July 1998 and the full results for 1995-97 were published in October 1999.

Many of these problems with the collection and analysis of the MPHS data could probably have been avoided or greatly reduced if the sample size had been smaller and more time had been allowed for planning the survey. Overall, the objectives of that survey were probably too ambitious given the capacity constraints that the GSO faced at that time.

On a more positive note, the MPHS for 1999, which is based on interviews conducted in January 2000, appears to be proceeding much more smoothly. [Its summary results were published in June 2000.](#) The GSO expects that the full report for this survey will be finished by the end of 2000. Part of this is no doubt due to the smaller sample size of the 1999 MPHS: 25,000 households instead of 45,000. Moreover, several improvements have been made to the survey that appear to have increased the quality and timeliness of the data. In particular:

- All interviewers were GSO staff, not newly hired temporary workers.
- Data entry was done at the provincial level, which provided some opportunities to revisit households if the data entry program detected problems with or inconsistencies in the data collected.
- There were many more supervisors per interviewer, about one supervisor for every three interviewers; this was particularly useful in that supervisors could check each questionnaire in the communes after the interview was completed and could also revisit some of the households to see whether the interviewer had incorrectly recorded some aspect of the socioeconomic conditions of the households.
- [5% of surveyed households has been reinterviewed.](#)

Despite these improvements in the 1999 MPHS, there are some potential shortcomings:

The survey questionnaire was shortened so that it is basically an income and expenditure survey, with almost no other socioeconomic data. In particular, unlike the 1994-97 MPHS surveys, almost no data were collected on health, education, housing or employment.

The GSO regards the smaller sample size as too small for it to be representative at the province-level; whether this is in fact the case will be discussed in detail below. The point here is that this is how the GSO regards the survey and indeed it has informed the provinces that the sample sizes are too small to be representative at the province-level.

The questionnaire requires interviewers to do a large number of mathematical calculations, mainly to check whether income is greater than expenditures. The burden of these calculations was made clear upon our visit to the Thanh Hoa Province Statistical Office. The director of that office estimated that the typical interview took about 8 hours at the household, followed by another 8 hours of the interviewer doing the various calculations, using scratch paper, to do the computations. This same comment was later made by the HCMC Province Statistical Office

Many parts of the MPHS questionnaire itself are little more than tables for the interviewer to fill out without necessarily asking precise questions. As explained below, this can lead to

inconsistent answers among interviewers. In addition, it forced households to answer questions that are in fact very difficult for them to answer, such as asking in a single question how much the household has spent on health fees of all types, for all household members, in the past 12 months.

The sample drawn for the 1999 MPHS was a randomly drawn sample of households that were interviewed in the 1994-97 MPHS, which in turn were based on lists not updated since 1994.

3. Plans for Future Surveys.

The GSO budget is allocated on a year by year basis, so that although they make plans for surveys 2-3 years in advance, they do not know for certain whether they will have the funds to implement them until about one year before the survey is to be done. Thus the GSO is not accustomed to making long-range survey plans. However, there are some definite plans for 2001, as well as some more general ideas about what surveys will be done in 2002 and beyond. The following paragraphs summarize future survey plans:

a) MPHS for 2000. The Social and Environmental Statistics Department of the GSO has proposed to undertake the MPHS again for the year 2000. Following the design of the 1999 MPHS, the interviews will be conducted in January 2001 and the reference period for all households will be January to December, 2000. The sample size will probably be 25,000 households, though perhaps somewhat larger. The final approval to carry out this survey has not yet been given by the leadership of GSO, although it is likely to be given very soon.

b) National Health Survey. The Ministry of Health (MOH) has been planning a major health survey. The plans are almost complete. MOH has asked GSO to conduct the survey on their behalf. The current plan is to conduct this survey in March, 2001. The sample size will be approximately 36,000 households.

c) Labor Force Surveys. As explained above, MOLISA has conducted the labor force survey, in collaboration with the GSO, each year from 1997 to 2000. These surveys will almost certainly continue in the future, and it is likely that beginning in the year 2001 the will become the complete responsibility of the GSO (a decision on this will be made later this year).

d) Population Dynamics Survey. This survey, which was described above, will continue to be carried out by the GSO on an annual basis.

III. Sample Designs for Household Surveys in Vietnam

1. Background

The General Statistics Office (GSO) of Vietnam wishes to embark on a program of household surveys for the next decade (2001-2010). This agenda can be divided into two parts, plans for the MPHS surveys in the next decade and plans for other surveys. This section makes recommendations for the sample design of future household surveys conducted by the GSO. Many of these recommendations are very general and apply to almost all future household surveys, while others (such as explicit suggestions regarding the sample size) are made specifically for future MPHS surveys.

This section first reviews past experience of the GSO regarding sample designs of household surveys, pointing out both strengths and weaknesses. It then goes on to make general recommendations on the sample design of future household surveys, followed by explicit suggestions regarding the MPHS surveys for 2002 to 2010.⁶ Finally, it provides some explicit suggestions for donor agencies. These recommendations are based on extensive discussions with the staff of various departments, line ministries, and subject-matter specialists within the national statistical system. While all recommendations regarding sample design are presented in this section, the details of the reasons underlying them and a comprehensive discussion of the sample design recommendations are contained in an appendix at the end of this report.

2. Sample Designs of Previous Household Surveys.

A number of household surveys and censuses have been conducted in Vietnam in the past decade, as explained in Section II. Most of these surveys have been conducted by the GSO or in collaboration with GSO departments and staff. These surveys highlight many of the current strengths of the GSO. In particular, the GSO has made significant progress in moving from a system of mainly administrative reporting to one based on household surveys and censuses, so that it now has an efficient data collection machinery with considerable experience in survey and census implementation. It also has trained staff with experience gained from involvement with many surveys, from questionnaire design to supervision of fieldwork, the establishment of a database, and data analysis. Finally, the GSO has a network of statistical offices at the provincial and district levels, which have a clearly defined technical and administrative relationship with the GSO. Also the GSO appears to have a good working relationship with the various ministries and other government agencies.

Household surveys conducted in Vietnam in the past decade have had some common sample design features. Thus, even though the surveys differ in target populations, content, sample size, and objectives, they generally have the following features:

- (i) They have employed a multi-stage stratified area probability sample design.
- (ii) They define strata using some combination of administrative regions (groups of provinces), often classified by urban status (urban/rural).
- (iii) The primary sampling units (PSUs) are communes in rural areas and residential clusters in urban areas.
- (iv) The secondary sampling units are villages within communes or blocks within wards.
- (v) The ultimate sampling units are households. The information is collected on all members of selected households.
- (vi) Because of the need to obtain adequate sample sizes in each sampling domain, some strata are oversampled (disproportionate allocation), with the resulting bias corrected by the construction of weights.
- (vii) There is a strong tendency to use very large sample sizes in order to reduce sampling errors, yet explicit sampling errors are rarely computed.

⁶ As explained in the following section, plans for the 2000 MPHS are too far along to be significantly altered at this point, and the report also recommends that the next MPHS be done in 2002, not 2001.

- (viii) Little attention is paid to non-sampling errors, and any possible relationship between sample size and non-sampling errors.

This general approach to sample designs in recent household surveys conducted in Vietnam reveals both strengths and weaknesses. The most important strengths are:

- (i) Most GSO surveys are based on valid probability sample designs, often recommended by internationally recognized experts.
- (ii) The samples can generate unbiased estimates of important characteristics of interest at the national level, regional level, and urban and rural areas within these domains.
- (iii) Although sampling errors are rarely calculated, the information needed to calculate them is available.
- (iv) GSO staff at the national level understand many of the basic concepts of sample design, and are very quick to grasp new, more complex concepts when they are introduced.

Despite these strengths, there are many weaknesses regarding the sample designs of household surveys conducted in Vietnam. The most important weaknesses are:

- (i) GSO staff at the national level are not familiar with some important aspects of statistical sampling, and their implications in developing a sample design that can generate accurate and reliable survey data.
- (ii) GSO staff have only limited experience in variance estimation techniques for complex survey data.
- (iii) Survey designs show excessive focus on sampling errors almost to the exclusion of the equally important concept of non-sampling error.
- (iv) Some surveys have limited technical documentation of their sample designs.
- (v) With the exception of the VLSS surveys, the sample sizes are usually very large. Two very serious problems with large sample sizes are that they are more expensive to implement (relative to smaller ones) and they are prone to producing data of poor quality due to non-sampling errors. Non-sampling errors come about because the quality of field procedures is more difficult to monitor, and the implementation of post-survey quality control procedures is much more expensive and time-consuming.
- (vi) While there are legitimate reasons for using a three-stage design (communes/wards as the first-stage units, villages/residential clusters as the second-stage units, and households as the ultimate sampling units), such a design is more complicated than necessary. Each stage of sample selection introduces another level of variability in the construction of household level weights, and some penalty is inevitably paid in the precision of survey estimates. The sample design can be simplified by using a two-stage PPS design, as explained below.
- (vii) The designs use a uniform cluster size across reporting domains. This has cost implications. Considering the difference in travel costs between urban and rural areas of Viet Nam as reported by GSO staff, it would be more efficient to use larger cluster sizes in the rural areas than in the urban areas. Such larger clusters can be

based on enumeration areas (EAs), supervisor areas (groups of EAs), or even communes. If transportation problems are encountered during data collection, then some of the areas can be segmented into more manageable parts. The process of segmentation of areas needs to be discussed at a future date.

- (viii) Most of the sampling designs used sampling frames based on the 1989 Population Census. In the light of the considerable changes in the Vietnamese population and in the number and nature of administrative subdivisions at the provincial and commune levels, the 1989 frame can be justifiably considered out of date, especially for surveys conducted during the last five years. The main problem with an out-of-date sampling frame is that there is no reliable information on the population size of the EAs. This precludes the selection of first-stage units with probability proportional to size (PPS) in terms of the number of households or number of people. The penalty for this is that the resulting weights will be highly variable and hence the precision of the survey estimates will be lower (standard errors will be higher), relative to a PPS design. This especially the case if it turns out that the sizes of the first-stage units are widely variable.

3. General Recommendations for Sample Designs of Future Household Surveys.

The following sample design recommendations are based on careful consideration of the current situation of household surveys in Vietnam, the data requirements that must be met by future surveys, and an assessment of the constraints under which the proposed program of surveys will be implemented. These recommendations are general, because they are geared toward almost all future household surveys in Vietnam, including the MPHS surveys to be conducted in the next decade. In addition, another subsection immediately below makes very specific recommendations for the sample design of MPHS surveys from 2002 to 2010. In both sets of suggestions, the recommendations attempt to capitalize on the strengths of the national statistical system and of past surveys, while at the same time minimizing the weaknesses.

- (i) The GSO should use a stratified two-stage area probability design. The primary sampling units (PSUs) should be enumeration areas (EAs) in urban areas and supervisor areas (groups of EAs) in rural areas, as identified and used in the 1999 Population and Housing Census, assuming the availability of good-quality maps covering all such geographic units in the country. The difference in the choice of enumeration areas between urban and rural areas reflects differences in the cost and difficulty of travel between urban and rural areas.
- (ii) The maps used in the 1999 Population and Housing Census should be put into a computerized database, which will provide easier access to those maps (and easier modifications of those maps) for the purpose of selecting future samples. The GSO will probably need donor assistance for this task.
- (iii) For any survey that is intended to provide accurate estimates at the province level (which includes future MPHS surveys), the sampling frame of enumeration areas should be stratified on the basis of the 61 provinces prior to sample selection. Within each province, there should be explicit stratification for urban and rural areas. Furthermore, if one or more important socio-economic and demographic variables can be identified from the 1999 census, such as economic zone, then they can be

incorporated into an implicit stratification. Implicit stratification helps reduce the variability in survey estimates.

- (iv) At the first stage of sampling, enumeration areas (urban areas) or supervisor areas (rural areas) should be selected with probability proportional to the number of households in the area. Sampling of PSUs should be done separately within each province. To improve the precision of survey estimates at the provincial level oversampling should be carried out in small provinces in order to obtain larger sample sizes than would be obtained from proportional allocation.
- (v) At the second stage, housing units would be selected systematically within each sampled area in such a way that the overall probability of selection of each sampled housing unit is the same. All households within each sampled housing unit should be interviewed, and data should be collected on all permanent and semi-permanent residents of these households. In later rounds of the implementation of the survey, the second stage of sample selection should be preceded by a listing operation. However this listing of households will be restricted only to sampled PSUs.
- (vi) Non-response at the household level should be handled by adjusting the sample weights to account for non-responding households. In past surveys in Vietnam, a reserve sample was used primarily for the substitution of households that did not participate in the surveys. However, attempts to substitute for non-responding households take time, and errors can be made in the process. For example, a replacement may be made using a convenient household rather than the household specifically designated to serve as the replacement for a non-responding household. This procedure of adjusting sample weights for non-response is the standard practice in major surveys throughout the world. The actual procedures for doing such adjustments should be discussed at a later date, along with other aspects of the implementation of the proposed sample design. GSO staff should be given training in avoiding non-response and in adjustment methods for non-response.
- (vii) The GSO should construct a Master Sample of PSUs, which offers many advantages for a long-range program of multiple surveys. In particular, there should be potential for considerable savings in costs related to start-up activities and PSU selection for future surveys. Master samples are widely used in developed countries, and also in some developing countries (such as South Africa; see Pettersson, 1998). Statistics Sweden, which has been working in Vietnam for many years in the areas of national accounts and 'economic' statistics, has also recently proposed that GSO should construct a Master Sample. It should, however, be appreciated that the design and implementation of a master sample is a major undertaking. If this recommendation is adopted the GSO must ensure that sufficient resources and competent staff are assigned to the project. (Note also that a master sample also provides considerable flexibility in the selection of households, depending on the objectives of the surveys. For instance, a rotating panel design can be built into the selection of households, which will permit efficient estimation of changes or trends in some indicators and at the same time reduce response burden and time-in-sample bias, relative to a full panel survey of households).

- (viii) The GSO's Institute for Statistical Sciences should undertake a rigorous study of the impact of increasing the number of households by the provincial statistical offices (beyond the sample size allocated by the national GSO) on provincial estimates of key variables of interest. This augmentation of sample sizes has been done by a number of provinces in recent MPHS surveys, ostensibly because the allocated sample sizes are deemed not to be "representative" of the provinces. In addition to the technical arguments against very large sample sizes, it would also be useful to compare the estimates based on the two samples (the GSO sample and the augmented sample) on the basis of the size of the estimates, the standard errors of the estimates, and various indicators of non-sampling errors.
- (ix) The GSO, in collaboration with the Institute for Statistical Sciences, needs to do a better job of explaining the costs and benefits of large sample sizes to the provincial statistical offices, in order to curtail the excessive demand for large sample sizes at the provincial level. Demands for precise estimates for all provinces and for urban and rural subdomains within provinces should be carefully reviewed on the basis of their costs and benefits. Both producers and users of household survey data must be aware of the implications of large sample sizes for survey costs and non-sampling errors. Further consideration should also be given to the option of constructing reporting domains from a judiciously defined grouping of the provinces. Such a grouping should be based on analytic interest at such a level.

4. Specific Recommendations for Sample Designs of Future MPHS Surveys.

The mission's discussions with government ministries and donor agencies clearly showed that the future MPHS surveys must satisfy the following two criteria:

- MPHS surveys must generate data that are deemed to be "representative" not only at the national and regional levels, but also at the provincial level.
- While donor agencies may provide technical assistance to the GSO for preparatory activities and general capacity building, the recurrent costs of implementing future MPHS surveys must be borne by the Government of Viet Nam.

Regarding the first criteria, the concept of "representativity" must be clarified. In statistical terms, this means that estimates should be unbiased and reasonably precise. An estimate is unbiased if its expected value is equal to the true value of the variable being estimated. That is, the only reason that the estimate may differ from the true value is due to the fact that it is based on a random sample of total population and thus contains some random sampling error. Precision is the size of that sampling error. As the sample size increases, so does the precision of the estimate. The real issue in any discussion of "reliability" is how much precision is adequate. For example, if the poverty rate in a particular province is estimated to be 40%, the question is whether a 95 percent confidence interval of 35% to 45% is sufficiently accurate for policymaking purposes.

Needless to say, these two criteria impose limitations on the types of sample designs that can be implemented in future MPHS surveys, given the financial and technical capacity of the GSO. Yet we believe that a satisfactory sample design can be developed that meets these constraints. All of the recommendations made immediately above apply to the sample design of future MPHS

surveys, especially recommendations (i) – (vii). However, some additional specific recommendations can also be made:

- (i) For the 2002 MPHS survey, the GSO should use an overall sample size of approximately 30,000 households. With oversampling in small provinces, this will yield sample sizes of 400 or more for about three quarters of the provinces. We consider these sample sizes adequate for most decision making purposes regarding funding or other resource allocation decisions at the province level because (as shown in Table 1) a sample size of 400 generates a confidence interval of 43%-57% for a variable whose true value is 50% (assuming a somewhat conservative design effect of 2).
- (ii) Smaller provinces should be oversampled to ensure that each province has at least 300 households (with the possible exception of Bac Can province, which has only 283 households even with the highest over-sampling rate). The oversampling should be done in such a way that the overall sample size of 30,000 households is not exceeded. The degree of oversampling will depend on the desired level of precision for survey estimates at the national level. A detailed plan is provided in the appendix on sampling issues.
- (iii) The sample will be drawn randomly and independently within each stratum based on provinces and urban/rural domains within provinces. Therefore, the sample will provide unbiased estimates within these strata. As for the precision of the survey estimates based on the sample, the allocation proposed in the appendix will provide precise estimates for about three fourths of the provinces (those with sample sizes of at least 400 households). However, because of small sample sizes, the survey estimates will not be very precise for about one fourth of the provinces, and for urban rural domains within all provinces (except Ho Chi Minh City, Thanh Hoa, and Hanoi, where sample sizes are large enough even for urban and rural subdomains) In some provinces where the survey estimates may not be very precise, they may be sufficiently precise for decision-making purposes.
- (iv) With the overall sample fixed at 30,000 households, the options for increasing the precision of estimates for urban/rural domains are very limited. Over-sampling in provinces at rates higher than those proposed in Table 3 is not advisable. The dangers of over-sampling are discussed in the appendix. If rural and urban comparisons and estimates are a high priority in the program of surveys being planned, then the following options can be considered:
 - For a start, produce estimates at the province level. Consider using several years' worth of data to produce multi-year-average estimates for urban / rural subdomains, provided that multi-year averages make sense from a subject-matter point of view.
 - Pool data across groups of contiguous small provinces in the same region.
- (v) In urban areas, approximately 10 households should be sampled from each enumeration area (EA). In rural areas, about 20 households should be sampled from each supervisor area (a group of enumeration areas). Since about 75 percent of

Vietnamese live in rural areas, this implies that between 1500 and 2000 PSUs will be needed.

The GSO should provide short term training to officials from the provincial statistical offices to explain the costs and benefits of large sample sizes.

