

Chapter 2

Methodology

2.1. Sampling Frame and Sample Size

The SAVY sample is a national representative sample of youth (persons ages 14-25) living in households across the eight economic regions of Viet Nam. The sample was drawn from the sub-sample of 45,000 households in the 2002 Viet Nam Living Standards Survey (VLSS 2002), with a multi-staged and stratified design. The youth in the SAVY sample design are sufficient to represent the nation as a whole, as well as the urban and rural sectors separately. The largest cities (Hanoi, Ho Chi Minh City [HCMC]) were over-sampled in order to provide for increased statistical power in that segment of the total population of youth.

Forty-two out of 61 provinces were selected for the SAVY sample¹, using the probability proportional to size (PPS) method to maintain representativeness. At the next stage of sampling, enumeration areas (EAs) in each province were selected. In those EAs sampled, all youth aged 14 through 25 were identified (i.e. those born between 1978 and 1989, males and females, married and non married from the 20 households that had been selected for the Vietnamese Living Standards Survey [VLSS] 2002). The youth cohort represents all youth, but not those living in special arrangements, such as barracks, re-education centers, social protection centers, factories and dormitories.

The 61 provinces in the VLSS 2002 sample included 2,250 EAs, and the 42 provinces selected for SAVY included 1,643 EAs. From these, a total of 446 EAs were selected for the SAVY sample. These EAs contained 8,920 households corresponding to a population of 40,140 (about 4.5 persons per household). Since youth aged 14-25 account for 24.5% of the total population (the figure in the 1999 census), the anticipated number of youth in the SAVY sample was approximately 9,835. If the mobilization rate (percentage of eligible youth actually interviewed) was 90% then the number of youth interviewed would be estimated to be about 8,850. In the actual SAVY field experience, the mobilization rate was 85% and the number of completed

interviews was 7,584.

The sample is therefore representative, and provides sufficient cases for analysis at the national level, within urban and rural sectors at the national level, by gender at the national level, and for each of the regions. Further detail on the sampling methodology is provided in the Appendix.

The SAVY mobilization rate is not low by the standard of such surveys of youth, and is only slightly lower than the GSO had expected. It should be noted that this is the first time this methodology has been used for a national survey. SAVY has certain characteristics that could lower the mobilization rate, particularly in the fact that its sampling frame was derived from another survey using household lists created one year earlier, in the way that local people's committees were called upon to mobilize youth, and in the relatively high rate of geographic mobility within the youth population.

Among those who did agree to go to the central location for interviewing, almost no one refused to answer the questions or fill in the self-completed part of the survey. The survey method (face-to-face interview with a self-administered second part), the quality of interviewers and the organization of the field work, including extensive supervision, were important factors that ensured the quality of the SAVY data.

2.2. Questionnaire

The questionnaire was designed through a very dynamic process, where experience from previous surveys was examined and opinions of young people were actively solicited to ensure quality and relevance. This process also helped to define the methodology and implications for fieldwork planning.

A number of stakeholders' agencies, including research institutes, were involved in the development of the questionnaire. This process ensured broad participation and ownership of the questionnaire and the survey.

The questionnaire design took place in two stages. In the first stage, experienced researchers, and others interested in the survey as stakeholders, were convened to a workshop by the MoH. Potential

topics, and the possible phrasing of questions using the questionnaire bank from previous studies in the region as reference, were fully discussed. Since some of the topics were deemed to be more sensitive than others, it was recommended that the questionnaire should be organized into two parts, one for an interview and the other for self-completion. On the basis of that workshop, a draft questionnaire was created for review by the workshop members and numerous others in stakeholder agencies, as well as by young people through a series of consultations.

Eight focus group discussions were conducted in Hanoi and HCMC, with around 60 young people of different ages in the 14-25 range who were either married or unmarried and either attending or not attending school. Participants gave detailed feedback about the terminology, the ways in which questions were posed and the sequencing of the questions, as well as which specific questions or issues they would prefer to respond to on their own, rather than with an interviewer. This process resulted in the rephrasing of a number of questions and changes to the self-completed section.

Preliminary training was conducted for field-testing of the questionnaire. Participants came from the GSO Office in Tuyen Quang, Hue and HCMC, representing the north, south and central regions of Viet Nam. A group of 50 young males and females, either married or unmarried and either attending or not attending school, participated in the interviewers' practice session. In the debriefing discussions, these young people expressed their feelings about the interviews, the questions asked, what they liked and did not like about the process, seating arrangements, ideas of what topics/issues they thought might still be missing in the draft questionnaire, and what they thought would be needed to make good interviewers. Field testing with around 180 young people from six communes in these three provinces then took place.

The second stage involved further vetting of questionnaire sections and was coordinated by the GSO. The review meeting following the field trips recommended the need for another field testing exercise, particularly because little experience had been gained from testing with urban young people and interviewing ethnic minority young people through interpreters. Following the second round of field-testing in Hanoi and Yen Bai, the feedback was incorporated to finalise the questionnaire for the

interviewers training. At the training, further revision and refinement of the questionnaire occurred prior to the field work.

The resulting questionnaire consisted of a total of more than 200 questions. SAVY experts then further modified the questionnaire in order to ensure the best phrasing possible, and to avoid technical terms. The first section was conducted as a face-to-face interview, with general questions categorized into topics. The second part of the questionnaire – and the part that makes the survey special – was an anonymous self-administered section, including 52 sensitive questions that youth preferred to answer in private. Originally, it was envisaged that the self-completed section would contain between 10-15 questions, but it became much longer as the youth consulted suggested that a lot more questions they perceived to be sensitive should be placed there. The questionnaire could be completed in 60 to 80 minutes, though this was longer for those unable to read the questionnaire and who required translation.

The specific information collected through the questionnaire includes:

- Personal demographics
- Schooling, education
- Vocational training, Work and employment
- Puberty: knowledge and behaviors about reproductive health
- Dating and friendship
- HIV/AIDS
- Injury, illness and physical health
- Attitudes, perceptions and behaviors
- Social factors and emotional wellbeing
- Mass media
- Future aspirations.

The complete questionnaire is found in the Appendix.

2.3. Survey Method

Data from young people were collected by assembling them together in a public space away from their homes (such as a people's committee office, or the cultural house of the hamlet), but still arranging for privacy during interviews. The fact that youth were asked to gather in one place had the potential to affect the response rate. To minimize this, the GSO



coordinated very closely with the Women's Union, the Youth Union and the local government, informing them about the purpose and importance of the survey so that they would then mobilize youth effectively.

An interviewer of the same sex, sitting side-by-side, interviewed young people, a method that had previously been tested during the training and testing phases. The face-to-face interview was of assistance to respondents who were generally not familiar with questionnaires and the coding of the responses. The interviewers checked the respondents' ability to self-complete the sensitive part (through interviewing and checking levels of education) and gave clear instructions on how to fill in the questionnaire before handing it over to the respondent.

After finishing the anonymous questionnaire, respondents were asked to post them in the box provided. This procedure was designed to ensure and demonstrate the privacy of the information.

2.4. Training and Field Data Collection

A central steering committee for the survey was established with members from the GSO, MoH, WHO, UNICEF, YU and WU.

Since this was a sociological survey asking for information covering a wide range of areas, including sensitive issues, the training to ensure high quality field staff was considered crucial to the quality of the results. Three training courses, over five-day periods including one day of practice, were organized by the GSO for 150 recruited data collectors. These collectors

were GSO staff from provincial offices.

Skills covered in training included how to communicate with youth, the specific content of the questionnaire, the underlying intent of each of the questions, and specific instructions for coding responses. The training methods included classroom instruction, group discussion and hands-on practice. More than 150 youth from different social backgrounds were invited to the sessions so that interviewers could practice interview skills and fieldwork procedures. Feedback from the young people was very valuable for both the instructors and the trainee interviewers. Only trainees who passed a post-training examination were allowed to participate in the survey.

Field data collection took place over 53 days. The trained interviewers were organized into 19 teams. The field schedule was prepared detailing the expected itinerary and travel of each survey team between EAs, districts and provinces. The People's Committee at commune level, through the YU and WU, provided notice to young people and mobilized them to participate in the survey.

2.5. Survey Errors and Limitations of the Data

The statistics in this report are estimates (also referred to as indicators) derived from SAVY, which is a sample survey. There are two types of error that occur in sample survey estimates: sampling and non-sampling. Sampling errors occur because the observations are made from only a sample of, rather than the entire, population. Non-sampling errors can be attributed to many sources. The joint effects of sampling and non-sampling errors determine the accuracy of the sample survey results. When a sample survey is designed using probability sampling techniques, it is possible to assess the sampling error associated with any particular estimate by using the survey data themselves.

2.6. Non-sampling Error

Non-sampling error is more difficult to quantify than sampling error because it has many components, each one of which requires its own evaluation study to assess effects on the survey results. Among the many sources contributing to non-sampling error are (1) non-

response from some of the households and/or persons selected to participate in the sample; (2) conceptual and definitional difficulties in the design of the questions that are asked of the respondents; (3) inability or unwillingness to provide correct information on the part of respondents; (4) mistakes by interviewers and data entry staff in recording or coding the data obtained; and (5) other errors of collection, processing and coverage. The evaluation of non-sampling error is also constrained because the statistical theory for doing so is underdeveloped compared to that for sampling errors, and also because there is an absence of knowledge about the true values in the target populations under study. For these reasons the designers and producers of large-scale surveys, including SAVY, rarely provide empirical results showing the type and magnitude of non-sampling error that may be present. Survey budget constraints effectively make its assessment infeasible.

Instead, efforts are usually directed at controlling and minimizing non-sampling error through such means as using previously validated questionnaire items, pre-testing of new questions, pilot testing of survey methods and operations, careful and intense training of interviewers, sample verification of data entry and coding, plus close supervision, observation and spot-checking in the field. All these steps were taken during the SAVY operations. Nevertheless it is important to describe, where known, the kinds of non-sampling error, or biases, that may be present in the data and to indicate, qualitatively, what the effects of these may be.

An important source of non-sampling error in SAVY is non-response. Its magnitude is measured by the non-response rate, calculated as $(1 - I/n)$, where I is the number of youth 14-25 for whom interviews were obtained and n is the number selected in the sample. In SAVY there were 7,584 interviewed youth and 9,989 selected-for a non-response rate of $(1-7584/9989)$, or 24.1%. This level of non-response is in line with the experience of other surveys focused on youth age groups. Moreover, the SAVY field design eliminated non-response at the household level, an additional source of non-response in most other surveys. The cited level of non-response can have a significant biasing effect on the results because without additional information we can only assume that the non-responding youth are similar, in terms of their characteristics and distribution, to the responding ones—an assumption which cannot be independently

verified. Comparisons between survey distributions and distributions available from other sources, notably the Health Surveys and the national censuses, suggest only a very limited bias is involved, though one exception should be noted. The SAVY sample of young people may slightly under-represent those in the older part of the age range who are married or who are working, or who are not enrolled in school. This is apparent in comparisons of age and sex-specific percents single, enrolled, and working, in SAVY versus in the 1989 and 1999 censuses. This bias should not have much effect on analysis of SAVY data, but must be kept in mind nevertheless.

2.7. Standard Errors

The particular sample used in SAVY is one of a large number of all possible samples of the same size that could have been selected using the same sample design. The particular value of the estimate – the point estimate – derived from each of the different samples would differ from each other. The deviation of a sample estimate from the average of all possible samples is called sampling error. While it is not possible to calculate the actual sampling error since we only have data from one of the possible samples, the standard error of a given estimate as calculated in this report is nevertheless an estimate of the sampling error. The estimated standard error also partially measures the effect of some non-sampling errors such as that which can be attributed to variability among interviewers and coders but does not measure any systematic biases in the data.

The point estimate from the sample for a given variable or indicator, and an estimate of its standard error, permit us to construct interval estimates with prescribed confidence that the interval includes the average result of all possible samples. To illustrate, if all possible samples were selected, each were surveyed under the same conditions and an indicator and its estimated standard error were calculated from each sample, then approximately 95% of the intervals from two standard errors below to two standard errors above the indicator would include the average value of all possible samples: the so-called 95% confidence interval. Details about standard errors can be found in the Appendix.

1. At the time of the sample selection there were 61 provinces; however at time of printing Viet Nam has 64 provinces.