

The Caucasus Research Resource Centers (CRRC)

Data Initiative 2006

Brief Description of the Survey

1. Background	1
2. Survey procedures and methodology	2
2.1. General	2
2.2. Sampling methodology: some details	3
3. Questionnaire design and field work	4
4. Data archiving and delivery	4
4.1. General procedures and coding	4
4.2. Weighting of data	5
4.3. Datasets and their availability	5

1. Background

The Data Initiative (DI) is a cross-border, comparative effort initiated by the Caucasus Research Resource Centers (CRRC) to assess the public opinion/attitudes and various socioeconomic and political developments in the South Caucasus (SC) region. The CRRC teams of Armenia, Azerbaijan and Georgia initiated coordinated data gathering efforts in the fall of 2003 to collect reliable data on the region.

- In 2004, the survey was conducted only in the capital cities. In total, 4,461 respondents were surveyed in Yerevan, Baku and Tbilisi. The data created by DI-2004 are representative at the level of capital cities of Armenia, Azerbaijan and Georgia.
- The DI of 2005 tended to be geographically more representative of the SC region. The centers collected data not only in the capitals, but also in one region in each country: Kotayk Marz (region) in Armenia, Shida Kartli region in Georgia, and Aran region - Mugan zone in Azerbaijan. In each country, half of 1,500 interviewed households were selected from the mentioned regions and the rest - from the capital cities. In the capital cities the panel datasets of respondents were created based on the DI-2004 respondents' lists. Thus, the data created by DI-2005 are representative at capital city level in each country and at the level of the mentioned three regions.
- In 2006, the centers set a goal to increase representativeness of collected data at the country level, and the SC region in general. Thus, the DI surveys were implemented in all regions of Armenia, Georgia and Azerbaijan. More than 2,000 households were surveyed in each country, representing both urban and rural areas. The data created by DI-2006 are representative both at the national level and at the level of the capital city and urban-rural areas in each country. In the capital cities the panel survey was carried out based on the DI-2004 and DI-2005 respondents' lists.

Thus, the CRRC DI database for 2004-2006 can allow researchers to:

- a) Analyze socioeconomic, demographic and political developments, as well as other trends during 2004-2006 in each SC country and the SC in general;
- b) Analyze current situation in each SC country, and
- c) Make cross-country analyses.

The table below provides a brief description of the survey in each country/year.

Table. Brief description of the CRRC DI surveys

Country\year		2004	2005	2006
Armenia	<i>Total number of respondents</i>	1,500	1,500 (750 + 750)	2,065 (715 + 1,350)
	<i>Location</i>	Yerevan	Yerevan + Kotayk Marz	Yerevan + all regions
	<i>Sampling base</i>	The households were randomly selected based on electricity users' lists. Electricity supply branches were used as a general frame for the sampling design, and 1,500 respondents were interviewed in the selected households.	Yerevan: 750 respondents interviewed in Yerevan were selected from the list of respondents of 2004 (each second respondent). Kotayk marz/region: 750 households were randomly selected (based on the lists of electors) and one respondent was interviewed in each household.	Yerevan: 715 respondents interviewed in Yerevan were selected from the lists of respondents surveyed during 2004-2005. All regions: 1,350 households were randomly selected in all 10 Armenian regions based on the lists of electricity users and one respondent was interviewed in each household.
Azerbaijan	<i>Total number of respondents</i>	1,489	1,500 (750 + 750)	2,400 (622 + 1,778)
	<i>Location</i>	Baku	Baku + Aran/Mugan	Baku + all regions
	<i>Sampling base</i>	The households were randomly selected based on census general frame, i.e. census district lists. 1,489 respondents were interviewed in the selected households.	Baku: 750 respondents interviewed in Baku were selected from the list of respondents of 2004. Aran region -Mugan zone: 750 households were randomly selected (based on census district lists) and 750 respondents in the selected households were interviewed.	Baku: 622 respondents interviewed in Baku were selected from the list of respondents surveyed during 2004-2005. All regions: 1,778 households were randomly selected in all 9 regions from the census district lists and one respondent was interviewed in each household.
Georgia	<i>Total number of respondents</i>	1,472	1,500 (750 + 750)	2,400 (600 + 1,800)
	<i>Location</i>	Tbilisi	Tbilisi + Shida Kartli	Tbilisi + all regions
	<i>Sampling base</i>	The households were randomly selected based on census general frame, i.e. census district lists. 1,472 respondents were interviewed in the selected households.	Tbilisi: 750 respondents interviewed in Tbilisi were selected from the list of respondents of 2004 (each second respondent). Shida Kartli region: 750 households were randomly selected based on census district lists. 750 respondents were interviewed in the selected households.	Tbilisi: 600 respondents in Tbilisi were selected from the lists of respondents surveyed during 2004-2005. All regions: 1,800 households were randomly selected in all 9 regions based on census district lists and one respondent was interviewed in each household.

2. Survey procedures and methodology

2.1. General

Survey procedures and methodology were developed mostly during the first year of DI surveys. In 2004, the CRRC set up the following three formal working groups comprised of scholars from the

three countries: Sampling Working Group (SWG), Questionnaire Working Group (QWG) and Data Archiving Working Group (DAWG). On an operational level, working groups were formed from each country to address the following issues: unified sampling design, questionnaire design, interviewer training and field procedures, data coding and data archiving.

The CRRC DI survey was designed and implemented according to the internationally accepted standards. Particularly,

- Multistage cluster sampling was employed through stratification. The number of sampled primary sampling units (PSUs) was defined in proportion to the number of households within each assigned stratum. The number of interviews was determined according to the number of people within each assigned strata.
- Actualization (block listing) was carried out in each selected cluster.
- The second level sampling was designed and Final Sampling Units (respondents) were identified using a Kish Table.
- A unified questionnaire was developed by the international group of experts.
- Interviewers and supervisors were trained in administering the actual field work, presenting the goals and aims of the questionnaire in a consistent manner across countries, and conducting interviews with a minimum risk of introducing sources of bias or error.
- Qualified data entry operators were recruited and trained to support the data collection process. Finally, raw data were processed, cleaned and posted on Internet for use by researchers.

2.2. Sampling methodology: some details

2004: In June, 2004, approximately 1,500 households in each capital city were randomly selected for interview using a carefully planned set of statistically reliable procedures. Multistage cluster sampling was employed through stratification. It consisted of three main phases: first level sampling, actualization, and second level sampling. The Primary Sampling Units in Baku and Tbilisi were census districts, while in Yerevan those were electricity supply branches as the information on census districts was not accessible. The Secondary Sampling Units were households, and the Final Sampling Units - respondents. Sampling within each household was implemented using the Kish Table. Only household members aged 18 years and over were eligible to be interviewed.

2005: During the first level sampling in 2005, a panel survey was designed for the capital cities of Armenia, Georgia and Azerbaijan based on the list of randomly selected households identified within the framework of 2004 DI program. 750 households in each capital were selected from the list developed in 2004. In addition, approximately 750 households in each targeted region of Armenia, Georgia and Azerbaijan were randomly-selected for interviews based on multistage cluster sampling, and considering the proportions of urban/rural populations. The selection of households in Shida Kartli region (Georgia) and in Aran region - Mugan zone (Azerbaijan) were implemented based on census districts, while the respondents in Kotayk region (Armenia) were selected from the lists of electors.

2006: During DI 2006, a panel survey was implemented in the capital cities of Armenia, Georgia and Azerbaijan based on the lists of randomly-selected households identified within the framework of DI 2004-2005. More than 600 households were surveyed in Yerevan, Tbilisi and Baku. In addition, more than 1,700 households were surveyed in Georgian and Azerbaijani regions based on census districts, and 1,350 randomly selected households were interviewed in the Armenian regions based on electricity users' lists.

Stratified cluster sampling using proportional stratification techniques was implied during the DI 2006 survey for the regions. Three levels of stratification were applied. At the first level capital

cities, urban (excluding capitals) and rural areas were considered as strata. Thus, the samples are representative for each of the aforementioned strata and the proportional stratification is ensured. At the second level of stratification the regions in the three countries were considered as strata and proportional stratification at regional level and urban-rural areas in each region was applied. At the third level, the respondents from regions were selected (one from each household) based on Kish tables.

3. Questionnaire design and field work

The survey instrument included a questionnaire which consisted of more than 120 questions structured into the following blocks:

- **General description of households** (number of its members, their relation to the chief of household /HH/, gender, date of birth, education)
- **Demographic data** of the respondents (marital status, ethnicity, citizenship, occupation, sphere of employment, etc.)
- **Education** (enrollment in educational programs, source of financing for education, perspectives and factors for finding a corresponding job, etc.)
- **Migration** (number of HH members emigrated during past three years, destination countries, residence status, reasons for emigration, employment spheres, etc.)
- **Health** (smoking and alcohol use habits, knowledge on Sexually Transmitted Infections and AIDS, and the ways for their transmission; attitude towards people living with AIDS, drug addicts and homosexuals; respondents' opinion on drinking/smoking/sexual behavior; experience and reasons for seeing a doctor; etc.)
- **Political behavior/outlook** (interest in politics, sources of information, participation in political elections, assessment of general politics in the country, public/social policy priorities, co-operation with neighboring countries; Armenia's/Georgia's/Azerbaijan's relations with NATO and various ethnic minorities)
- **Social Institutes**: respondents' trust towards various social institutes (parliament, police, courts, church, mass media, president, etc.); assessment of effectiveness of various organizations (UN, the World Bank, OSCE, USAID, CIS, etc.); affiliation with political parties, religions and NGOs, etc.
- **Crime**: respondents' sense of security at home, work place, in the street; self assessment of the risks for being involved in crime as victims; experience in crime, or other forms of crime, etc.
- **Economic status**: possession of certain household items, HH budget generation, sources of income, HH expenditures, experience with credits/loans, self assessment of HH socioeconomic status and its dynamics over the past years, projections for future.

The main mode of data collection has been face-to-face in-home interviewing, which lasted 30-40 minutes in average. The response rate was about 80 percent in all regions/years. To increase representativeness of collected data at the country level, the collected data for households and HH members were weighted in accordance with the proportions implied during the stratification and clusterisation.

4. Data archiving and delivery

4.1. General procedures and coding

CRRC DI field supervisors coordinated the collection of data in the regions. After the required number of interviews were completed, the answers including those to the open ended questions in the questionnaires were entered and coded in coordination with all three CRRC offices. The

majority of coding of verbatim responses was carried out by field supervisors (coders) post-field. Coded items included languages spoken at home, citizenship, countries and reasons for relocation; reasons for studying or finding a decent job; countries for emigration, reasons for not participating in parliamentary/presidential elections; countries that Armenia/Georgia/Azerbaijan needs to cooperate with in various spheres; political/religious affiliation, nationality, occupation and industry, reasons for not working, etc. Where no coding frame was available, verbatim listings of the responses were made and the frame developed.

4.2. Weighting of data

In order to be able to generalize the survey results for the whole country (-ies), the CRRC dataset contains (longitudinal) weights for each wave of data. In general, there are separate weights for respondents and for households.

For the weighting of data in accordance with the proportions implied during the stratification and clusterization, the following steps were carried out:

- The weight of each actualized household in each stratum was calculated;
- Then, the weight of each sampled household in each sampled cluster of each stratum was calculated.
- Taking into account that the sample frame in Armenia was based on the electricity consumers' list, which deviated to some extent from the numbers of households identified by the census, an adjustment of weights was made in order to have more accurate figures.
- In order to obtain the weights for the respondents, a similar procedure was implied taking into account the adult population size in each actualized cluster.
- As the sample frame in Armenia was based on the electricity consumers' list, which slightly deviated from the population size (and therefore, adult population) according to the census, an adjustment of weights was made in order to have more accurate figures.

4.3. Datasets and their availability

After the data enterers finished the data entry process in each country/year, the data archiving experts merged the datasets of year 2006, then a combined 2004-2006 regional database was produced in SPSS format. To make it easier to analyze the data on demography and migration blocks, as well as on some questions regarding the income of HH members, the data archiving experts further transposed the SPSS datasets into two types: one for HH members and one for the respondents. All databases contain variables with the weights of households, respondents and household members.

The survey datasets (in SPSS format), as well as the corresponding code books, data analysis guides, the questionnaire and the detailed description of survey methodology are available online at www.crrccenters.org and at the web pages of each center (www.crrc.am, www.crrc.ge, www.crrc.az) for use by social science researchers and the public at large locally and globally.

As a complementary to this program, CRRC awards fellowships and provides training and guidance to local researchers for analyzing the survey data.