



Caucasus Research Resource Centers

A Program of the Eurasia Foundation

Data Initiative 2007

UNIFIED SAMPLING METHODOLOGY

BACKGROUND	1
GENERAL PRINCIPLES	3
SAMPLING FRAMES	3
The Case of Armenia	5
The Case of Azerbaijan	6
The Case of Georgia	7
SAMPLING STRATEGY: AN OVERVIEW	7
DETAILED SAMPLING METHODOLOGY	9
Step 1: Stratification on the macro-level and calculation of the initial sample size	9
Step 2: Calculation of Sample size with DEFF and non-response rate adjustments	9
<i>Design Effect (DEFF) calculations</i>	10
<i>Estimation of Non-Response Rates</i>	11
Step 3: Calculation of the Final Sample Size	13
Step 4: Stratification	14
Step 5: Clusterization	14
<i>The case of Armenia</i>	15
Step 6: Final estimation of sample size by strata	15
USM STRENGTHS:	17
USM LIMITATIONS:	17

List of Tables

Table 1: Brief Description of CRRC DI Surveys, 2004-2006	2
Table 2: List of the Strata in each country	8
Table 3: Calculation of the Initial Sample Size for each stratum and for the whole country (without DEFF)	9
Table 4: Variables from DI 2006 database to estimate DEFF for DI 2007.....	10
Table 5: Estimation of DEFF for each country (based on aggregated clusters)	11
Table 6: Estimated non-response rates for DI 2007 in Armenia	12
Table 7: Estimated non-response rates for DI 2007 in Azerbaijan	12
Table 8: Estimated non-response rates for DI 2007 in Georgia	13
Table 9: Final Sample Size for Armenia	13
Table 10: Final Sample Size for Azerbaijan	13
Table 11: Final Sample Size for Georgia	13
Table 12: Geographical Stratification by Quadrants	14
Table 13: Total Number of Aggregated PSUs	14
Table 14: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Armenia	16
Table 15: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Azerbaijan	16
Table 16: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Georgia	16

Background

The Data Initiative (DI) is a cross-border effort initiated by the Caucasus Research Resource Centers (CRRC) to collect data on a wide variety of social, political and economic indicators in the South Caucasus (SC) region.

The DI is part of a comprehensive capacity building effort across the South Caucasus to increase quantitative methodological skills in the social sciences. It therefore involves both international and local expertise, but is committed to developing local capacity to carry out high level survey work.

Furthermore, while CRRC strives to use the DI a tool for collecting useful information on the region, it also sees the DI as a tool to experiment with different methodological approaches in terms of areas such as sampling and question design. Since we are constantly aiming to improve what we do, we highly value user feedback.

The CRRC teams in Armenia, Azerbaijan and Georgia began to collect data on the region in the fall of 2003. CRRC carried out the first survey in 2004.

- In 2004, the survey was conducted only in the capital cities of the South Caucasus. In total, 4,461 respondents were surveyed in Yerevan, Baku and Tbilisi. The data generated by DI-2004 are representative at the level of capital cities of Armenia, Azerbaijan and Georgia.
- In 2005, CRRC expanded its efforts to collect data not only in the capitals, but also in one region in each country: the Kotayk region in Armenia, the Aran region - Mugan zone in Azerbaijan, and the Shida Kartli region in Georgia. In each country, half of the 1,500 interviewed households were selected from the abovementioned regions and the other half from the capital cities. In the capital cities 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 three abovementioned regions.
- In 2006, the centers increased the representativeness of the collected data. The DI survey was implemented in all regions of Armenia, Azerbaijan and Georgia controlled by the central government (with the exception of Naxcivan in Azerbaijan). The conflict zones not under control of the central government (Nagorno-Karabakh, Abkhazia and South Ossetia were not surveyed). More than 2,000 households were surveyed in each country, representing both urban and rural areas. The data generated by DI-2006 are representative at the national level, the level of the capital city and the level of urban-rural settlement types in each country. In the capital cities, the panel survey was carried out based on the DI-2004 and DI-2005 respondents' lists.

The CRRC DI database for 2004-2006 allows researchers to analyze:

- a) longitudinal trends within and among SC countries;
- b) the current situation within and among SC countries.

The **Table 1** below provides a brief description of the survey in each country from 2004-2006.

Table 1: Brief Description of CRRC DI Surveys, 2004-2006

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	Yerevan + all regions
	<i>Sampling base</i>	The households were randomly selected based on the electricity users' lists. Electricity supply branches were used as the general frame for the sampling design, and 1,500 respondents (one in each household) were interviewed in the selected households.	Yerevan: The 750 respondents interviewed in Yerevan were selected from the list of respondents from 2004 (each second respondent). Kotayk region: 750 households were randomly selected (based on the voter lists), and one respondent was interviewed in each household.	Yerevan: The 715 respondents interviewed in Yerevan were selected from the lists of respondents surveyed in 2004-2005. All regions: 1,350 households were randomly selected in all ten 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 the census general frame, i.e. census district lists. 1,489 respondents (one in each household) were interviewed in the selected households.	Baku: The 750 respondents interviewed in Baku were selected from the list of respondents from 2004. Aran region - Mugan zone: 750 households were randomly selected (based on census district lists), and one respondent was interviewed in each household.	Baku: 622 respondents interviewed in Baku were selected from the list of respondents surveyed in 2004-2005. All regions: 1,778 households were randomly selected in all nine 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 the census general frame, i.e. census district lists. 1,472 respondents were interviewed in the selected households.	Tbilisi: The 750 respondents interviewed in Tbilisi were selected from the list of respondents of 2004 (each second respondent). Shida Kartli region: The 750 households were randomly selected based on census district lists, and one respondent was interviewed in each household.	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 ten regions based on the census district lists. One respondent was interviewed in each household.

General Principles

Considering the differences in sampling approaches used in the previous years, the Centers set a goal to unify the sampling methodology for the DI survey in the three countries in 2007, which would allow having more comparable data in the South Caucasus. With this in mind, CRRC set out to create a sampling methodology for 2007 fieldwork in Armenia, Azerbaijan and Georgia based on a more unified sampling methodology.

In March and April, 2007, the representatives of the three CRRC centers, along with sampling experts met in Tbilisi to discuss the possibilities of having a unified sampling methodology in 2007; to revise the survey instrument (questionnaire)¹, and to introduce ways to contain costs associated with the DI survey. The following changes were made to the DI in 2007 in order to create a unified sample in line, to the extent possible, with international best practices. The following macro-level decisions were made:

1. To abolish the panel methodology, since it had only partially been implemented;
2. To remove reserve lists and increase sample size to take into account non-response;
3. To aggregate and use larger primary sampling units (PSUs)² in 2007, which would facilitate fieldwork and would also allow saving money for the travel between clusters in the regions of those countries;
4. To implement much more rigorous block-listing to update the Primary Sampling Units (where possible);³
5. To select respondents in the household based on "last birthday" method instead of Kish Table.

The Unified sampling methodology for the DI survey in 2007 is presented in this document.

Sampling Frames

CRRC aims to survey the entire *de facto population* of the South Caucasus, who possess a permanent residence. Such a survey excludes homeless people, active military officers and others who do not have a permanent address.

CRRC were advised to use Census tracts (if available) for selecting PSUs to be included in the survey. Census tracts provide information on the lists of household addresses and should be accompanied by maps which delineate the exact borders (street names in the cities, and notes on the geographic maps in rural areas) of each tract.

¹ DI 2006 questionnaire contained several sections of questions (Migration, Economic status, etc.) that did not provide reliable data. In addition to extensively editing these sections, the Centers worked with several partner organizations, including the International School of Economics at Tbilisi State University (ISET), who through a World Bank grant co-funded part of the DI related to economics.

²

³ Block listing was only implemented in Georgia.

However, Census tracts maps are not available to CRRC in all South Caucasus countries. Therefore, during the March 30–April 2, 2007 DI meeting the DI sampling working group (SWG) experts, headed by Dr. Cynthia Buckley, decided to obtain detailed maps, where possible, of the SC countries (including city blue prints) and use these together with the lists of HH addresses to implement a block listing procedure in order to reconstruct the borders of each census tract. This would allow receiving more accurate and updated lists of HHs living in the selected clusters for further sampling purposes.

The Case of Armenia

CRRC-Armenia has been requesting the Census tracts from Armenian National Statistical Service (NSS) since 2004. However, the NSS has, to date, refused to provide the requested data “due to the confidential nature of the requested information,” as written in their response letters.

Since Armenia could not obtain the required information from the NSS, CRRC-Armenia considered using the voters’ lists that have been available on the website of Armenian Central Electoral Commission (<http://www.elections.am/?lan=e>) since 2005. CRRC-Armenia piloted this sampling frame during the DI 2005 survey in Kotayk region/*marz* of Armenia. As the voters’ lists include lists of adults, and the DI survey sampling is based on households, CRRC-Armenia made substantial efforts to aggregate individual voters into households unites based on those lists of individuals.

However, despite the amount of efforts required to reformat the database, the voters’ lists in Armenia, the results list was still felt to be highly inaccurate (even after the Parliamentary elections of 2007) as the list contains names of migrants, and even names of the dead. Indeed, since the CRRC survey is interested in *de facto* rather than *de jure* household, the voter’s lists proved problematic.

Therefore, for 2007, CRRC-Armenia decided to continue using the database of electricity users, which had been used in the past. An updated version of this database was provided by Armenian Electricity Networks Company⁴ (AENC) in April 2007. This database includes lists of electricity counters with the names and addresses of corresponding users – heads of HHs. According to Armenia Demographic and Health Survey (DHS) data of 2005 (see www.measuredhs.com), 99.9% of HHs in Armenia have access to electricity supply, and therefore, the database of electricity consumers provides a comprehensive list of households in Armenia.

The AENC database of individual electricity users covers entire country, excluding large factories and enterprises and is located in one Excel file with 53 worksheets that represent corresponding electricity sub-networks in the country. 36 of these sub-networks coincide with the former Soviet administrative regional divisions of the country (excluding Yerevan). Some regions have additional sub-networks, which represent former Soviet administrative divisions, which were later merged. Additionally, Yerevan has ten sub-networks. (See **Table A1** in the Appendix for more details.)

After Armenian independence, the new government aggregated the smaller former Soviet regions (*rayon*), upon which the electricity sub-networks were based into larger districts called *marzes*. However, the larger *marzes* maintained the former administrative boundaries of the Soviet regions. Each sub-network, therefore, lies entirely within one *marz*, region, which makes stratification along *marz* boundaries easily feasible using the electricity grid.

⁴ A privately held company owns the electricity provision in Armenia (the firm has a 90% stake owned by the Russian state electricity company).

The variables (columns) of the AENC database include the following variables:

- a. Electricity branch name
- b. Electricity sub-network name (name of former Soviet administrative district)
- c. Transformer #
- d. Transformer name
- e. Transformer street address
- f. Transformer city/village
- g. Account Number of electricity user
- h. User's last name, first name.
- i. Street address (or village name in the villages).
- j. Notes that aid in cleaning the database, (i.e., A=garage, *Pomp* = water pump)

Based on this last information, CRRC-Armenia can divide the database by urban/rural types of settlements based on the unified methodology since the database contains the name of the settlement. Each settlement in Armenia is coded by the government as either urban or rural. CRRC, regionally, follows the governmental designation of population points as either urban or rural.

Although AENC is a database of HHs who are electricity consumers, it also contains information on electricity counters designed for apartment building elevators, lights and garages of some households, as seen in Variable J above. These counters comprise less than 3% of cases. However, based on Variable J, the CRRC-Armenia sampling expert designed a program in the corresponding database (in Microsoft Excel format) to clean the database of all counters that were not associated with residences, such as garages, water pumps, barns and elevators.

In previous years, CRRC-Armenia had utilized a form of updating the electricity lists to determine whether they were still accurate. In 2006, they only found 4% error in the cleaned data. Therefore, it was decided not to perform any form of block listing in the 2007 in Armenia, since the lists had already proved highly accurate.

The Case of Azerbaijan

CRRC-Azerbaijan, similar to CRRC-Georgia, chose to use the Census data from State Statistical Committee (SSC) for implementing the DI 2007 survey in Azerbaijan. Last Census was conducted in Azerbaijan in 1999. Although this data was outdated, another database for sampling frame was not available.

Average size of Instructor Areas (IA) during the Census was 600; these constituted DI 2007 PSUs (clusters). However, according to the SSC of Azerbaijan, the Census tract maps are not available in the country. Therefore, CRRC-Azerbaijan sampling experts selected PSUs (IA) throughout the country based on calculations presented later in this document. CRRC-Azerbaijan was able to obtain the list of HHs in the IAs which was sampled from SSC. Although it was highly

desirable to conduct block listing in Azerbaijan, due to outdated and hence less reliable Census information, practically it was not possible to implement. Hence, random selection of HHs in each of the selected IAs was performed based on the original data from the SSC.

Furthermore, because of political infeasibility, Naxcivan was removed from the study and all calculations of sample size.

The Case of Georgia

As in the case of Azerbaijan, sampling frame of DI 2007 survey in Georgia were results of 2002 Census. The average number of HHs in aggregated clusters in Georgia was 400. The Georgian SWG expert compiled a list of all IAs for the DI 2007 and selected a sample of these based on the principles outlined below. Lists of HH addresses for these PSUs were obtained from the SDS.

According to the information provided by the State Department for Statistics (SDS) of Georgia, Census tract maps consisting of SDS 2002 Instructor Areas (IA) are not available. CRRC-Georgia obtained maps of the selected PSUs from GeoScope, a private mapping company. The two documents were jointly used for block-listing (see Block Listing Instructions for more information) which was piloted in June, 2007 and implemented countrywide in August, 2007. As a result of block-listing, lists of HH addresses per each cluster were updated, and 50 HHs in each PSU were randomly selected to be interviewed (see more below).

Sampling Strategy: an Overview

As mentioned above, the DI 2007 sampling design is based on multistage cluster sampling with preliminary stratification. In this section, an overview of the sampling strategy and steps of the implemented sampling design is presented (to be described in detail in the following sections).

The question of geographical stratification needed to be determined for the DI 2007. As it was agreed, sampling in each region of the countries, which had been done in 2006, did not make sense, given that there was no statistical significance to the sample in any one region. Poverty levels were considered, but reported poverty levels among administrative units were too tightly clustered to generate meaningful strata. As regional differences may play a role in determining variation in economic and social conditions, particularly in Georgian and Azerbaijan, directional quadrants were discussed as a viable stratification method and adopted.

Hence, all countries have been stratified by three macro-strata, Urban, Rural and Capital, followed by division of each of the countries into four directional quadrants: North-East, North-West, South-East, and South-West. The capital cities in each country represent a separate stratum, hence, there are total of 9 strata in each country.

Table 2: List of the Strata in each country

Northwest – Urban	Northeast – Urban
Northwest – Rural	Northeast – Rural
Southwest – Urban	Southwest – Urban
Southwest – Rural	Southeast –Rural
Capital	

Based on the sampling frames, the countries have been divided into N number of primary PSUs, with the average PSU size of approximately 500 households in Armenia, 400 in Georgia and 500 in Azerbaijan. On average, 50 households were randomly sampled in each PSU (regardless of the actual PSU size) for an interview. Fifty households were selected in order better calculate inter and intra-cluster effects. The number of PSUs selected in each of the nine strata was proportionate to the total number households in each quadrant.

The initial sample size was calculated based on 95% Confidence Interval, 5% Error Margin (degree of precision), 0.5 population proportion.⁵ In addition, the Design Effects (DEFF) were estimated based on the variance of 6 variables from DI 2006 (with the exception of Azerbaijan, where the variable *p1_rec* was excluded from the DEFF calculations, as it has a large variation significantly affecting/increasing the sample size).

The Centers also decided to include the expected non-response rates in the final sample size calculation as opposed to the reserve lists⁶ used in the former years (**See Tables 6-8**).

The data gathered during the DI 2007 survey is representative at four levels (as it was during the DI 2006 survey):

1. Capital cities
2. Other urban areas (non-capital)
3. Rural areas
4. The national level

Although the subdivision of the countries into for geographical quadrants (plus the capital) is a more cost-effective approach (and reasonable in Azerbaijan and Georgia to account for geographic diversity⁷), the Centers (especially Armenia) did not target at providing representativeness at those quadrants' levels.

The final sampling units (FSUs) are households randomly selected in each PSU. In each selected household (HH), a Household Interview and Individual Interview were conducted. During the Household Interview, up to two competent adult household members were asked to provide data on their HH general characteristics and economic conditions and behavior. The respondent for the

⁵ This would guarantee that no variable variance is ignored in the study. The highest variance in the DI 2006 survey had the variable *gender* (about 50/50), which was decided to take as the population proportion for initial sample size calculation.

⁶ During the previous DI surveys the centers used lists of HHs randomly selected together with the main (required minimum number of HHs) lists of HHs to serve as reserve lists so that the HHs in the main lists could be replaced with others in those lists in case of non-responses.

⁷ This technique is less warranted in Armenia, as the country displays a high level of homogeneity in all of its geographic regions.

Individual interview was selected among adult HH members using the *last birthday* method; this interview collects information about social and political attitudes, as well as education and ethnic identity.

Detailed Sampling Methodology

Step 1: Stratification on the macro-level and calculation of the initial sample size

All countries have been stratified by three macro-strata: *Urban, Rural* and *Capital*. In none of the South Caucasus countries does type of settlement directly relate to the population size in this settlement. In Armenia and Azerbaijan, the division between urban and rural settlements is made by the government, and each settlement is clearly assigned to either rural or urban type. In Georgia, in addition to rural settlements (villages) and urban settlements (cities and towns), there is also third type of settlement, called *daba*, which officially is neither urban nor rural, but is rather considered somewhere in between these two types. For the DI 2007 sampling strategy, *dabas* are considered urban settlements because they have administrative functions characteristic of urban settlements (and are administrative centers of their respective *rayons*), which results in urban-type jobs available in these settlements (despite the fact that for the population of *dabas* agricultural occupations are characteristic more than for the population of towns and cities).

Although in all of the SC countries the government data also distinguishes between "small" vs. "large" within urban settlement, in terms of DI 2007 we do not apply this division; rather the DI aggregates large and small urban areas as it is only concerned with stratification by the three macro-strata: *Urban, Rural* and *Capital*.

The Initial (minimum) sample size for each stratum in each country is calculated based on 95% Confidence Interval (CI), 5% Error margin (degree of precision), and 0.5 population proportion⁸. **Table** below illustrates the calculation of this initial sample size for each macro-stratum, as well as for the whole country; calculations are the same for each country.

Table 3: Calculation of the Initial Sample Size for each stratum and for the whole country (without DEFF)

Macro-strata	Half-Width Desired	Population Proportion	Desired Confidence Level	Sample size (without DEFF)
Capital	0.05	0.5	95%	385
Urban	0.05	0.5	95%	385
Rural	0.05	0.5	95%	385
<i>Total:</i>				<i>1155</i>

Step 2: Calculation of Sample size with DEFF and non-response rate adjustments

⁸ This guarantees that no variable variance is ignored in the study. The highest variance in the DI 2006 survey had the variable *gender* (about 50/50), hence, it has been decided to take as the population proportion for initial sample size calculation.

Since the DI 2007 sampling methodology involves cluster sampling, it is necessary to calculate the Design Effects (DEFF) that represents the intra-cluster and inter-cluster variations of the key variables in the study. In addition, non-response rates should be estimated, and the initial sample size has to be recalculated taking into account these two factors. These calculations are presented in this step.

Design Effect (DEFF) calculations

The most reliable way to calculate Design Effects is to base these calculations on the data from the previous similar studies. Nevertheless, several other options for DEFF calculations were explored, given the change in methodology associated with the DI 2007. However, since no other similar survey could be located in all three of the SC countries, it was decided as the best viable option.

While targeted sample size per cluster in DI 2007 is 50, it was 6 to 15 (depending on the settlement type) in DI 2006. Hence, it was decided to aggregate DI 2006 clusters for DEFF calculation, so that number of interviews per aggregated cluster was approximately 50. On average, five neighboring clusters were aggregated, after which the following variables chosen by SWG experts from DI 2006 database were used to estimate the DEFF for DI 2007 in the three macro-strata (capitals, urban settlements and rural settlements). While this is not the optimal method for the calculation of DEFF, given the fact the PSUs aggregated from 2006 were not all contiguous, it was considered the best method given the current possibilities.

In order to estimate the design effects for the three types of settlements (strata), the mean of 6 variables (in case of Azerbaijan – the mean of 5 variables) in each type of settlement was used. Furthermore, cluster data had not properly been documented in 2006 in Baku, so no calculations about Design Effects could be made. Therefore, two was chosen. These variables are presented in **Table 4** below.

Table 4: Variables from DI 2006 database to estimate DEFF for DI 2007

<i>Variables</i>	<i>Descriptions</i>
resp_sex	Respondents' sex (male, female).
age60	Respondents' age, aggregated variable for two age groups (up to 60; 60+ years old).
Edu	Respondents' education, aggregated variable for two groups (complete secondary and lower; higher than secondary education).
h1	Question "Have you ever smoked regularly?" (Yes/No).
h11_a	Question "During the last 12 months, have you been sick and felt the need of a doctor?" (Yes/No).
p1_rec ⁹	Question "To what extent are you interested in politics?" aggregated into two groups (very/somewhat interested + refuse to answer; not very/not at all interested).

Estimations of design effects for each country are presented in the **Table 5** below. Stata commands for these calculations are available upon request.¹⁰

⁹ CRRC-Azerbaijan has not used variable *p1_rec*, as it has a large variation, which significantly increases the sample size (See DEFF calculations for Azerbaijan with variable *p1_rec* in Table A2 in the Appendix).

¹⁰ Stata commands for DEFF calculation in Armenia see in the Appendix.

Table 5: Estimation of DEFF for each country (based on aggregated clusters)

Armenia¹¹	Resp_sex	age60	edu	h1	h11_a	p1_rec	Average DEFF
Yerevan	2.06	2.04	3.40	1.06	1.41	1.50	1.9
Urban	1.39	1.35	1.93	1.65	2.77	3.50	2.1
Rural	1.69	1.41	1.83	2.04	2.80	4.66	2.4

Azerbaijan¹²	Resp_sex	age60	edu	h1	h11_a	p1_rec	Average DEFF
Baku	NA	NA	NA2	NA	NA	NA	2
Urban	1.36	1.89	2.48	1.59	5.16	NA	2.5
Rural	1.09	1.76	1.91	1.44	3.04	NA	1.8

Georgia¹³	Resp_sex	age60	edu	h1	h11_a	p1_rec	Average DEFF
Tbilisi	2.1	2.8	2.7	2.1	1.5	2.2	2.2
Urban	1.1	1.8	2.7	3.3	3.5	2.7	2.5
Rural	1.4	2.2	4.5	1.6	5.4	4.6	3.3

Estimation of Non-Response Rates

In the DI 2007 sampling strategy, expected non-response rates affect the calculation of the final sample size, as opposed to the reserve lists¹⁴ used in the previous years. Estimation of expected non-response rate is based on the data available in the given country. Namely, in Armenia, where there is respective data from DI 2006, estimation of expected non-response rate is based on this data. In Azerbaijan and Georgia, however, where such data is not available¹⁵, the estimation of non-response rates is based on the information provided by Departments of Statistics of these countries, with certain expert adjustments explained in this section. **Table 7 and 8** below illustrate the estimation of non-response rates in each country.

In Armenia, DI 2004 data was used for Yerevan (as it was the first year of the panel survey in the capitals) and DI 2006 data for the regions of the country. Data from other similar surveys (the Household Survey carried out by the Armenian National Statistical Service, WB and UNDP in 2002-2003) were also used to estimate non-response rates for each stratum. The following Table represents the estimation of non-response rates in Armenia for DI 2007 survey, as well as the

¹¹ CRRC-ARM calculated DEFFs based on DI 2006 database by merging 3 clusters in the regions and 6-8 clusters in Yerevan to have 45-50 respondents in each cluster.

¹² CRRC-AZE calculated DEFFs based on DI 2006 database aggregating 5-6 clusters in urban settlements and 3 clusters in rural settlements. Moreover, because of some coding errors made in DI 2006 survey in Baku city, the DEFF for this capital has been taken as theoretically justified average number – 2.

¹³ CRRC-GEO calculated DEFFs based on DI 2006 database by merging 4-6 clusters containing approximately 50 HHs in each.

¹⁴ In the previous DI surveys, the Centers used reserve lists of randomly selected HHs together with the main (required minimum number of HHs) lists of HHs, to serve as reserve lists so that the HHs in the main lists could be replaced with HHs from the additional lists in case of non-response. This technique brings in unwanted bias into the sample.

¹⁵ CRRC-Azerbaijan and CRRC-Georgia could not estimate the rate of non-response based on previous DIs, because the non-response information was no maintained. This error has been corrected in 2007 and full non-response rates for this will be coded and entered for better non-response rate calculations in the future.

formulas for the calculation of the sample size with the account for expected non-response rate in each stratum.

Table 6: Estimated non-response rates for DI 2007 in Armenia

	Estimated non-response rate	Multiplication factor¹⁶
Yerevan	30%	1.43 (1:0.7)
Urban	25%	1.33 (1:0.75)
Rural	20%	1.25 (1:0.8)

In Azerbaijan, the corresponding information has been provided by the SSC and is based on SSC Household Budget survey. Since SSC has conducted a panel survey the last two years, CRRC has taken the latest non-response rate from the 2005 SSC survey.

However, there are two significant differences between the procedures of the SSC survey and CRRC survey:

1. SSC pays 2 USD to each respondent per quarter to remain in the panel.
2. Members of local administration are actively involved in the process of conducting the survey.

None of these conditions were satisfied in terms of DI 2007. Considering these facts, an extra 6% has been added to the official non-response rates of the SSC survey in each macro-stratum. **Table 7** presents estimation of non-response rate in Azerbaijan.

Table 7: Estimated non-response rates for DI 2007 in Azerbaijan

	SSC non-response rate	Addition to SSC non-response rate	Final non-response rate	Multiplication factor
Baku	20%	6%	26%	1.35 (1:0.74)
Urban	8%	6%	14%	1.16 (1:0.86)
Rural	4%	6%	10%	1.11 (1:0.9)

In Georgia, results of the SDS ongoing Household Budget Surveys were used to estimate the non-response rate. Namely, average annual non-response rates for the three types of settlements have been calculated.

Extra 3% has been added to each of these rates, since SDS is paying the HHs for their participation in the survey. Expected levels of non-response in each macro-stratum are presented in the **Table 8** below:

¹⁶ If we are expecting, for example, 30% non-responses, then we need to multiply the initial sample size by 1.43, i.e. increase it by 43%.

Table 8: Estimated non-response rates for DI 2007 in Georgia

	SDS non-response rate	Addition to SDS non-response rate	Final non-response rates	Multiplication factor
Tbilisi	29%	3%	32%	1.47 (1:0.68)
Urban	23%	3%	26%	1.35 (1:0.74)
Rural	13%	3%	16%	1.19 (1:0.84)

Step 3: Calculation of the Final Sample Size

The final sample size in each stratum of each country has been calculated based on the initial sample size, with DEFF and non-response rate adjustments. The following Tables present results of these calculations:

Table 9: Final Sample Size for Armenia

Stratum	Half-Width Desired	Population Proportion	Desired Confidence Level	Sample size (without DEFF)	DEFF	Sample size (with DEFF)	Non-Resp. adj.	Multiplication factor	Sample size (adjusted)
Yerevan	0.05	0.5	95%	385	1.9	732	30%	1.43	1,046
Urban	0.05	0.5	95%	385	2.1	806	25%	1.33	1,075
Rural	0.05	0.5	95%	385	2.4	924	20%	1.25	1,155
Total:				1,155		2,462			3,276

Table 10: Final Sample Size for Azerbaijan

Stratum	Half-Width Desired	Population Proportion	Desired Confidence Level	Sample size (without DEFF)	DEFF	Sample size (with DEFF)	Non-Resp. adj.	Multiplication factor	Sample size (adjusted)
Baku	0.05	0.5	95%	385	2.0	778	26%	1.35	1040
Urban	0.05	0.5	95%	385	2.5	962	14%	1.16	1116
Rural	0.05	0.5	95%	385	1.8	712	10%	1.11	790
Total:				1,155		2,452			2,946

Table 11: Final Sample Size for Georgia

Stratum	Half-Width Desired	Population Proportion	Desired Confidence Level	Sample size (without DEFF)	DEFF	Sample size (with DEFF)	Non Resp. adj.	Multiplication factor	Sample size (adjusted)
Yerevan	0.05	0.5	95%	385	2.2	860	32%	1.47	1,268
Urban	0.05	0.5	95%	385	2.5	971	26%	1.35	1,314
Rural	0.05	0.5	95%	385	3.3	1264	16%	1.19	1,504
Total:				1,155		3,095			4,086

Step 4: Stratification

Table 12 illustrates the grouping of administrative regions of the SC countries into directional quadrants for the initial stratification.

Table 12: Geographical Stratification by Quadrants

Quadrants	Countries/ Administrative Regions		
	Armenia	Azerbaijan	Georgia
North-East	Gegharkunik	Guba-Khachmaz	Mtskheta – Mtianeti
	Kotayk	Dagliq Shirvan	Kakheti
	Tavush	Absheron	Shida Kartli
North-West	Aragatsotn	Shaki-Zagatala	Samegrelo – Zemo Svaneti
	Lori	Ganja-Gazakh	Imereti
	Shirak		Racha – Lechkhumi
South-East	Syunik	Lankaran-Astara	Kvemo Kartli
	Vayots Dzor	Aran	Samtskhe – Javakheti
South-West	Ararat	Kalbajar-Lachin	Adjara
	Armavir	Dagliq Garabaq	Guria
Capital	Yerevan	Baku	Tbilisi

As mentioned above, population in each of these quadrants has been divided into urban and rural; in addition, the capitals represent a separate stratum. Naxcivan is omitted from the study, as are all territories not under control of the central government.

Step 5: Clusterization

Clusterization of the sampling frame (formation of PSUs for sampling on the basis of the sampling frame) has been performed on the next stage. The distribution of the total number of clusters by strata/country is presented in **Table 1** below.

Table 13: Total Number of Aggregated PSUs

STRATA		Armenia	Azerbaijan	Georgia
Capital		588	1040	872
North-East	Urban	161	308	167
North-East	Rural	186	122	614
North-West	Urban	219	353	411
North-West	Rural	166	252	672
South-East	Urban	57	443	190
South-East	Rural	55	383	429
South-West	Urban	90	12	139
South-West	Rural	178	33	316
TOTAL:		1700	2946	3810

The case of Armenia

While in Azerbaijan and in Georgia the clusters were aggregated based on Census Instruction Areas, the clusters in Armenia are based on groups of households using the same or neighboring electricity transformation stations. There are 53 electricity networks in Armenia that include about 100-400 transformation stations in each, and each transformation station provides electricity to up to 700 households (electricity users) in its neighborhood, with the average of 150-200 households/users in urban areas and 50-100 HHs in rural areas. The clusterization for DI 2007 survey in Armenia has been implemented the following way:

- a) Rural areas with less than 400 households were combined with their neighboring villages - to form aggregated clusters with 400-600 households in each;
- b) Rural areas with more than 600 households were split into smaller clusters (based on proximity of transformation stations) to have 400-600 households in each;
- c) Urban areas were split by transformation stations. Clusters consisting 500 households in average were formed based on geographical proximity of transformation stations;
- d) Yerevan was split into 10 electricity networks serving its various districts first. Then each of them will be split into transformation stations. Clusters consisting of 500 households in average will be formed in Yerevan based on geographical proximity of those transformation stations.

Step 6: Final estimation of sample size by strata

The final sample size for each settlement type (stratum) was divided between 4 quadrants and the capital proportionally to their sizes/shares in the total number of urban/rural households in the country.

The number of clusters required for sampling in each stratum was calculated by adjusting the final sample size to the strata proportionate to their share (PPS) in overall population. As mentioned above and agreed earlier, the average number of interviews per cluster should be 50. Thus, in order to get the number of clusters to be selected per strata, the total number of respondents to be interviewed in each stratum has been divided by 50. Afterwards, final correction of sample size per strata has been made.

Tables 14 through 16 illustrate the allocation of the final sample size of each stratum to the corresponding quadrants/strata. As a result, DI 2007 sample for Armenia, for example, has been distributed into 64 clusters, with a total sample size of 3200 interviews. Taking into consideration expected non-response rates, 2400-2500 interviews are expected to be completed.

Table 14: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Armenia¹⁷

STRATA	Settl. type	Total N of CLUSTERS	Sample size (including Non-resp.)	Average sample size in PSUs (cluster)	Number of PSUs to be sampled	FINAL SAMPLE SIZE
Yerevan	Capital	588	1133	50	23	1150
North-East	Urban	161	310	50	6	300
North-East	Rural	186	358	50	7	350
North-West	Urban	219	422	50	8	400
North-West	Rural	166	320	50	6	300
South-East	Urban	57	110	50	2	100
South-East	Rural	55	106	50	2	100
South-West	Urban	90	173	50	3	150
South-West	Rural	178	343	50	7	350
TOTAL:		1,700	3,276		64	3,200

Table 15: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Azerbaijan

STRATA	Settl. type	Sample size	Average sample size in PSUs (cluster)	Number of PSUs to be sampled	FINAL SAMPLE SIZE
Baku	Urban	1040	50	21	1050
North-West	Urban	353	50	7	350
North-West	Rural	252	50	5	250
North-East	Urban	308	50	6	300
North-East	Rural	122	50	2	100
South-West	Urban	12	50	1	50
South-West	Rural	33	50	1	50
South-East	Urban	443	50	9	450
South-East	Rural	383	50	8	400
Total:		2946		60	3000

Table 16: Allocation of the final sample size of each stratum to the corresponding quadrants/strata in Georgia

STRATA	Settl. type	Sample size	Average sample size in PSU	Number of PSUs to be sampled	FINAL SAMPLE SIZE
Tbilisi	Urban	1268	50	25	1250
North-East	Urban	249	50	5	250
North-East	Rural	466	50	9	450
North-West	Urban	550	50	11	550
North-West	Rural	546	50	11	550
South-East	Urban	274	50	5	250
South-East	Rural	290	50	6	300
South-West	Urban	241	50	5	250
South-West	Rural	202	50	4	200
Total:		4086		81	4050

¹⁷ These numbers are based on cleaned database of Electricity users in Armenia as of April, 2007.

USM Strengths:

- Unified approach/strategy in all three countries;
- Use of the most reliable sources as sampling frames;
- Updated original data whenever possible;
- Representativeness at the level of countries, rural settlements, urban settlements, capitals;

USM Limitations:

- Unavailability of exactly the same sampling frames in all countries;
- Difficulties in aggregation of DI 2006 PSUs/clusters in Azerbaijan;
- Impossibility of conducting block listing in Azerbaijan;

Again, we would like to repeat that CRRC strives to use the DI a tool for collecting useful information on the region but also as a tool to experiment with different methodological approaches in terms of areas such as sampling and question design. We are constantly trying to improve what we do, so if you have any questions or concerns, please do not hesitate to contact us.

Appendices

Table A1: Armenian Quadrants by Electricity Branches, Electricity Networks, former Regions, and current Marzes

Quadrants	EI.Branch	EI.Network	ex.Region	Marz
North-East	Aghstev	Berd	Smashadin	Tavush
	Aghstev	Dilijan w/ ljevan	ljevan	Tavush
	Aghstev	ljevan	ljevan	Tavush
	Aghstev	Noyemberyan	Noyemberyan	Tavush
	Geghama	Abovyan	Abovyan	Kotayk
	Geghama	Arzni (w/ Abovian)	Abovyan	Kotayk
	Geghama	Chambarak	Krasnoselsk	Gegharkunik
	Geghama	Charentsavan w/ Hrazdan	Hrazdan	Kotayk
	Geghama	Gavar	Kamo	Gegharkunik
	Geghama	Hrazdan	Hrazdan	Kotayk
	Geghama	Martuni	Martuni	Gegharkunik
	Geghama	Nairi	Nairi	Kotayk
	Geghama	Sevan	Sevan	Gegharkunik
	Geghama	Vardenis	Vardenis	Gegharkunik
North-West	Amberd	Aparan	Aparan	Aragatsotn
	Amberd	Aragats	Aragats	Aragatsotn
	Amberd	Ashtarak	Ashtarak	Aragatsotn
	Amberd	Byurakan w/ Ashtarak	Ashtarak	Aragatsotn
	Debet	Gugark	Gugark	Lori
	Debet	Spitak	Spitak	Lori
	Debet	Stepanavan	Stepanavan	Lori
	Debet	Tashir	Kalinino	Lori
	Debet	Tumanyan (Alaverdi, Tumanyan)	Tumanyan	Lori
	Debet	Vanadzor w/ Gugark	Gugark	Lori
	Ghars	Akhuryan	Akhuryan	Shirak
	Ghars	Amasia	Amasia	Shirak
	Ghars	Ani	Ani	Shirak
	Ghars	Artik	Artik	Shirak
	Ghars	Ashotsk	Ghukasyan	Shirak
	Ghars	Gyumri w/ Akhuryan	Akhuryan	Shirak
Ghars	Talin	Talin	Aragatsotn	
South-East	Araqs	Eghegnadzor	Eghegnadzor	Vayots Dzor
	Araqs	Vayk	Azizbekov	Vayots Dzor
	Tatev	Goris	Goris	Syunik
	Tatev	Kapan	Kapan	Syunik
	Tatev	Meghri	Meghri	Syunik
	Tatev	Qajaran w/ Kapan	Kapan	Syunik
	Tatev	Sisian	Sisian	Syunik
South-West	Araqs	Ararat	Ararat	Ararat
	Araqs	Artashat	Artashat	Ararat
	Araqs	Masis	Masis	Ararat
	Musaler	Armavir	Oktemberyan	Armavir
	Musaler	Baghramyan w/ Oktemberyan	Oktemberyan	Armavir
	Musaler	Ejmiatsin	Ejmiatsin	Armavir
	Musaler	Musaler Gorts. w/ Ejmiatsin	Ejmiatsin	Armavir
	Musaler	Sardarapat w/ Ejmiatsin, Armavir	Ejmiatsin	Armavir
Capital	North	Kentron		Yerevan
	North	Mashtots		Yerevan
	South	Arabkir		Yerevan
	South	Sari Tagh		Yerevan
	South	Shengavit		Yerevan
	West	Erebuni		Yerevan
	West	Gircaranayin		Yerevan
	Yerevan	Nor Nork		Yerevan