THE CRISIS MONITORING SURVEY: METHODOLOGY, SAMPLING AND RESPONSE RATES[1]

The report assesses the impact of the recession on household welfare and the effectiveness of Government policies in mitigating its costs on vulnerable households using a Crisis Monitoring Survey (see Box below).

Box A1.1 Bulgaria Crisis Monitoring Survey (CMS):
Questionnaire and key design parameters

- Three waves panel survey (every 6 months) to track the impact of the crisis over time. The baseline data will be collected during the first wave, in the field in February 2010. The other waves were fielded in October 2010 and February 2011.
- *The CMS is a focused, shorter multi-topic household survey.* The survey will collect information about household demographics (roster); labor market participation and earnings; housing; durables; access to and receipts of social protection programs; informal safety nets & remittances; other incomes; credit; self-reported impact of the crisis; coping and mitigation mechanisms.
- *Focused on how Bulgarian households are affected and are coping with the crisis.* Through this survey, we try to map the coping strategies used by households in Bulgaria, such as: increase in labor supply; informal employment; reduction in expenditures; postponement of investments; sell of assets; reliance on formal/informal credit.
- *Checking whether the social protection system offers adequate protection to the poorest during crisis time[1].* The survey documents whether the unemployment benefits and the income-tested program work as automatic stabilizers (increase their coverage during crisis time); to identify whether income-eligible households are denied access to the GMI or heating allowances on other grounds than income (for example, because of the 12 months time limit for the GMI, or the requirement to register first as unemployment for 9 months before being able to apply for the GMI program); and to identify whether there are arrears in payments, or other factors that postpone payment or delay the eligibility determination process.
- *Dis-aggregated by different socio-economic groups.* The survey will allow to disaggregate the results across the *income* distribution (focus on income[1] and assets ownership, no consumption); ethic majority vs. minorities (including a Roma booster).
- *Sample size.* The survey aims for a sample size of 2400 households, with a booster of 300 households for the Roma minority.

The empirical instrument used to track the effects of the crisis on households was a Crisis Monitoring Survey, implemented jointly by the OSI Sofia and the World Bank (CMS). The CMS aims (i) to influence policy making with "real time" information; (ii) to understand the transmission channels of the crisis (e.g. via labor markets, financial markets; product markets; government services and transfers); (iii) to measure the impact of the recession on household welfare and HD outcomes (to go beyond administrative data, or impact on the "average

household"), across the income/welfare distribution; (iv) to document and describe the coping strategies of the households affected by the crisis; and (v) to determine whether the formal and informal safety nets are effectively mitigating the impact of the crisis. In particular, the survey will investigate in-depth how households are using the labor market to mitigate the impact of the crisis; whether the formal social protection programs, including those intended to act as automatic stabilizers, are protecting households against sliding into poverty; and the effectiveness of the informal safety nets. The CMS does not duplicate the Government's own efforts and systems to monitor the impact of the recession on household welfare, but complements them.

The value added of the CMS for the Government. The objective of the CMS is to produce just-in time information to feed in the policy making cycle on who is most affected by the crisis in a broad range of outcomes across the income distribution, and to profile the coping and mitigation mechanisms used by different types of households. In particular, the survey will investigate in depth how households are using the labor market to mitigate the impact of the crisis; whether the formal social protection programs, including those intended to act as automatic stabilizers, are protecting households against sliding into poverty; and the effectiveness of the informal safety nets. The CMS does not duplicate the Government's own efforts and systems to monitor the impact of the recession on household welfare, but complements them.

Box A1.2: Crisis Monitoring Surveys

Crisis Monitoring Surveys around the Region

Countries in Eastern and Southeastern Europe have been particularly affected by the global economic crisis. In response to crisis, governments, donors, and civil society organizations have been cooperating to conduct Crisis Monitoring Surveys that assess social and economic impacts of the crisis on households and individuals. Crisis Monitoring Surveys rely on modules that are specifically tailored to assess household circumstances in a crisis and provide real-time, nationally representative data that can inform policy.

Many countries in the region have well-developed data collection systems that include administrative data, labor force surveys, and household budget surveys. Crisis Monitoring Surveys, however, provide a unique instrument to assess:

- *The main transmission channels* of the crisis (such as labor markets, access to credit, remittances/informal transfers, relative prices, government services) through which it affects income. In particular, the Crisis Monitoring Survey labor modules are uniquely designed to capture how the crisis affects households in the labor market.
- *The mitigation strategies* adopted by households and the extent to which they negatively affect welfare and increase vulnerability including indebtedness
- *The existing safety nets* (including government assistance and informal transfers) and the extent to which they respond to the crisis.
- *The impacts on welfare* that cannot be quantified through data, including: reduced expenditures on health, education, and food

Stand-alone Crisis Monitoring Surveys are being (or already have been) conducted in: Turkey, Montenegro, Romania, Armenia, Georgia, Bulgaria and Tajikistan. A number of countries (Latvia, Croatia, Serbia, and Armenia) have also included Crisis Monitoring Modules into routinely conducted Labor Force or Household Budget Surveys. In Romania, Serbia, and Turkey, representative household surveys have been combined with non-representative qualitative data collection to understand the effect of the crisis on vulnerable groups, including Roma.

Bulgaria Crisis Monitoring Survey

In Bulgaria, the World Bank partnered with the Open Society Institute (OSI) Sofia to conduct a Crisis Monitoring Survey in three waves: February 2010, September 2010 and February 2011. The Bulgaria Crisis Survey combines modules that have been used in crisis surveys in a number of countries (including crisis-specific labor, credit, and coping strategies modules) with uniquely detailed modules on income and social assistance.

The survey questionnaire that was used to generate the data follows the philosophy of integrated household surveys. It consists of separate components containing both household and individual modules.

- The definition of "*household*" used in this survey is as follows: People who live under the same roof, share common budget and have at least one meal per day together.
- A "h*ousehold member*" is someone who has lived in the household at least three months in the last 12. This includes someone who has moved in within the last three months, not necessarily a usual member, but someone who lives under the same roof and shares the same

budget. Always considered as household members are the head of the household, newly married couples and babies.

- The *head of the household* is defined by the household members themselves.

A. Questionnaire design

The questionnaire consists of 10 modules.

- **Roster**—data was collected on individual level and each household member's profile was registered (demographic characteristics, marital status, education, health, ethnicity and religion).

- **Labor**—data was collected in two parts on individual level from each household member age 15 and over, who was asked whether he or she worked in the last 4 weeks. Those who had were asked about position, salary, working hours, social insurance, etc. Those who had not were asked why not, whether they had looked for a job, if yes, what kind of job, if no, why.

- **Housing**—data was collected on a household level and included questions about living conditions (type of dwelling, total area, availability of electricity, water supply, heating and so on), and whether there were any unpaid bills.

- **Assets**—data was collected on a household level and included questions about ownership of goods and durables (TV, stove, refrigerator, car, computer, telephone, etc.).

- **Social assistance and unemployment**—data was divided into three parts and collected information on both individual and household levels. All types of social assistance and other benefits (pensions, maternity benefits, scholarships) were registered.

- **Informal transfers and support**—the data was collected on a household level about all types of support (in money or in kind) for the last 12 months from the country and abroad.

- **Other income**—data was collected on a household level regarding annual additional income (for example, income from rent, agriculture, sales etc.)

- **Credit**—data was collected on a household level and detailed information about credits and mortgages was gathered.

- **Effects of crisis and coping strategies**—data was collected on a household level concerning how the household members were affected by the crisis (losing their jobs, unpaid salaries and so on) and ways to cope (reduced consumption of objects of common use).

- **Subjective questions**—data was collected on individual level and some economical and other attitude questions were asked (for example, "How would you describe the situation in the country at this moment?")

B. Sample design

- **Sample types and sample sizes**

The survey was planned and realized as a panel survey. Three waves were conducted—first one in February 2010, the second one in September 2010 and the third one in February 2011.
The universe under study consists of all households in Bulgaria (NSI, Census 2011, N=2,856,740). The survey used two different types of sampling strategies. The main one was nationally representative of households in Bulgaria. In the sample design plan, 2,400 households had to be interviewed. In each of the three waves this plan was realized as follows:

	households	household members
First wave (Feb. 2010)	2,384	6,653
Second wave (Sept. 2011)	2,298	6,225
Third wave (Feb. 2011)	2,329	6,180

The universe of the booster study consists of all the households in segregated settlements or areas of compact segregate population. The sample is representative for the households living in such compact areas. In the sample design plan, 300 households had to be interviewed. In each of the three waves this plan was realized as follows:

	households	household members
First wave (Feb. 2010)	296	1,099
Second wave (Sept. 2011)	293	1,115
Third wave (Feb. 2011)	296	1,077

- **Main sample design**

The main sample was created in two stages.

First, the population was stratified by *district* (NUTS 3) and *type of settlement*. In Bulgaria, there are 28 administrative districts. For the *type of settlement* three categories were defined—*rural*, *urban* (with population under 50,000) and *metropolitan* (with population over 50,000). Bulgaria's capital, Sofia, is include in the *metropolitan* category. In this way 28 x 3 = 84 categories (*strata*) were defined and proportional allocation was made, i.e. the number of households included in the sample from each stratum was calculated as a proportion of the total number of households in the stratum. The method of selecting settlements from each stratum is simple random sampling with replacement, weighted by the number of households in the settlement. Hence, some of the bigger cities could be drawn out more than once. For example Sofia was selected 40 times in the main sample.

The second stage chose voting stations in each settlement. In this methodology, voting stations were used as a type of cluster. In each cluster, 10 households had to be interviewed. In one voting station there could be put down various numbers of voters. So, voting stations were selected with probability proportional to the number of voters in each station. In each cluster, (voting station), 20 household addresses were selected randomly from the list of all addresses in

the station. The first 10 addresses are a kind of *main list*, i.e. they have to be visited obligatorily. If there is a refusal in some of this main list, it has to be replaced with an address from the *list of reserves* (the last 10 addresses).

There were some exceptional cases from this sampling rule. In seven (rural) settlements, there were no street names, no house names and no other way to identify the exact address of the residents. For those settlements and for those, from the booster, GPS sampling was used. This is a geographical method to choose the households, which have to be interviewed, randomly. The sampling model will be explained in details in the item for the booster.

Table 1. Main sample—number of clusters by districts (NUTS 3) and types of settlements

(NUTS 3)	Total number of clusters	Rural	Urban	Metropolitan
Blagoevgrad	10	4	4	2
Burgas	13	4	3	6
Dobrich	6	2	1	3
Gabrovo	4	1	1	2
Haskovo	8	2	3	3
Kardzhali	5	3	1	1
Kyustendil	5	1	2	2
Lovech	5	2	3	0
Montana	5	2	3	0
Pazardzhik	9	3	3	3
Pernik	4	1	1	2
Pleven	9	3	2	4
Plovdiv	22	5	4	13
Razgrad	4	2	2	0
Ruse	8	2	1	5
Shumen	6	2	1	3
Silistra	4	2	2	0
Sliven	7	2	2	3
Smolyan	4	2	2	0
Sofia district	8	3	5	0
Sofia-city	42	2	0	40
Stara Zagora	11	3	2	6
Targovishte	4	2	2	0
Varna	15	3	2	10
Veliko Tarnovo	9	3	4	2
Vidin	3	1	0	2
Vratsa	6	2	2	2
Yambol	4	1	1	2
Total	**240**	**65**	**59**	**116**

- **Booster on Roma segregated communities**

For the booster, an expert database was used. It contains basic information for all segregated neighborhoods in the country like *locality* (district, municipality and settlement), an experts' approximation for the *number of population, number of households, number of houses* and other characteristics. The planned booster size sample was 300 households. We used this expert database for simple random sampling without replacement of segregated neighborhoods, weighted by their population. In this way, we selected 30 segregated neighborhoods in 20 districts. In each district, 10 randomly sampled households had to be interviewed.

Table 2. Booster—locality of the segregated neighborhoods

District	Municipality	Settlement
Blagoevgrad	Gotze Delchev	Gotze Delchev
Burgas	Burgas	Burgas
Burgas	Burgas	Rudnik
Varna	Varna	Varna
Veliko Tarnovo	Svishtov	Svishtov
Vratsa	Biala Slatina	Biala Slatina
Vratsa	Borovan	Borovan
Dobrich	Dobrich	Karapelit
Kyustendil	Kyustendil	Kyustendil
Lovech	Lukovit	Lukovit
Montana	Lom	Lom
Pazardzhik	Pazardzhik	Pazardzhik
Pazardzhik	Peshtera	Peshtera
Pazardzhik	Pazardzhik	Govedare
Pernik	Pernik	Pernik
Plovdiv	Plovdiv	Plovdiv
Plovdiv	Марица	Kalekovetz
Ruse	Ruse	Ruse
Ruse	Slivo pole	Kosharna
Sliven	Sliven	Sliven
Sofia district	Botevgrad	Novachene
Sofia district	Etropole	Etropole
Sofia-city	Sofia	Sofia
Stara Zagora	Stara Zagora	Stara Zagora
Stara Zagora	Kazanlak	Kazanlak
Stara Zagora	Stara Zagora	Kalitinovo
Targovishte	Targovishte	Ostrec
Haskovo	Dimitrovgrad	Dimitrovgrad
Haskovo	Dimitrovgrad	Krepost
Shumen	Shumen	Shumen

GPS sampling was used to identify the households in each cluster, because of the very specific living conditions in the segregated areas. For example, one neighborhood may be spread out an open field and there are neither addresses, nor streets, but only buildings out of regulation. Another example is neighborhoods, where at one single address there are two, three or more different buildings, inhabited by independent households. These specific conditions in segregated neighborhoods do not allow using a list of addresses from voting stations or any other kind of addresses.

The GPS sampling strategy included the following steps

1.   **First**, get geographical coordinates of the four framing points of the neighborhood or the settlement, if the whole settlement is of this type. Framing points are the most northerly, westerly, easterly and southerly points of the residential area. Connecting these four points, a rectangle appears and the neighborhood/settlement is inscribed in it.

2.   **Second**, select 20 geographical coordinates in this rectangle. This procedure is made by random number generator, as each coordinate is determined by its longitude and latitude. If some of the selected points belong to the rectangle, but don't belong to the neighborhood, they have to be removed from the list and new random coordinates have to be generated. For each selected coordinate, the nearest "door" (i.e. nearest dwelling) is chosen. This dwelling (house, or any kind of building) has to be described in some way that allows the interviewer to find it. If there is an address, it have to be written down, if not—each specific symbol has to be noted down, like household name, the color of the building, type of the door, post number and so on. Finally, there is a list with 20 randomly chosen "addresses"—10 for the main list and 10 for the additional one.

3.   If there are blocks of flats in the neighborhood, people who live there will have smaller chance to be selected than those who live in a house. The reason is that the procedure picks out the nearest door of dwelling to a chosen geographical coordinate in 2D. This problem could be solved by dividing the population into 2 parts—those who live in houses, and those who live in flats. Then, a list of all flats with the total number of people, who live there, has to be made. The total number of addresses (main and additional) also has to be divided into 2 parts proportionally to the population in the houses and in the flats respectively. In addition, two independent samples have to be drawn out—one for the houses (GPS sample), and one for the flats (simple random sample from the list of flats).

4.   To guarantee the representativeness of this procedure, an equal chance for each household selection has to be given. This could be done by creating a lattice over the area of the neighborhood with size of a cell equal to the size of the largest yard in it. Then select only some of the cells from this lattice by geometrical rule. The idea is to reduce the influence of the larger yards and to generate points (coordinates) randomly only in the selected sells.

The main advantage of this method in comparison with *"random walk"* is that with *"random walk"* the enumerator has the possibility to make his/her own choice, i.e. to select one household instead of another. By contrast, GPS sampling is not subjective.

C. Main sample characteristics

Both samples (main and booster) are representative for Bulgaria on a household level. In the roster, there is a detailed description of all household members. It contains information for basic demographic characteristics as sex, age, marital status, health insurance status, education, ethnicity, religion. Therefore, the main sample is representative also for the entire population in Bulgaria.

To examine the quality of the sample, we compare information, gathered by NSI in the official census 2011 and from this survey. The distribution of the population by some main demographic characteristics is the same as in the roster for this survey.

Table 3. Distribution by sex

	Population (NSI)		Main sample, 1st wave		Main sample, 2nd wave		Main sample, 3rd wave	
Sex	Count	Valid percent	Count	Valid percent	Count	Valid percent	Count	Valid percent
Male	3580337	48,7 percent	3255	48,9 percent	3013	48,4 percent	2992	48,4 percent
Female	3770897	51,3 percent	3396	51,1 percent	3212	51,6 percent	3186	51,6 percent
Total	7351234	100 percent	6651	100 percent	6225	100 percent	6178	100 percent

Table 4. Distribution by age groups

	Population (NSI)		Main sample, 1st wave		Main sample, 2nd wave		Main sample, 3rd wave	
Age	Count	Valid percent	Count	Valid percent	Count	Valid percent	Count	Valid percent
0—17	1172208	15,9 percent	1114	16,8 percent	1035	16,6 percent	979	15,8 percent
18—64	4789967	65,2 percent	4308	64,8 percent	4011	64,5 percent	3964	64,1 percent
65+	1389059	18,9 percent	1225	18,4 percent	1177	18,9 percent	1237	20,0 percent
Total	7351234	100,0 percent	6647	100,0 percent	6223	100,0 percent	6180	100,0 percent

D. Response rates

- **Planned and realized interviews**

Almost all of the planned interviews were realized in the main sample, as well as, in the booster. We deleted the information from several completed questionnaires because they contained

insufficient information, most of the questions had incorrect answers, or otherwise did not conform to the methodology of the survey.

There are several clusters with fewer than 10 interviews. That is because the enumerator could not end the fieldwork in time, or because there were regions (clusters) in which very few of the households agreed to give information about their families.

Table 5. Planned and realized interviews by type of samples

	Main sample		Booster	
	Count	percent	Count	percent
First wave (Feb. 2010)	2 384	99 percent	296	99 percent
Second wave (Sept. 2011)	2 298	96 percent	293	98 percent
Third wave (Feb. 2011)	2 329	97 percent	296	99 percent

- **Using additional addresses**

For each cluster, there was a list of 10 addresses that had to be visited by the interviewer and an additional 10 addresses in reserve. If any of the first 10 addresses did not exist, was locked for a long time or the people categorically refused to be interviewed, the additional ones came into use. According to the instructions, the interviewer had to visit each address in the main list three times, unless the building (or flat) was obviously uninhabited. The interviewer had a protocol: to note down what happened at each visit to each address of the list. At addresses where the interview did not take place, the interviewer noted the reason.. Once an interview was done, the questionnaire got an ID that showed whether the address was on the original list or not.

- **Percent of people who agreed to participate in the second wave**

The survey was designed to be a panel and the idea is that it will be repeated in six months (or in a year). In the first wave, at the end of the questionnaire there was a question for the respondent if he/she agreed to take part in the second survey wave or not. If he/she agreed, the interviewer wrote down his/her phone number or e-mail.

Table 7. Agreement for the second wave by types of samples

Agree for second wave	Main sample		Booster	
	Count	Total percent	Count	Total percent
Yes	1529	64 percent	139	47 percent
No	661	28 percent	109	37 percent
Non-response	194	8 percent	49	16 percent
Total	2384	100 percent	297	100 percent

In the main sample there were more refusals and non-found people than in the booster. On the contrary, people from the booster didn't want to participate in the second wave in a bigger extent than those from the main sample. Still, over 62 percent of all respondents in the first wave were ready to take part again in September, a rather high percentage.

Before the beginning of the second wave, we decided to visit each household from the first wave, nevertheless it had given agreement for second participation or not. Actually, there were households who had said they didn't want to be visited again but agreed to be interviewed a second time when the interviewers stopped by again. For that reason, we decided not to ask for permission for a third interview,   but to visit all the households from the first wave and possibly the second wave as well.

E. The panel component

This survey was planned and realized as panel survey in three waves. Households that were interviewed in February 2010, i.e. first wave, were asked if they agreed to be visited again after 6 months.

In the second wave, six months later, all the 2, 384 households were visited again whether or not they agreed to take place in the survey again. To complete the list with households from second wave to 2,400, some additional addresses were drawn out in the same sampling procedure, in the same clusters as in the first wave. Therefore, in the second wave, there were two types of households, taking place in the survey:

✓   **first**, these who were interviewed in the both waves;
✓   **second**, these only from the second wave (additional ones).
In the third wave, September 2011, the same scheme was followed. There were three types of addresses
✓   All the households from the first wave, nevertheless they were interviewed in the second, or not;
✓   Addresses of the households only from the second wave;
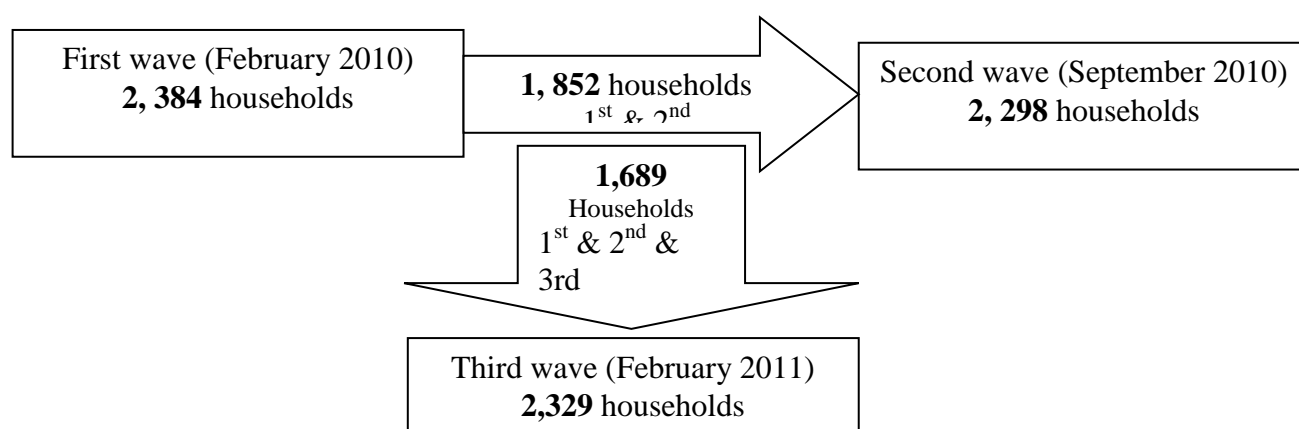✓   Additional addresses—to complete the list to 2, 400 households.

The enumerator had to visit first households from the first wave, then these, which were visited only in the second wave, and at last, if there were less than 10 interviews done in the cluster, to get addresses from the list with additional ones. Therefore, in the third wave, there were four types of households, taking place in the survey:
✓   From all the 3 waves;
✓   From the first and the third wave;
✓   From the second and the third wave;
✓   Only from the third wave.

To be able to make conclusions from the panel, i.e. to observe changes at an individual level, we used information the households that were interviewed in all three waves, i.e. first type from the list above.

We have a very good response rate for the panel survey. In the second wave, all 2, 384 households were visited again. From this population, 1,852 households were interviewed in the second wave, 78 percent of the original respondents. The remaining 22 percent were completed with new addresses and each household that was not interviewed was replaced with a new one in the same cluster. Therefore, the distribution of the sample by *district, settlement type* and *clusters* is the same in the second and third waves as in the first one.

In the third wave (September 2011) 1,689 of the households was interviewed for the third time, i.e. this is the number of participants in the three waves of the survey, which is 71 percent of the initial population of 2, 384 households.

First wave (February 2010)
2, 384 households

1, 852 households
1st & 2nd

Second wave (September 2010)
2, 298 households

1,689
Households
1st & 2nd &
3rd

Third wave (February 2011)
2,329 households

The quality of the panel is examined by comparing the main demographic characteristics of the population in the panel with these from the official Census 2011, conducted by NSI. There are no significant differences between the distribution of the two populations by sex and age.

Sex	Population (NSI)		Population in the panel	
	Count	Valid percent	Count	Valid percent
Male	3 580 337	48,7 percent	2 105	48,2 percent
Female	3 770 897	51,3 percent	2 265	51,8 percent
Total	7 351 234	100 percent	4 370	100 percent

Age	Population (NSI)		Population in the panel	
	Count	Valid %	Count	Valid percent
0—17	1 172 208	15,9	722	16,5
18—64	4 789 967	65,2	2 785	63,7
65+	1 389 059	18,9	863	19,7
Total	7 351 234	100,0	4 370	100,0