Russia 2000: Survey Information

Summary table

Generic information	
Name of survey	Russia Longitudinal Monitoring Survey (RLMS)
Institution responsible	Russian Institute of Nutrition / Carolina Population Center at the University of North Carolina at Chapel Hill / Institute of Sociology of the Russian Academy of Sciences
Frequency	Annual
Survey year / Wave	Round IX (2000)
Collection period	From 21 September to 23 December 2000
Survey structure	Cross-sectional and longitudinal
Coverage	Private households in most of the territory
Geographic information	8 main regions
Files delivered	Several files at the household and individual level
Sample size	
Households	4,006 households
Individuals	9,074 individuals aged 14 and over , and 2,023 children under 14
Sampling	
Sampling design	*Initial sample* by multi-stage probability: - first stage: the raions (PSUs) were allocated into 38 strata (geographical factors, level of urbanization and ethnicity) and one raion was selected from each stratum using "probability proportional to size"; - second stage: in rural areas of the selected PSUs, a list of all villages was compiled to serve as SSUs; in urban areas, SSUs were defined by the boundaries of 1989 census enumeration districts; - third stage: 10 households were selected from each SSUs. *Following waves:* same dwellings of the initial sample ("old" or "new" households living in them), even if they did not respond the preceding wave.
Sampling frame	In rural areas: reliable lists of households existing in the villages In urban areas: list of dwellings developed by survey conductors
Questionnaires	The RLMS survey instruments include: household questionnaire, adult questionnaire, child questionnaire and community questionnaire
Standard classifications	
Education	0 to 11 years of schooling plus 6 different national higher degrees
Occupation	4-digit ISCO-88 standard
Industry	Not available
Income	
Reference period	30 days preceding the interview date, which varies between 23/09/00 and 21/12/00
Unit of collection	All sources are collected at the household level, while earnings, unemployment benefits and pensions are collected also at the individual level
Period of collection	Mostly monthly income (without information on number of months), home production yearly
Gross/net	Variables are recorded net of taxes and contributions
Data editing / processing	
Consistency checks	Extensive interviewers training, quality data entry (with 1% of entries corrected).
Weighting	Cross-sectional household weights that correct for (1) the probability of selection for each sample household; (2) household non-response based on geographic and other known characteristics of sample households; (3) non-coverage biases in the frame used to derive the original sample of dwellings and individuals . Individual weights also correct for non-response at the level of the individual.
Imputation	Yes (imputation of missing values by sample median).

This document draws extensively upon the web site of the RLMS (see http://www.cpc.unc.edu/projects/rlms/home.html).

Table of contents:

A. General characteristics

Official name of the survey/data source:
Russia Longitudinal Monitoring Survey (RLMS)

Administrative Unit responsible for the survey:

Carolina Population Center at the University of North Carolina at Chapel Hill
WWW: http://www.cpc.unc.edu/projects/rlms/home.html
Hotline: rlms@unc.edu

The RLMS is a household-based survey designed to measure the effects of Russian reforms on the economic well-being of households and individuals. In particular, determining the impact of reforms on household consumption and individual health is essential, as most of the subsidies provided to protect food production and health care have been or will be reduced, eliminated, or at least dramatically changed. These effects are measured by a variety of means: detailed monitoring of individuals' health status and dietary intake; precise measurement of household-level expenditures and service utilization; and collection of relevant community-level data, including region-specific prices and community infrastructure data. Data have been collected ten times since 1992.

In the initial two years of this effort, a main goal of the RLMS was to work with the Russian State Statistical Bureau (Goskomstat) and the All-Russia Center of Preventive Medicine to upgrade the systems in place for monitoring these issues. A breakdown in the collection of statistical data was occurring throughout the former Soviet Union. In addition, it had become clear that Russian Federation data collection systems did not provide a representative profile of the economic and social dimensions of the population. In particular, adequate monitoring of the poor did not take place. Among the accomplishments of Phase I was the creation of the first national sample frame, allowing surveys to be representative at the national level. More recently, this sample frame has been extended to develop samples representative at the regional and oblast levels.

For Phase II, begun in 1994, the RLMS switched collaborators in Russia and the emphasis of the work changed from institution-building to providing timely, high-quality information. The new RLMS sample is smaller, but the number of primary sampling units was doubled to enhance the representativeness of the survey.

Funding for the RLMS has been provided mainly by the United States Agency for International Development and the National Institutes of Health. Additional support has come from the National Science Foundation, the World Bank, the Swedish Ministry of Foreign Affairs (through the Stockholm Institute of Transition Economies), and the University of North Carolina at Chapel Hill.

Data are now available through anonymous FTP. In order to receive data, please see http://www.cpc.unc.edu/projects/rlms/data.html from where it is possible to access an order form for obtaining data sets in SAS XPORT format.

B. Population, sampling size and sampling methods

In Phase II of the RLMS, a multi-stage probability sample was employed. First, a list of 2,029 consolidated raions was created to serve as primary sampling units (PSUs). These were allocated into 38 strata based largely on geographical factors and level of urbanization, but also based on ethnicity where there was salient variability. As in many national surveys involving face-to-face interviews, some remote areas were eliminated to contain costs; also, Chechnya was eliminated due to armed conflict. From among the remaining 1,850 raions (containing 95.6% of the population), three very large population units were selected with certainty: Moscow city, Moscow Oblast, and St. Petersburg city constituted self-representing (SR) strata. The remaining non-self-representing raions (NSR) were allocated to 35 equal-sized strata. One raion was then selected from each NSR stratum using the method "probability proportional to size" (PPS). That is, the probability that a raion in a given NSR stratum was selected was directly proportional to its measure of population size.

The NSR strata all have approximately equal sizes because they were purposefully designed that way to improve the efficiency of estimates. The target population (omitting the deliberate exclusions described above) numbers over 140 million inhabitants. Ideally, one would use the population of eligible households, not the population of individuals. As is often the case, we were obliged to use figures on the population of individuals as a surrogate because of the unavailability of household figures in various regions.

Although the target sample size was set at 4,000, the number of households drawn into the sample was inflated to 4,718 to allow for a non-response rate of approximately 15%. The number of households drawn from each of the NSR strata was approximately equal (averaging 108), since the strata were of approximately equal size and PPS was employed to draw the PSUs in each one. However, because we expected response rates to be higher in urban areas than in rural areas, the extent of over-sampling varies. This accounts for the differences in households drawn across the NSR PSUs. It also accounts for the fact

that 940 households were drawn in the three SR strata--more than the 14.6% (i.e., 689) that would have been allotted based on strict proportionality.

Since there was no consolidated list of households or dwellings in any of the 38 selected PSUs, an intermediate stage of selection was then introduced, as usual. Professional samplers will recognize that this is actually the first stage of selection in the three SR strata, since those units were selected with certainty. That is, technically, in Moscow, St. Petersburg, and Moscow Oblast, the census enumeration districts are the PSUs. However, it is cumbersome to keep making this distinction throughout the description, and we shall follow the normal practice of using the terms "PSU" and "SSU" loosely. Needless to say, in the calculation of design effects, where the distinction is critical, we have maintained the proper distinction. The selection of second-stage units (SSUs) differed depending on whether the population was urban (located in cities and "villages of the city type," known as "PGTs") or rural (located in villages). That is, within each selected PSU the population was stratified into urban and rural substrata, and the target sample size was allocated proportionately to the two substrata. For example, if 40% of the population in a given region was rural, 40 of the 100 households allotted to the stratum were drawn from villages.

In rural areas of the selected PSUs, a list of all villages was compiled to serve as SSUs. The list was ordered by size and (where salient) by ethnic composition. PPS was employed to select one village for each ten households allocated to the rural substratum. Again, under the standard principles of PPS, once the required number of villages was selected, an equal number of households in the sample (10) was allocated to each village. Since villages maintain very reliable lists of households, in each selected village the 10 households were selected systematically from the household list. In a few cases, villages were judged to be too small to sustain independent interviews with 10 households; in such cases, 3 or 4 tiny villages were treated as a single SSU for sampling purposes.

In urban areas, SSUs were defined by the boundaries of 1989 census enumeration districts, if possible. If the necessary information was not available, 1994 microcensus enumeration districts, voting districts, or residential postal zones were employed--in decreasing order of preference. Since census enumeration districts were originally designed to be roughly equal in population size, one district was selected systematically without using PPS for each 10 households required in the sample. In the few cases where postal zones were used, one zone was likewise selected systematically for each 10 households. However, where voting districts were used, to compensate for the marked variation in population size, PPS was employed to select one voting district for each 10 households required in the urban sub-stratum.

Given the lack of reliable official lists of households within the urban SSUs, the list of households from which ten households were selected had to be developed. First, a list of dwellings was made. Where more than one household was known to exist within a single dwelling (that is, in the communal apartments and enterprise dormitories that are relatively commonplace in the Russian Federation), the list was amended so that each household (or space within the dwelling) was enumerated in advance of selection. Then,

the required number of households was drawn systematically, starting with a random selection in the first interval.

As described above, the sample frame was essentially based on dwellings in urban areas and households in rural areas. In conducting Rounds VI, VII, VIII, IX, and X interviewers in both urban and rural areas attempted to conduct interviews in the same dwellings (or spaces within communal apartments and dormitories) that fell into the Round V sample. They returned to each Round V dwelling even if the household in the dwelling had refused to participate during previous rounds, and even if they found out that the household whom they interviewed in previous rounds had moved to a new dwelling prior to the interview.

Since the change in housing stock was minuscule between late 1994 and late 1995, this procedure insured that the results in 1995 were approximately as representative as they were in 1994. The response rate was nearly the same: 84% in Round V; 80% in Round VI--both respectable figures in survey research requiring such substantial face-to-face interviews about every member of every household. Furthermore, by returning to every dwelling we actually obtained interviews from some 200 households who had declined to participate in Round V. This should eventually permit some analysis of the nature of non-response in Round V--an analysis that would be more sophisticated than merely comparing the demographic characteristics of households to those in the census.

It is especially important to notice that this procedure did not appreciably vitiate our ability to conduct panel analyses with Round V and VI data. First, it goes without saying that the data set renders it quite easy to identify households and people who participated in both rounds. Second, as it turned out, only 250 households (6.3%) from Round V moved from their dwellings and were thus lost to Round VI--a low level of attrition for a panel survey of this sort. Nevertheless, we did gather data on their new addresses whenever possible in anticipation of a supplementary study to follow up on them.

As stated above, the household response rate exceeded 80%. As in Round V, individual questionnaires were obtained from over 97% of the individuals listed on the household rosters. The response rates did indeed vary across PSUs depending on the proportion of households in rural areas. However, since we anticipated that in over-sampling, the actual proportion of completed household interviews compares well to the proportion of the population in each stratum. The distribution of household size in the sample, within both rural and urban localities, corresponds well to the figures from the 1989 census. Bear in mind that single-member households are excluded from the comparison because the census includes many institutionalized people, while our sample explicitly excludes them. Thus, there is no valid basis for comparison.

The multivariate distribution of the sample by sex, age, and urban-rural location compares quite well with the corresponding multivariate distribution of the 1989 census. Of course, due to random sampling error and changes in the distribution since the 1989 census, we would not expect perfect correspondence. Nevertheless, there is usually a difference of only one percentage point or less between the two distributions.

Another way to evaluate the adequacy (or efficiency) of the sample is to examine design effects. An important factor in determining the precision of estimates in multi-stage samples is the mean ultimate cluster (PSU) size. All else being equal, the larger the size, the worse the precision. In Rounds I-IV of the RLMS, the average cluster size approached 360--a large number dictated by constraints imposed by our collaborators. Thus, although the sample size hovered around 6,000 households, precision was less than we would have liked for a sample of that size. In Rounds I and III of the RLMS, the 95% confidence interval for household income was about ±13%.

In the Phase II sample, the situation was considerably better. Although there were only 4,000 households, the mean size of clusters was much smaller than in Phase I. There were 35 PSUs with about 100 households each; even this was an improvement over the average of 360 in the design of the RLMS Rounds I-IV. However, in the three self-representing areas, the respondents were drawn from 61 PSUs. Recall that Moscow city and oblast, as well as St. Petersburg city, were not sampled, but were chosen with certainty. Therefore, the first stage of selection in them was the selection of census enumeration districts. Thus the mean cluster size in all the sample was about 42, i.e., 4,000/(35+61). Given these much smaller cluster sizes, we had reason to expect that precision in this survey would be as good as it was in Rounds I-IV despite the smaller sample size. This, in fact, turned out to be the case in Rounds V-X.

B. Data collection and acquisition

Data collection period

Beginning in 1994 and ongoing, the RLMS has collected five rounds of data in the second phase of the project.

Data Schedule for the RLMS Phase II (Rounds V-X)

	Training	**Collection**	**Data Entry**	**Data Cleaning***
Round V	10/94	11/94 to 12/94	12/94 to 1/95	12/94 to 4/95
Round VI	10/95	10/95 to 12/95	11/95 to 1/96	12/95 to 4/96
Round VII	10/96	10/96 to 12/96	11/96 to 1/97	12/96 to 4/97
Round VIII	10/98 to 11/98	10/98 to 1/99	12/98 to 2/99	1/99 to 5/99
Round IX	**9/00 to 10/00**	**9/00 to 12/00**	**11/00 to 1/01**	**12/00 to 4/01**
Round X	9/01 to 10/01	9/01 to 12/01	11/01 to 1/02	12/01 to 4/02

* Data analysis begins 7-10 days after data cleaning is completed.

Survey instruments

RLMS survey instruments were designed by an interdisciplinary group of Russian and American social science and biomedical researchers with extensive experience in survey research. Particular care was taken to collect data that would allow us to answer policy-relevant questions concerning the design and impact of programs and policies affecting a

wide range of social sector outcomes. The survey is designed to allow various modules of questions to be included from round to round.

Interview methodology

In both urban and rural substrata, interviewers were required to visit each selected dwelling up to three times to secure the interviews. They were not allowed to make substitutions of any sort. The interviewers' first task was to identify households at the designated dwellings. "Household" was defined as a group of people who live together in a given domicile, and who share common income and expenditures. Households were also defined to include unmarried children, eighteen years of age or younger, who were temporarily residing outside the domicile at the time of the survey. If perchance the interviewer identified more than one household in the dwelling, he or she was obliged to select one using a procedure outlined in the technical report. The interviewer then administered a household questionnaire to the most knowledgeable and willing member of the household.

The interviewer then conducted interviews with as many adults as possible, acquiring data about their individual activities and health. Data for the children's questionnaires were obtained from adults in the household. By virtue of the fact that an attempt was made to obtain individual questionnaires for all members of households, the sample constitutes a proper probability sample of individuals as well as of households, without any special weighting. Actually, the fact that we did not interview unmarried minors living temporarily outside the domicile slightly diminishes the representativeness of the sample of individuals in that age group.

Interviewing quality control

In Phase II, it was the responsibility of local supervisors to gather the necessary information for sampling in accordance with written instructions, to arrange for training facilities, to invite people to be trained, to supervise their work, and to check the completed questionnaires. All local supervisors consulted by telephone with representatives in Moscow who could answer their questions in advance.

All interviewers underwent a demanding training regimen, outlined below. Any trainee whose performance during training revealed him or her to be unsuited for the job was dismissed before field work began.

1. Lectured on the general principles of face-to-face interviewing. We provided a 70-minute video tape entitled "Introduction to Interviewing" to insure that all interviewers received the same instructions and examples. Where there was no available VCR, we rented video salons.
2. Required interviewers to read through the entire questionnaire in advance, then to fill out the questionnaire themselves.
3. Showed interviewers an example of a good interview with commentary, again using a video tape. The tape include a section on the diet portion of the questionnaire.
4. Introduced them to the written questionnaire specifications, entitled "Interviewer Instructions."
5. Played the role of respondent while trainees took turns reading questions as they would in an actual interview.
6. Had the interviewers practice interviewing in groups of three. One assumed the role of interviewer; another, the role of respondent; the third, the role of observer, watching to see whether the interviewer was working properly. The trainer and perhaps some other experienced interviewers circulated among the triads to observe.
7. Gave the interviewers written exercises that tested their ability to react properly to certain difficult situations in administering the questionnaire.
8. Reviewed the administrative procedures pertaining to the survey.
9. Gave the trainees practice in persuading respondents to participate by having them role play.
10. Required interviewers to complete at least one practice interview with a household that was not in the sample--preferably not a household related to them, although they were allowed to practice with relatives first.
11. Examined their work after each of their first three interviews or more, until they demonstrated that they were competent.

Data entry

In Phase II, when questionnaires were returned to local supervisors, those supervisors were required to examine them to locate problems that could best be remedied in the field, e.g., by returning to get key demographic information or cleaning ID numbers so

that the roster of individuals located in the household questionnaire matched those on the individual questionnaires from that household. The questionnaires were then transported to Moscow, where yet another ID check was performed.

In Moscow, coders looked through all questionnaires to code so-called "other: specify" responses. However, open-ended questions (e.g., occupation questions) were not coded at this time. Instead, their texts were fully entered as long string variables. (Please note that these character variables are not available at this time.) Entering the open-ended answers as character variables offers several advantages. First, it allows data entry to begin immediately, with no delay for coding. Second, it permits the use of computer programs to assist in coding the string variables. Third, this method allows any user of the original data sets to recode the character variables to suit his or her purposes without going back to the paper copies of the questionnaires.

All data entry was handled in-house using the SPSS data entry program on PCs. For the first survey of Phase II, Round V, the first pass of data entry began on December 20, 1994, and finished on February 1, 1995. The second (verification) pass overlapped with the first to speed up the process. It began on January 15, 1995, and was completed on February 8, 1995 (with the exception of the diet data). The second pass revealed an error rate of 1% in each pass. Rounds VI, VII, VIII, IX, and X used a similar timeframe.

C. Definition of the survey units

Household (or family)
All people living together and having common income and expenditures; unmarried children under the age of 18 not living with the household because of study reasons, are to be included in the household.

Head of household
The family member with the best knowledge of the affairs and who worries of the family and of its present income and expenditures.

D. Contents

The household questionnaire consists of 5 sections:

- Family information (household roster with main demographics of all members)
- Living conditions (housing assets)
- Farming and animal husbandry (consumed, given away and sold own production)
- Expenditures (very detailed diaries: food in the last 7 days, services and utilities in the last 30 days, clothing and durables in the last 3 months)
- Income (fuel benefits, children's benefits, 10 categories of gratuitous money, earnings, 12 categories of payments).

The adult questionnaire (for individuals aged 14 and over) includes 4 sections:

- Migration
- Work (detailed information about primary and secondary current job and additional paid activities, including cash and non-cash earnings, satisfaction indicators, values questions, education characteristics, income from pensions and unemployment benefits, job search activities, main occupation)
- Medical services
- Interviewer's remarks

The child questionnaire (for individuals aged under) includes 4 sections:

- Migration
- Care of children
- Medical services
- Health evaluation
- Diet
- Medical measurement
- Interviewer's remarks

The community questionnaire contains the following information for each survey site:

- Demographic Characteristics
- Types of Housing Available
- Transportation and Communications
- Health Care Facilities
- Public Dining
- Employment Opportunities
- Municipal Services
- Minimum and Maximum Prices for a Number of Food Items

E. Quality of data

Cross-Sectional and Longitudinal Design and Analysis

Data from the RLMS may be used in two types of analyses.

A. Repeated Cross-Section Analysis

As its name implies, the RLMS is a longitudinal study of populations of dwelling units. Rounds V-VII are designed to provide a repeated cross-section sampling. Barring the construction of major new housing structures, renewed contact with a fixed national probability sample of dwelling units provides high coverage cross-sectional representation. The repeat visit at each round to a static sample of dwelling units also introduces a correlation between successive samples that leads to improved efficiency in longitudinal analyses comparing aggregate statistics.

The repeated cross-section design is far and away the simplest alternative for the RLMS. The sampling is cost efficient, easy to maintain, and easy to update when needed. The design supports both efficient cross-sectional and aggregate longitudinal analyses of

change in the Russian household population. Updates to the sample, including a full replenishment of the probability sample of dwelling units, will not seriously disrupt the longitudinal data series.

B. Longitudinal or "Panel" Analysis

The primary disadvantage of a repeated cross-section design is that it does not enable micro-level analysis of longitudinal change at the household or individual level. The exception is the potential to link households and individuals who remain in the original dwelling unit over time, but such a "panel" may be vulnerable to selection bias when reasons for moving are correlated with the dependent variable of interest (see 2.B. below).

A true panel design in which sample households and individuals are followed and interviewed at each wave would be preferred if the sole purpose of the RLMS were to study micro-level change. The original sampling plan for Rounds V-VII did not call for households to be followed if they moved from the Round V sample dwelling unit. Likewise, individual household members who moved away were not to be followed. At each round, the RLMS interview was completed with the household and its members in the original sample dwelling unit. Consequently, the RLMS is not a true panel design, although Round VII departed from the original protocol and followed some households and individuals who moved. 1

Sample Attrition

The first question is the nature of attrition in the RLMS samples and its impact on cross-sectional and longitudinal analysis of the data.

A. Attrition Effects on the Analysis of the Repeated Cross-Section Data

Sample attrition due to nonresponse cannot be avoided. Table 1 summarizes RLMS Round V interview completion rates for the original sample of dwelling units in the eight regions that comprise the survey population. These are not response rates; each denominator includes dwelling units that were vacant or uninhabitable at the time of the Round V interviews. Overall, interviews were completed in 84.3% of the original national probability sample of n=4718 dwelling units.

Table 1: RLMS Round V Interview Completion Rates*

Region	**N**	**Dwelling Interview (%)**
Moscow/St. Petersburg	686	60.2
North/Northwestern	319	88.7
Central/Central Black Earth	923	84.8
Volga/Viask/Volga Basin	770	89.4
North Caucuses	538	87.6
Urals	619	91.0
Western Siberia	416	92.6
Eastern Siberia/Far East	447	87.0
TOTAL	***4718***	***84.3***

St. Petersburg	222	67.1
Moscow	464	56.9

* Including vacancy, no contact, refusal.

Interview completion rates outside St. Petersburg, Moscow City, and Moscow Oblast range from 84.8% in the combined Central/Central Black Earth region to 92.6% in Western Siberia. Rates in the highly urban Moscow/St. Petersburg region are much lower. In part, these rates may reflect higher vacancy rates in metropolitan areas, but clearly lower household contact and response rates also come into play. Lower rates in Moscow and St. Petersburg were anticipated at the design stage, and initial allocations to these strata were increased to offset expected losses from refusal and noncontact. This is one form of what we might call "designing for nonresponse." The over-sampling strategy is beneficial in that it means reduced variability in the final analysis weights (due to the offset in the product of higher sample selection probability and lower response propensity); however, over-sampling eliminates the potential for bias only if attrition is occurring at random within the final weighting adjustment cells.

If independent samples were developed for each round of the repeated cross-section design, attrition in one round would be independent of (although possibly similar in nature to) that in other rounds. However, since the RLMS uses a static sample of dwellings across multiple rounds, the impact of nonresponse and attrition is the net effect of several factors. Round V attrition bias can arise only from differential nonresponse and noncontact for subclasses of households that occupy the original sample of dwelling units. The potential for nonresponse bias in cross-sectional analysis or contrasts involving the Rounds VI and VII data is a complex function of: (1) initial nonresponse in Round V; (2) net difference in characteristics of households and individuals who move out of or into sample dwellings; (3) nonresponse on the part of old households continuing to reside in sample dwelling units; and (4) nonresponse on the part of new households currently living in sample dwelling units.

Time did not permit analysis of each of these factors. Instead, I performed several simple analyses of the net effect of household turnover and nonresponse on the marginal sample distributions (unweighted) of population characteristics that should not change significantly over time.

Table 2 compares the unweighted distribution of the Round V-VII interview households by region, settlement type, characteristics of household head, and household size. The general observation is that the combined influence of nonresponse attrition and household turnover does not seriously distort the geographic distribution of the sample or its size or household-head characteristics. The distributions for the geographic variables indicate that, between Round V and Round VII, there is a decline in the nominal representation of households in the Moscow/St. Petersburg region, reflected in a decline in the proportion of sample households from the urban domain. Households with a male head aged 18-59 may be subject to slightly higher than average attrition/net loss in replacement. If we focus only on these characteristics, the problem is not serious.

Table 2: Net Attrition/Recruitment Effect on Cross-Sectional Composition of Household Sample

	Percent by Category		
Subpopulation	**Round V**	**Round VI**	**Round VII***
REGION			
Moscow/St. Petersburg	10.4	9.2	8.5
North/Northwestern	7.1	7.2	7.3
Central/Central Black Earth	19.7	19.4	20.1
Volga/Viask/Volga Basin	17.3	17.6	17.9
North Caucuses	11.8	12.0	12.2
Urals	14.2	14.8	14.7
Western Siberia	9.7	9.8	9.4
Eastern Siberia/Far East	9.8	10.2	10.0
SETTLEMENT TYPE			
Urban	70.2	69.3	68.4
PTG	5.4	5.6	5.8
Rural	24.4	25.1	25.8
HOUSEHOLD HEAD			
Older child (7-18)	0.1	0.1	0.1
Male (18-59)	64.8	63.6	63.2
Female (18-54)	10.8	11.2	11.7
Male (60+)	11.6	11.8	11.9
Female (55+)	12.7	13.4	13.2
HOUSEHOLD SIZE			
1	17.6	18.7	19.0
2	26.9	26.1	26.6
3	23.8	23.7	24.0
4	21.0	20.0	19.7
5	7.0	7.6	6.6
6+	3.8	3.9	4.1

* Including households followed to new residences.

Table 3 gives a similar comparison of the unweighted marginal frequencies for individual sample members interviewed in Rounds V-VII. Again, the combined effects of attrition and change in dwelling unit occupants result in a net decline across rounds in the proportion of sample individuals from the Moscow/St.Petersburg region and an associated decline between Rounds V and VII in the percent of sample individuals from urban areas. We also find a modest decline in the proportion of males aged 0-19 between Rounds V and VII.

In summary, the net effect of nonresponse attrition and change in dwelling unit occupants across rounds on the marginal characteristics of the observed cross-sectional samples is modest. Loss in nominal "sample share" between Rounds V and VII is greatest for residents of Moscow/St. Petersburg--a loss in representation that is readily corrected with the combined sample selection/nonresponse adjustment factors that have been computed for each round. It is important to note that the simple analysis described here cannot demonstrate that no uncorrected attrition bias remains. The potential for uncorrected nonresponse bias can be specific to the dependent variable under study. Nevertheless, it appears that, with the nonresponse and post-stratification adjustments developed by Michael Swafford, the potential for serious attrition bias in repeated cross-section analysis is small.

Table 3: Net Attrition/Recruitment Effect on Cross-Sectional Composition of Individual Sample

	Percent by Category					
Subpopulation	**Round V**		**Round VI**		**Round VII***	
REGION						
Moscow/St. Petersburg	10.5		9.0		8.0	
North/Northwestern	7.2		7.2		7.0	
Central/Central Black Earth	18.1		17.8		18.6	
Volga/Viask/Volga Basin	17.0		17.3		17.6	
North Caucuses	13.4		13.9		14.1	
Urals	14.4		14.9		14.7	
Western Siberia	9.9		9.8		9.7	
Eastern Siberia/Far East	9.6		10.1		10.2	
SETTLEMENT TYPE						
Urban	69.3		68.2		66.8	
PTG	5.5		5.7		6.2	
Rural	25.2		26.0		27.0	
AGE GROUP/SEX	**M**	**F**	**M**	**F**	**M**	**F**
0-19	14.5	14.0	14.3	14.0	14.0	14.0
20-39	13.9	15.6	13.6	15.3	13.6	15.3
40-59	11.1	13.6	11.4	13.6	11.3	13.7
60-79	5.5	9.5	5.5	9.8	5.5	10.2
80+	0.4	1.8	0.4	1.9	0.5	1.9

* Including individuals followed to new residences.

B. Attrition Effects on Simulated "Pure Panel" Analysis

The intent behind the RLMS design is that data be analyzed as repeated cross-sections of the Russian population. An interesting question is, "How misleading would it be to conduct pure panel analysis of households and individuals observed in Rounds V and VI or in Rounds V-VII?" The obvious problem is that by definition analysis can include only

households and individuals who continue to reside in the original sample dwelling units and who participate in two or three consecutive rounds of the study.

Table 4: Attrition Effects for Round V Household Panel, Round V Characteristics for Retained Sample

	Percent by Category		
Subpopulation	**Round V Panel**	**Round VI Panel**	**Round VII Panel***
REGION			
Moscow/St. Petersburg	10.4	8.4	7.5
North/Northwestern	7.1	7.4	7.3
Central/Central Black Earth	19.7	20.1	20.6
Volga/Viask/Volga Basin	17.3	18.3	18.8
North Caucuses	11.8	11.8	12.2
Urals	14.2	14.8	15.0
Western Siberia	9.7	9.6	9.6
Eastern Siberia/Far East	9.8	9.6	9.0
SETTLEMENT TYPE			
Urban	70.2	67.2	65.7
PTG	5.4	5.6	5.6
Rural	24.4	27.2	28.8
HOUSEHOLD HEAD			
Older child (7-18)	0.1	0.1	<0.1
Male (18-59)	64.8	64.6	64.5
Female (18-54)	10.8	10.1	10.0
Male (60+)	11.6	12.0	12.3
Female (55+)	12.7	13.2	13.1
HOUSEHOLD SIZE			
1	17.5	17.0	16.0
2	26.9	27.2	27.8
3	23.8	23.1	22.9
4	21.0	21.4	21.5
5	7.0	7.2	7.6
6+	3.8	4.1	4.2
NUMBER OF CHILDREN <7			
0	78.5	78.8	78.5
1	17.8	17.5	17.7
2+	3.7	3.7	3.8
NUMBER OF CHILDREN 7-18			
0	65.2	64.6	64.1
1	22.4	22.5	22.6
2+	12.4	12.9	13.3

NUMBER OF WORKING-AGE MALES			
0	35.2	35.5	35.5
1	55.0	54.3	54.3
2+	9.8	10.2	10.2
NUMBER OF WORKING-AGE FEMALES			
0	34.7	35.5	35.6
1	56.4	55.5	55.6
2+	8.9	9.0	8.8

* Including households followed to new residences.

Tables 4 and 5 give a partial answer to the question. The second column in each shows a multinomial distribution or median value of a characteristic as measured for the Round V sample of cooperating households. The third column gives the same statistic (again the Round V characteristic) but computed only for households that participated in both Rounds V and VI. The final column gives the statistic based on Round V measures only for households that participated in all three rounds.

Here, as was the case for cross-sectional analysis, the most notable effect of attrition is the loss in the percentage of sample households from the Moscow/St. Petersburg region and the broader urban domain. Between Rounds V and VII there is also a modest loss in the relative percentage of single-person households. Round V-VII attrition does not appear to seriously distort the relative distribution of households by count of children or numbers of working men and women.

Table 5 shows the impact of Round V-VII attrition on the financial characteristics of the household "panel." It suggests that households that move out of their original residences or decline to participate at Round VI, or Rounds VI and VII, have higher median incomes and expenditures than households that remain in their original residences and continue to cooperate in the RLMS.

Table 5: Attrition in the Round V Panel, Round V Income Statistics for Respondents and Nonrespondents at Later Rounds

	Round V Panel		**Round VI Panel**		**Round VII Panel**
Statistic	**R**	**R**	**NR**	**R**	**NR**
Round V Median Household Income	354,564	349,000	396,490	344,000	395,095
Round V Median Household Expenditure	466,593	465,552	474,404	463,657	498,451
Round V Median Income, % Poverty	2.024	1.995	2.179	1.976	2.138

Table 6 repeats the Table 4 analysis for a "panel" of individual respondents. As with households, nonresponse and movement have the greatest impact on the percent of individuals from the Moscow/St. Petersburg region and the more general urban domain. Attrition effects on the relative age/sex distribution produce a general aging of the

"panel" of individuals. Consistent with the finding for households, nonresponse and movement result in losses of "panel" members from the higher economic ranks. Interestingly, there is only a slight disproportionate tendency for individuals who are unemployed at Round V to leave the sample at Round VI or VII. Those who remain at Rounds VI and VII are slightly older and are more likely to have had a normal body weight at Round V than are those who left after Round V.

Table 6: Attrition Effects for the Round V Individual Panel

	Percent by Category					
Subpopulation	**Round V Panel**		**Round VI Panel**		**Round VII Panel***	
REGION						
Moscow/St. Petersburg	10.5		8.0		7.0	
North/Northwestern	7.2		7.5		7.1	
Central/Central Black Earth	18.1		18.4		19.1	
Volga/Viask/Volga Basin	17.0		18.3		19.0	
North Caususes	13.4		13.5		13.5	
Urals	14.4		15.0		15.5	
Western Siberia	9.9		9.8		9.9	
Eastern Siberia/Far East	9.6		9.5		9.0	
SETTLEMENT TYPE						
Urban	69.3		66.0		64.4	
PTG	5.5		5.8		5.9	
Rural	25.2		28.2		29.7	
AGE GROUP/SEX	**M**	**F**	**M**	**F**	**M**	**F**
0-19	14.5	14.0	14.2	14.3	14.2	14.2
20-39	13.9	15.6	13.0	15.0	12.4	15.0
40-59	11.1	13.6	11.2	14.2	11.2	14.8
60-79	5.5	9.5	5.7	10.2	5.6	10.6
80+	0.4	1.8	0.4	1.8	0.3	1.7
ECONOMIC RANK						
1	12.6		13.2		13.1	
2	15.4		15.8		15.8	
3	24.0		24.3		24.5	
4	22.5		21.6		21.6	
5	19.4		19.5		19.5	
6	4.1		3.9		3.8	
7+	1.9		1.6		1.6	
NORMAL WEIGHT?						
% Yes	54.5		56.4		57.4	
UNEMPLOYED?						
% No	96.2		96.5		96.7	

MEDIAN AGE	34	36	36

Replenishing the RLMS Sample

As noted above, in the absence of housing construction, the original sample of dwelling units provides a cross-sectional representation of the Russian household population at each observed point. Of course, over a reasonable period there will be housing construction, and occupants of new units should be included in a sample that is to be nationally representative. Techniques such as those used in the U.S and Canada for sampling new housing construction could be employed to update the original sample of dwellings, but these techniques are complicated, and the necessary data (building permits, data from planning or housing agencies) may be difficult or expensive to collect today in Russia.

Most current housing construction in Russia is concentrated in multi-unit structures and development areas. It may be possible to replenish the sample by drawing a new sample of dwellings from the original enumeration lists compiled prior to Round V. New listings could be prepared for new housing structures located within the existing sample of second stage units (SSUs). A supplemental sample (at the correct rate for the SSU) could be selected from the new housing listing and combined with the sample from the listing of pre-existing housing to form an updated sample of dwellings.

Replenishment of the sample at some point may also be a good idea to avoid more serious problems of attrition among households that continue to reside in the original sample of dwelling units. The timing of replenishment will depend on several factors, not the least of which is cost.

Weights in Descriptive Analysis of RLMS Data

Analysis weights are essential for unbiased sample-based estimation of RLMS descriptive statistics such as population and subclass means, proportions, and totals. The construction of a descriptive weight for cross-sectional analysis involves a simple sequence of steps: (1) determine the probability of selection for each sample household; (2) based on geographic and other known characteristics of sample households, compute an adjustment for nonresponding sample households; and (3) compute a nonresponse-adjusted weight as the product of he reciprocal of the sample selection probability and the nonresponse adjustment.

Since the RLMS attempts to interview all individuals within sample households, the selection probability for an individual equals that for his household. An individual in a cooperating household may, however, choose not to give an interview. If data on individuals-- both cooperating and not--are known from household listings, the nonresponse adjustment factor in the analysis weight can be computed at the level of the individual. Fortunately, the majority of RLMS nonresponse at the individual level corresponds to noncooperation by the entire household, and the household nonresponse adjustment factor will capture most of the sample attrition loss at both levels.

If recent census data on households and individuals are available, a fourth post-stratification step can be added: scaling analysis weights so that the sum of weights for a defined subpopulation matches the corresponding census proportion (e.g., the weighted sample proportion of females, age 45 and older, in the Moscow/St. Petersburg region matches the corresponding proportion from the most recent census). The post-stratification of analysis weights serves two functions: (1) it can reduce the sampling variance of weighted estimates; more importantly, (2) it may correct noncoverage biases in the frame used to derive the original sample of dwellings and individuals.

There is considerable debate over the value of using weights in multivariate analysis. For example, in estimating linear or generalized linear models, many software programs allow the specification of weights for model fitting. Some statisticians argue that using weights is not necessary if the fixed effects that explain the variation in weights are included in the model. In RLMS data, the household characteristics that explain the greatest variation in weights are the geographic region and the urban/rural character of the civil division in which the dwelling is located. Variation in individual weights will reflect the geographic effects for households as well as differentials due to post-stratification of the sample by major geographic regions, age, and sex. Researchers who are interested in exploring the impact of RLMS weights on a multivariate analysis should consider the following test. Fit the model omitting the weights but including as fixed effects the household (region, urban/rural) or individual (region, urban/rural, age, and sex) characteristics. Without changing the specification, also estimate the model using the analysis weights. Compare the results to see if there are important differences in model parameters and/or interpretation. Differences in the unweighted and weighted versions could be due to added sampling variability introduced by the weighted estimation or could indicate that the model is not correctly specified.

Constructed variables

Several variables are constructed on the basis of the questionnaire direct variables. Constructed variable data sets are currently available for Rounds V-X only. In the absence of detailed documentation, the variable labels in each data set describe the contents of the variables.

The number of observations in each data set does not necessarily match the total number of observations in the original data files. For the economics data sets, filter criteria were established so that only families with complete economic information were included. The health data sets used the maximum number of non-missing observations per individual analysis. Thus, the health data sets vary in composition more than do the economics ones.

In the economics data sets, nominal ruble values are those figures that appear in the original data. Real ruble amounts are nominal values that have been adjusted to June 1992 rubles.

The chart below describes the different types of constructed-variable data sets available for Rounds V-X and the unit of observation for each.

Economics Data		**Health Data**	
Household Assets	Household	Adult Nutrition	Individual
Demographic Composition	Household	Caloric Intake	Individual
Household Expenditures	Household	Alcohol Consumption	Individual
Household Income	Household	Drug Availability	Individual
Household Poverty Line	Household	Immunizations	Individual
Labor Force Participation	Individual	Children's Nutrition	Individual
		Medical Problems	Individual
		Smoking	Individual

Imputation

Different steps of imputation are also carried out within the process of variable construction.

Constructed Individual Level income variables - The imputation procedure consists in:

- Calculate the median values of the inflation adjusted income variables of interest, subsetted by settlement type, gender, and age category
- Calculate the median values of the inflation adjusted variables for the entire sample
- Merge both back into main set with inflation adjusted variables
- If indicator variable says that the person engages in this activity, but they don't report any income from that activity, then replace the missing income value with the categorized imputed value. If the categorized imputed value does not exist, replace with the whole sample imputed value.
- Check the counts for the # of imputed values for a particular observation, categorized and whole sample.
- Re-inflate the imputed real variables back to their nominal values

In general, less than 1% of any particular variable is imputed, and it never materially affects the characteristics of the variable.

Constructed Household Level income variables - The imputation procedure consists in:

- After deflating, calculate the median of the real variables, subsetted by settlement type (urban or rural, and family size. Also calculate the median based on the entire sample, and save both to output sets.
- If the indicator variable for that income category says that the household should have income from that source, but the level variable is missing, then replace with the categorical median. If the above is true and the categorical median is missing,

then replace the level variable with the whole sample median. The program also collects counts of imputation for a few variables.

- Re-inflate the real level variables back to their nominal values, by multiplying by the appropriate inflation index.

Value of the home production of fruits and vegetables consumed or given away. Note that the construct sum(var1 var2 0) means that if all of the variables being added together are missing, then the sum is equal to zero. Also note that the P_* variables used are the site level prices that we collected in our community data.

Constructed Household Level expenditure variables - The imputation procedure consists in:

- Divide all of the monetary expenditure variables by their corresponding inflation index, in order to get June 92 rubles for imputation.
- Calculate the median of each real variable, by settlement type SETT_TYP and family size FAMSIZE, and by the entire sample. Save the resultant data sets and re-merge back into the main working data.
- If the indicator variable for a particular expense indicates that the household incurred that expense, but the actual amount variable is missing (for example, H7PURPOT is the yes/no response to whether the family purchased potatoes, while the variable H7PPOTAT is the amount of those expenses), then replace the missing value with the family size and settlement type specific median. If the above holds true and the categorical median is missing as well, then replace the missing amount with the median of the entire sample.
- Re-inflate the relative monetary variables back to their original values.

Information on the actual substitutions counts for a few variables is also collected, namely those indicating the purchase of various fuels and rent/utilities payments. For these variables at least, the total imputation count never exceeds 1% of the data.

F. Uses of the survey

The main results of the survey are published after each round by the survey organisers in two series: “Monitoring Economic Condition in the Russian Federation: the Russia Longitudinal Monitoring Survey” and “Monitoring Health Condition in the Russian Federation: the Russia Longitudinal Monitoring Survey” (see below for exact references).

Publications

Published (Or In Press) Articles

Entwisle, B., and P. Kozyreva (1997). "New Estimates of Induced Abortion in Russia." Studies in Family Planning 28(1):14-23.

Levin, S., B.E. Ainsworth, C.W. Kwok, C.L. Addy, and B.M. Popkin (In Press). "Patterns of Physical Activity Among Russian Youth: The Russian Longitudinal Monitoring Survey." European Journal of Public Health.

Lokshin, M., and B.M. Popkin (1999). "The Emerging Underclass in the Russian Federation: Income Dynamics 1992-96." Economic Development and Cultural Change 47:803-29.

Lokshin, M., K.M. Harris, and B.M Popkin (2000). "Single Mothers in Russia: Household Strategies for Coping with Poverty." World Development 28.

Lokshin, M., and M. Ravallion (2000). "Welfare Impacts of the 1998 Financial Crisis in Russia and the Response to the Public Safety Net." Economics of Transition 8(2).

Mroz, T., and B.M. Popkin (1995). "Poverty and the Economic Transition in the Russian Federation." Economic Development and Cultural Change 44:1-31.

Popkin, B.M., N. Zohoori, and A. Baturin (1996). "The Nutritional Status of the Elderly in Russia, 1992 through 1994." American Journal of Public Health 86(3):355-60.

Popkin, B.M., L. Kohlmeier, N. Zohoori, A. Baturin, A. Martinchik, and A. Deev (1997). "Nutritional Risk Factors in the Former Soviet Union." In Premature Death in the New Independent States. Jose-Luis Bobadilla, Christine Costello, and Faith Mitchell, Editors. Committee on Population, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, DC: Academy Press. pp 314-334.

Popkin, B.M., A. Baturin, L. Kohlmeier, and N. Zohoori (1997). "Russia: Monitoring Nutritional Change during the Reform Period." In Implementing Dietary Guidelines for Healthy Eating. V. Wheelock, Editor. London: Chapman and Hall. pp 23-46.

Ravallion, M., and M. Lokshin (In Press). "Who Wants to Redistribute Russia's Tunnel Effect in the 1990's?" Journal of Public Economics.

Tseng, M., H. Chakrborty, D.T. Robinson, and L. Kohlmeier (1997). "Adjustment for Iron Availability in Epidemiologic Studies." Journal of Nutrition 127:1456-68.

Zohoori, N. (1997). "Recent Patterns of Alcohol Consumption in the Russian Elderly, 1992-1996." American Journal of Clinical Nutrition 66(4):810-814.

Zohoori, N., P. Kozyreva, M. Kosolapov, M. Swafford, B. Popkin, E. Glinskaya, M. Lokshin, and D. Mancini (1998). "Monitoring the Economic Transition in the Russian Federation and its Implications for the Demographic Crisis--the Russian Longitudinal Monitoring Survey." World Development 26(11):1977-1993.

Completed Articles Submitted To Journals

Entwisle, B., and L. Watterson. "Family Planning and Abortion in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-96." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. March 1997.

Kohlmeier, L., M. Mendez, S. Shalnova, A. Martinchik, H. Chakraborty, and M. Kohlmeier (1995). "Deficient Dietary Iron Intakes among Women and Children in Russia: Evidence from the Russia Longitudinal Monitoring Survey (RLMS)."

Lokshin, M. (1999). "Household Childcare Choices and Women's Work Behavior in Russia." Policy Research Working Paper Series #2206. Development Research Group, The World Bank, Washington, DC.

Mroz, T., L. Henderson, M. Bontch-Osmolovsii, and B.M. Popkin. "Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2001." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. April 2002. (Russian text)

Mroz, T., L. Henderson, and B.M. Popkin. "Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2000." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. March 2001.

Mroz, T., D. Mancini, and B.M. Popkin. "Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-98." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. March 1999.

Ravallion, M., and M. Lokshin (1999). "Subjective Economic Welfare." Policy Research Working Paper #2106. Development Research Group, The World Bank, Washington DC.

Tudor-Locke, C., B.E. Ainsworth, L.J. Neff, C.L. Addy, and B.M. Popkin (2000). "Omission of Active Commuting to School and the Prevalence of Children's Health-Related Physical Activity Levels: The Russia Longitudinal Monitoring Study."

Tudor-Locke, C., B.E. Ainsworth, S.G. Trost, and B.M. Popkin (2000). "Physical Activity in Russian Children."

Vannappagari, V., and R. Ryder. "Monitoring Sexual Behavior in the Russian Federation: The Russia Longitudinal Monitoring Survey 2001." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. April 2002. (Russian text)

Zohoori, N., K. Gleiter, and B.M. Popkin. "Monitoring Health Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2000." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. March 2001.

Zohoori, N., K. Gleiter, and B.M. Popkin. "Monitoring Health Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2001." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. April 2002. (Russian text)

Zohoori, N., L. Henderson, K. Gleiter, and B.M. Popkin. "Monitoring Health Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-98." Report submitted to the U.S. Agency for International Development. Carolina Population Center, University of North Carolina at Chapel Hill, North Carolina. March 1999.

Poverty and income distribution

According to the Report "Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-2001", 26.5% of Russian households had an income below the poverty line, where the poverty level is defined as the subsistence level based on adjustments for economies of scale, oblast-level prices, and regional food baskets.

According to a study by Ovtcharova ("What kind of Poverty alleviation policy does Russia need, Research Paper of the Russian-European Centre for Economic Policy, May 2001), the Gini coefficient amounted to 0.401 in 2000.

Under the official methodology for determining the number of the poor, this group includes the entire number of people with income below the subsistence level (based on the value of the minimum market basket), amounting to 34.7% in 2000.