UNECE 2022

Improving survey methods in Kazakhstan

According to the recommendations of the UNECE Poverty Measurement: Guide to Data Disaggregation

Report prepared for UNECE by Gianni Betti

Table of Contents

1. Introduction

This report is prepared under a consultancy conducted under the Poverty measurement stream of the United Nations Development Account 13th tranche project "Strengthening social protection for pandemic responses". It has the main scope of providing technical assistance to implement the recommendations related to assessing and improving survey methods of the UNECE Poverty measurement: Guide to Data Disaggregation (UNECE, 2020).

In particular, the report objectives are focused on the improvement of the representativeness under conditions of increasing non-response, and in the advance the production of disaggregations of relevance to poverty and inequality, including in producing the corresponding SDGs indicators.

The report has been prepared in collaboration with the Bureau of National Statistics (hereinafter – BNS) of Kazakhstan, also to help optimise the survey design and enhance survey precision in the context of the pandemic.

The current version of the sampling design of the Household Budget Survey was developed in 2010 under the project of the Living Standard Measurement Study with the support of the World Bank.

Then, one of the main instruments for the implementation of the HBS, namely the questionnaires' set, was fully analysed during May 2018 in the Department of Labour and Living Standards Statistics of the Committee on Statistics of the Ministry of National Economy of the Republic of Kazakhstan; the main scope was to implement a comparative analysis of the country HBS questionnaire with the Model Questionnaire prepared by UNECE in 2017.

Moreover, another important source of information for this report has been the recent analysis of household survey and poverty measurement in the Republic of Kazakhstan in the context of a pandemic, conducted in 2021, which has highlighted the potential increase in non-response rates in household surveys due to the pandemic periods.

Taking into account all the above considerations, the present report has firstly taken into account and analysed the sampling methods currently employed by the BNS in conducting the household budget survey, by assessing its capacity for producing accurate and representative data. These issues have been addressed in Section 2 of the report itself.

Then, the report has the scope of further assist in reinforcing the use of statistical techniques to address the issues of non-response and coverage error, and ensure better representativity of the sample, taking into account the challenges posed by the pandemic. These techniques have been fully discussed in Section 3.

Finally, suggestions and recommendations related to assessing and improving survey methods of the UNECE (2020) Poverty Measurement: Guide to Data Disaggregation have fully discussed in Section 4.

The report has also the aim of providing recommendations and practical examples with the scope of transform the theory and technical sides into concrete actions. This has been done also by engaging proactively with the experts responsible for poverty measurement in the BNS and by taking into account the information provided by them. The work for the writing of the report has been carried out in close contact with the staff of the Social and Demographic Statistics section of the UNECE Statistical division.

2. Analysis of the sampling methods currently employed by the Bureau of National Statistics of Kazakhstan

a. Description

This section has the aim of describing the sampling methods currently employed by the BNS for the Household Budget Survey (hereinafter HBS) used in the context of data disaggregation on poverty measurement.

Such sampling methods have been developed in the context of the Joint Economic Research Project for 2010 “Development of a new sample for a household survey to assess living standards” co-funded by the World Bank. The HBS is also known as Living Standard Measurement Survey, as all such surveys developed in collaboration with the World Bank.

This description is based on a document provided by the BNS (translated in English by Hasanov, 2022), by documents from BNS, UNECE and World Bank web sites, and by the International HBS network; the section is composed of four parts: i) definition of the population under investigation and sample frame; ii) sample design; iii) sample size and iv) sampling over time.

i) The population under investigation consists in all households and their components (individuals) resident in the Republic of Kazakhstan. As in any sample household survey, the population living in collective housing such as: orphanages, dormitories, nursing homes, homes for children with disabilities, hotels, military barracks, hospitals, sanatoriums, prisons are not included in the study.

The main source for better representing such population under investigation is the information system “Statistical Register of Housing” (hereinafter – SRH), which is an important component of the integrated information system "e-Statistics" of Kazakhstan. As well described in Hasanov (2022), the use of this database has several advantages, which include the following: availability of a ready-made sampling frame; this excludes one of the significant costs related to the compilation of the sampling frame; availability of an ongoing updated database, which is necessary to rotate the households; availability of regional information on households.

ii) The sampling design of households is based on multi-probabilistic (random) sampling. This principle guarantees the independence of the selection of the sample units (households and individuals) and avoid deliberate its mistakes. In particular, the sample of households is formed by two-stage probabilistic (random) sampling using stratification and random selection procedures at each of the sampling stages. The stratification procedure is aimed at forming a representative sample of households that adequately reflects the regional characteristics of the population.

Generally speaking, such sampling design follows the principle of the optimal combination of costs and specified criteria for the accuracy of the results. In particular, in the first stage clusters (groups) of dwellings are selected in order to reduce the costs of the field work; on the other side, the selection of such clusters is performed by stratifying them on a territorial basis, so to guarantee representativeness of the population over such regions (accuracy of the results).

At the **first stage**, the general population is stratified according to the regional distribution, including the distribution into urban and rural areas. Thus, 31 strata are formed – these are selected urban and rural areas in sixteen regions of the country (considering that there are no rural areas in the cities of Astana and Almaty). This, despite to the general description reported in English in some international web sites (i.e. the international network of HBSs, https://catalog.ihsn.org/catalog/8742), where the description reports six generic strata (groups of regions).

A certain number of clusters (regional units) are selected as Primary Selection Units (hereinafter – PSU), defined by the first six digits of the code according to the Classifier of Administrative Territorial Objects (hereinafter – CATO). This first procedure identifies PSU within each stratum with a probability proportional to size (hereinafter - PPS), that is, the number of PSU in a stratum is formed depending on the number of households present in the stratum itself.

In the **second stage**, a certain number of households are then selected from each sampled PSU. The basis for the sampling at the second stage is the list of individual dwellings in the PSU. These dwellings to be visited during the survey are selected with equal probability from among the eligible dwellings in the PSU.

iii) Also the overall sample size is calculated on the basis of the principle of the optimal combination of costs and specified criteria for the accuracy of the results, introduced above.

In general terms, the algorithm used in order to define the national sample size is the following equation (1):

$$SE = \frac{\delta\sqrt{Deff}}{\sqrt{n}}$$

where, SE is the (target) sample standard error of the estimate of interest; δ is the dispersion in such estimate of interest; Deff (or design effect) is the impact (effect) of sampling design described above; and n is the overall (national) sample size to be drawn.

Such overall (national) sample size is defined on the basis of the (target) sample standard error and on the basis of design effect (Deff) for urban and rural areas, which have been set to 1.0 and 2.0, respectively. The assumption of Deff = 1.0 for urban areas is explained by the fact that the urban clusters of the survey are close to random sampling, since large cities are not subdivided into smaller territorial units. The assumption of Deff = 2.0 in rural areas is based solely on the experience of other countries.

The BNS has taken into account three scenarios for deciding the optimal sample size – in function of both costs and accuracy of results). The sample size has been set at 12,000 households (0.3% of the total population): this ensures to obtain results with an error of no more than 4% at the national level and no more than 7% at the regional level, as it is shown from Table 1.

Table 1. Determining the overall sample size

Standard Error Value		Sample size: number of households	Survey costs, million tenge
National level	by regions		
Option №1		24,000	800

<1%	<3%		
Option №2		18,000	600
<2%	<5%		
Option №3		**12,000**	**400**
<4%	**<7%**		

Source: HBS methodological note, translated by Hasanov (2022).

The allocation of the overall sample size among the strata (regions / urban rural) is determined in Table 2. First of all, the sample of 400 PSUs are allocated to each of the 31 strata according to PPS; then in each PSU, 30 dwellings are selected in order to reach a total of 400 x 30 = 12,000 dwellings.

Table 2. Determining the sample sizes across regions / urban and rural

#	Region	Number PSUs	including		Number of households	including	
			Urban	Rural		Urban	Rural
1	Akmola region	28	12	16	840	360	480
2	Aktobe region	26	12	14	780	360	420
3	Alma-Ata's region	24	8	16	720	240	480
4	Atyrau region	17	10	7	510	300	210
5	West Kazakhstan region	22	8	14	660	240	420
6	Dzhambul region	21	9	12	630	270	360
7	Karaganda region	32	20	12	960	600	360
8	Kostanay region	27	12	15	810	360	450
9	Kyzylorda region	20	8	12	600	240	360
10	Mangistau region	18	11	7	540	330	210
11	Pavlodar region	28	12	16	840	360	480
12	South Kazakhstan region	22	9	13	660	270	390
13	Turkestan region	20	7	13	600	210	390
14	East Kazakhstan region	30	15	15	900	450	450
15	Nur-Sultan	20	20	-	600	600	-
16	Almaty city	27	27	-	810	810	-
17	Shimkent	18	18	-	540	540	-
	Total Kazakhstan	**400**	**218**	**182**	**12000**	**6540**	**5460**

iv) The longitudinal sampling design over time does not consist in a classical series of independent cross-sectional surveys, but rather a rotating panel design.

This is due to two main factors: from one hand, it facilitates the fieldwork over years; from the other hand, the short period has the scope to reduce households' burden from participation in the survey, and the higher representativeness by the selection of new and fresh panels. However, it is important to notice that the process of systematic sample rotation does not relate to the replacement of households that refuse to participate in the survey.

Rotation of households (updating the sample) is carried out annually in the amount of 1/3 of the total number of surveyed households. That is, every year 1/3 of households are withdrawn from the sample and replaced by others. Thus, after 3 years, the sample is completely updated.

In order to implement this rotating design, at the end of each the year (in December), 10 households in each PSU are randomly selected and removed from the sample. Another 10 of the initially selected households are randomly selected and removed from the sample at the end of the following year, and the remaining 10 households at the end of the following year. Each time, 10 new households are randomly selected from the updated SRH database to take the place of the withdrawn ones.

b. Analysis and findings

The sampling methods currently employed by the BNS for the HBS generally complies with the sampling methods accepted in the world practice and ensures statistical data of high-quality. The sampling methods, implemented in 2010, have been monitored by World Bank staff in 2017 and evaluated as "still good".

The probability of selection of PSU (hi) in a certain stratum (h) over time follows standard formulas, i.e. in equation (2):

$$p_{hi} = \frac{s_h n_{hi}}{N_h}$$

where P_{hi} is the selection probability; s_h is the number of PSUs selected in stratum h; n_{hi} is the number of households in PSU hi, according to SRH; and N_h is the total number of households in the stratum, still according to the SRH.

Such probabilities change over time, taking into account the change in the overall population and its distribution over the regions and strata. In the last 5 years, distribution of population has moved from rural to urban areas, and this has properly been taken into account. As Table A.1 shows in the Annex, in 2015 about half of total population was in rural areas, and the number of selected PSUs in urban and rural areas was similar, 208 and 192 respectively. This proportion has changed over the last decade, as shown in Table 2, where the number of PSUs selected in rural areas has increased by 10 to 192.

The first area for improvements in the sampling design undertaken by the BNS for the HBS could be identified in the hypotheses made on the design effect (Deff) values for the urban and rural sub-samples. Apparently, in fact, the values of Deff = 1 for urban and Deff = 2 for rural, seems to be too small, and without a scientific base of evidence.

In accordance with and the support of the BNS and the Social and Demographic Statistics section of the UNECE Statistical division, in this report we have estimated such design effects on the basis of a real data set, and theory provided by the most relevant literature. First of all, the design effects have been estimated with recent method of Verma and Betti (2011), where Deff is defined as the ratio of the standard error estimated under the current design, divided by the standard error estimated as the sample would have been drawn under simple random sampling (SRS); the description is reported in Section 5.4 of Verma and Betti (2011). The standard errors have been estimated via one of the methods suggested by UNECE (2020, p. 149), namely the Jack-knife Repeated Replications method (in the modified version of Verma and Betti, 2011).

The real estimated design effects are **Deff = 1.24** and **Deff = 1.89** for urban and rural area respectively. Overall, the design effect for the whole HBS in the Republic of Kazakhstan could be seen as very similar to the one currently undertaken by BNS; instead, the specific values for the two urban / rural areas should be revised on the basis of the empirical evidence.

In fact, the hypothesis that the sampling design currently implemented in the urban areas is comparable to the simple random sampling (SRS) is quite strong.

As it is possible to observe by the probabilities of selection, and the consequent calculation of weights (see also Section 3 below), dwellings in different urban areas have different probabilities of selection, which substantially differs from the classical SRS. Moreover, the presence of unequal weights generates the so-called "Kish effect" (Kish, 1965) – also highlighted in the UNECE (2020, p. 133) guide – which let the standard errors and therefore the design effect increase.

In fact, the Kish effect for the 2020 HBS – urban areas – has been estimated in Kish = 1.10, which accounts for more than 40% of the overall design effect.

Such design effects have been estimated on the basis of a small data set provided to the Author by the BNS; the few need variables are stratum, PSU, rural / urban, weight and target variable of interest. Figure A.1. in the Annex shows the screenshot of the first rows of the SAS data set used for this purpose.

Recommendation n. 1. It is recommended to adopt the new Deff = 1.24 and Deff = 1.89 for urban and rural areas respectively, for the calculation of the optimal sample sizes, and their allocation over regions in the future HBS sampling strategy.

Moreover, it is recommended to re-calculate such design effects every 5 years in order to take into account possible changes in the structure of the population in such urban and rural areas of the Republic of Kazakhstan.

Another interesting aspect of the sampling design undertaken by the BNS for the HBS is the longitudinal dimension. In fact, as well described in the previous section, the design over time does not consist in a classical series of independent cross-sectional surveys, but rather a rotating panel design.

Such rotating panel design over 3-year period implies that about one third of the overall sample (i.e. about 4,000 households) is interviewed for three consecutive years. This is an enormous source of information on poverty dynamics, which is probably underexplored by researchers.

A simple way to analyse such rich longitudinal data set would be to study the distribution of years spent in poverty, i.e. the percentage / rate of individuals and households which have been in the state of poverty for: 0 years (never), 1 or 2 years (transient poverty), or all 3 years (persistent poverty).

Recommendation n. 2. As an example, it is recommended to explore the Eurostat web survey on the similar "Distribution of population by number of years spent in poverty within a four-year period" (https://ec.europa.eu/eurostat/web/products-datasets/-/ilc_li51).

Table A.2 in the Annex is an example of how reporting the distribution of number of years in poverty over a period of four years. For the case of Kazakhstan, the reduced sample size will let possible to estimate such indicators at the national (Kazakhstan) level, but not at the regional level; however, standard errors should still be acceptable for estimates based on homogeneous groups of regions. Other variable for disaggregation will be discussed in Section 4 below.

The issue of analysing poverty dynamics is very important for at least two main reasons; first of all, evidence from research in several regions of the world (see for instance Ravallion and Jalan, 2000) shows that the causes of persistent or chronic poverty are much different from those of transient or temporary poverty. In fact, the two most asked questions in this area of research are: " …for those who experience it, is poverty a lasting or a temporary condition? What types of individuals and households are more likely to endure either "permanent" or "transitory" poverty, and what programs are more likely to be effective in reducing the two types of need?

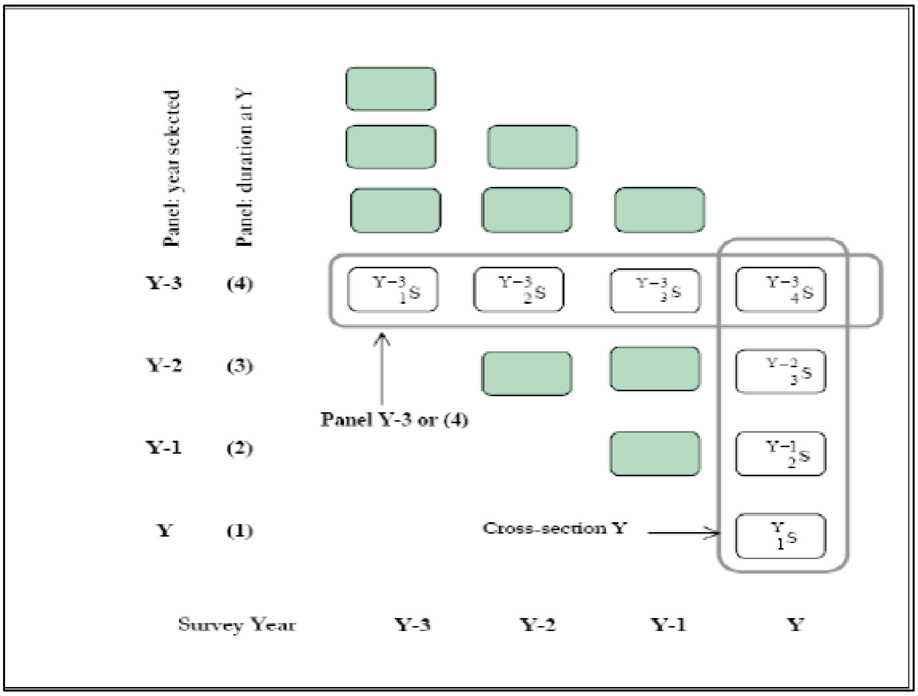
The answers to these questions have important bearing on the direction of government policy, and this is the second important issue of studying poverty dynamics: short-term measures such as income transfers are adequate to relieve temporary distress; long-term measures including structural changes in the labour market as well as investment in education, training, and special services are needed to address persistent poverty. For these reasons, an in-depth study of poverty dynamics could let the Government of the Republic of Kazakhstan to better target adequate anti-poverty policies.

The longitudinal aspect of the sampling design currently undertaken by the BNS for the HBS could be interesting for further improve easily the estimation of longitudinal estimates of poverty.

With the current design, the availability of the 3-year panel is present only every 3 years, i.e. when the rotating group selected three years ago is eventually discarded in the current year t. Maintaining this base structure, but introducing the 3-year length for any rotating panel, could easily and quickly improve the design.

This is described by means of the well-known rotating scheme adopted by the EU-SILC survey (4-year panels instead of 3-year, see Figure 1) described by Verma and Betti (2006) and well documented by the most recent Eurostat Doc 65 (Eurostat, 2020).

Figure 1. - EU-SILC standardised rotating scheme



Source: Verma and Betti (2006)

This design has also fully suggested by the UNECE (2020) guide. In particular, at page 119 describes the EU-SILC rotating panel, where the main characteristics are that this design yields a cross-sectional sample as well as longitudinal samples of various durations. In particular, the cross-sectional sample for year Y consists of four subsamples, 1-4, one subsample introduced each year from (Y-3) to Y.

A longitudinal sample consists of households who have remained in the survey since they were first introduced into it. Three overlapping longitudinal samples of different durations are formed: of two-year duration from subsamples (2+3+4), of three-year duration from subsamples (3+4), and of four-year duration from subsample (4).

Recommendation n. 3: It is recommended to transform the current design of the Household Budget Survey to one similar to the 4-year rotating panel of EU-SILC.

Such sampling design could have a positive effect also in reducing the so-called attrition, i.e. the unit non-response due to the respondents' burden. This is already partially taken into account by the current sampling design implemented by the BNS, but attrition could increase in the near future due to the change in the respondents' behaviour for the Covid-19 pandemic effects.

3. Addressing the quality of data: issues of non-response and coverage errors

This section aims at analysing a typology of errors in the Household Budget Survey data collected by the BNS; this typology distinguishes various categories of errors in measurement on the one hand, and errors arising in the process of estimation from the survey sample to the target population on the other hand. Important sources and types of errors – including conceptual, measurement, and unit and item non-response errors – are explained also in the light of the recent Covid-19 pandemic situation.

a. Description

Knowledge about data quality is required for their proper use and interpretation. Also, measures of data quality are important for the evaluation and improvement of survey design and procedures. Continued monitoring and improvement of data quality is particularly important in major continuing surveys such as the HBS conducted by the BNS. There are diverse forms and many different sources of errors in surveys, and various frameworks have been proposed for their classification. Different frameworks emphasise different aspects of the problem. None may be considered as 'the best', though some frameworks are more illuminating than others. Here we follow the framework introduced by Verma (1981), further elaborated in Husmanns *et al.* (1990) and Verma *et al.* (2010), and summarised by UNECE (2020, pp. 106-107) in the context of data disaggregation of poverty measures. This framework distinguishes between two groups of errors affecting the survey process:

(A) Errors in measurement: these arise from the fact that what is measured on the units included in the survey can depart from the actual (true) values for those units. These errors concern the accuracy of measurement at the level of individual units enumerated in the survey, and centre on substantive content of the survey: definition of the survey objectives and questions; ability and willingness of the respondent to provide the information sought; and the quality of data collection, recording and processing. This group of errors can be studied in relation to various stages of the survey operation.

(B) Errors in estimation: these are errors in the process of extrapolation from the particular units enumerated in the survey to the entire study population for which estimates or inferences are required. These centre on the process of sample design and implementation, and include errors of coverage, sample selection and implementation, non-response, and also sampling errors and estimation bias. The above categorisation, in terms of errors in measurement and errors in estimation, is more fundamental than the distinction usually made between sampling and non-sampling errors.

A third category, namely item non-response, has been added as an intermediate category between measurement and estimation errors.

Each group of errors may be further classified in more detail in order to identify specific sources of error, so as to facilitate their assessment and control. However, it is important to note that the various phases of a survey are closely related. While it is useful to classify the total survey error into components, errors cannot always be attributed to a particular type or source. The same or similar methods of assessment and control may be suited for measuring more than one type of error, and some of the indicators obtained may provide no more than a general or overall measure of data accuracy without being able to identify specific sources and types of error.

This typology of errors is very relevant in the context of a pandemic situation, as shown by several works in the literature. For instance, during the coronavirus pandemic, the UK Office for National Statistics (NSO, 2021) has published estimates of personal well-being showing the high impact that the pandemic has had on data collection (in terms of errors in measurement), on estimates of personal well-being (in terms of errors in estimation) and the comparability of these estimates.

Moreover, in a study of the practices of UNECE countries' national statistical offices in adapting to the COVID-19 crisis situation their household surveys that are used for poverty measurement, Hasanov and Hasanova (2020) have highlighted interesting issues which may have introduced systematic errors in data collection and poverty measurements. All these issues and aspects will be analysed in the following sub-sections, and the corresponding findings – with the help of selected case studies – will be utilised for making useful recommendations.

b. Analysis and findings

Before making an in-depth analysis of the data quality of the HBS conducted by the BNS, it is useful to further describe the steps undertaken during the fieldwork, and the post-survey steps.

As already introduced in Section 2, in the second stage of sampling, 30 households are randomly selected in each of the 400 PSUs, for a total of about 12,000 households. The basis for the sampling at the second stage is the list of individual dwellings in the PSU. The cluster (or clusters) of dwellings to be visited during the survey is selected with equal probability from among the eligible dwellings in the PSU. The probability (p_{hij}) of choosing a household (hij) in the PSU (hi) of a stratum (h) is determined by the following equation (3):

$$p_{hij} = p_{hi} \frac{m_{hi}}{n'_{hi}}$$

where, p_{hi} is the probability of choosing PSU hi given in equation (1); m_{hi} is the number of dwellings in PSU hi to be drawn (usually 30); n'_{hi} is total number of eligible dwellings in PSU.

There are cases during the fieldwork when it is not possible to interview households because the dwelling is not found or not occupied, or the household refuses to participate in the survey. Refusal is undesirable because it reduces the sample size and mainly because it is a source of potential sample bias and leads to skewed statistical results.

A first, effective, prevention of overcoming the problem of non-receipt of data is the careful documentation of each case. Since in the fieldwork it becomes always difficult in the Republic of Kazakhstan to comply with the "no replacement" principle, then a list of reserve households is provided in the amount of one third of the number of households in each cluster (i.e. 10 households per PSU). For these reasons, a total of 40 dwellings are selected in each cluster: 30 in the nominal sample and 10 substitutes.

The post-survey steps are limited to the weighting procedure, described here below. The implementation of this method is carried out by assigning to each responded household a statistical weight that takes into account the total number of households represented by the responded part that included into the sample. The weights for indicators of the standard of living of the population are calculated for each quarter of the HBS.

The weights are calculated using the distribution of surveyed households separately by urban and rural population for regions.

The probability weight (W_{hij}) of a household in PSU (hi) of stratum hi (h) is the reciprocal of its selection probability p_{hij} and is determined by equation (4) as follows:

$$w_{hij} = \frac{1}{p_{hij}} = \frac{N_h n'_{hi}}{s_h n_{hi} m_{hi}}$$

where, W_{hij} is the household weight; P_{hij} is the probability of selection; N_h is the total number of households in the stratum; n'_{hi} is the total number of eligible dwellings in PSU; S_h is the number of PSUs selected in stratum h; m_{hi} is the number of selected dwellings in PSU hi. Since $n_{hi} \approx n'_{hi}$ and $m_{hi} \approx 30$ in all PSUs, the sample is roughly self-weighted within each stratum.

The weights are rescaled such as the sum of the weights provides an estimate of the number of all households in a region and the country as a whole. Although their use makes it possible to maintain the compliance of the sampling population with the original principles of sampling, while in practice there are cases that violate this compliance (refusals to participate in the survey and other cases of non-receipt of data). Unit non-response from selected households is an important source of errors described in Section 2.

The method undertaken by the BNS is straightforward and is in line with international standards, but it is applied at PSU level; in fact, in order to compensate for cases of non-response, a simple weight adjustment scheme is applied by assigning larger weights to all responding households in surveyed PSU. The weights of all responding households in surveyed PSU are increased by the same coefficient. As described in Hasanov (2022), for instance if 90 percent of households in the PSU answered all questions, the weights for all respondents are increased by a coefficient of 1/0.9=1.11. All non-responding households are excluded from the sample by assigning an actual weight of zero for each household.

Recommendation n.4: Use larger areas for non-response adjustment. This will reduce the presence of extreme weights in some specific PSU and will reduce the current Kish effect (effect of weights), which determine larger variability in the estimates (standard errors).

As a practical example, a simple and implementable step that can lead to concrete actions could be the so-called Eurostat EU_SILC Doc 65 (Eurostat, 2020, p. 35): here it is recommended to perform non-response adjustment for “Homogenous groups” of about 100-300 households. It therefore should be much larger than 30 households in a PSU. A **practical** suggestion could be a group of homogenous PSUs (4-8), in order to reach the size of about 200.

Differently, the solution for addressing the problem of item non-response - i.e. the absence of data only on certain questions - consists of replacing missing data for an individual question with a value that is predicted based on other information available for the household or households in the survey. This is a “donor” imputation method, i.e. where another responding unit donates the missing information.

On this issue, it is highly recommended to implement other imputation methods which could guarantee the original variance present in the population, such as regressions, etc. A good suggested software is IVE-wave (see Raghunathan et al., 2001) or the Canadian Census Edit and Imputation System (CANCEIS).

Recommendation 5: Use “hot-deck” imputation methods for correcting for item non-response.

Two practical and implementable methods for performing “hot deck” imputation of item non-response is reported by the UNECE (2020) guide at pages 126-127. Examples are from EU-SILC in Albania (using IVE-ware) and Canada (using CANCEIS).

The step-by-step procedure used by the BNS for calculating the statistical weights does not take into account two important steps, namely calibration (or post-stratification) and trimming.

The calibration procedure is probably the most important step to face the problem of bias due to non-response, especially in a period of post-Covid pandemic situation.

In fact, unit non-response is not a completely at random phenomenon; instead, there are some specific household and individual characteristics, which determine higher or lower probabilities of non-response. The classical non-response adjustments based on territorial basis only (such as PSUs or groups of PSUs) does not guarantee to correct such bias.

The post-stratification or calibration procedures has the aim of considering the characteristics of households and individuals which are responsible of different non-response rates, and then post-stratify or calibrate for such characteristics.

Recommendation 6: Perform calibration or post-stratification.

In the middle term, calibration could be performed on the basis of the census when they will be available. UNECE (2020, pp. 131-133) report provides good recommendations on how to perform calibration or post-stratification.

Moreover, an example is provided by the calibration in the Canadian Income Survey.

As already mentioned, trimming or winsorisation refers to recoding of extreme weights to more acceptable values. The objective of trimming is to avoid excessive increase in variance due to weighting (the so-called Kish effect). It is important to realise that the process will introduce some bias. Even so, the aim is to seek a procedure which reduces the mean squared error. Though treatment of extreme or overly influential weights introduces some bias, the overall error may still be reduced.

Recommendation 7: Perform trimming on calibrated weights in order to avoid large increase in the variance or standard errors of final estimates.

Finally, one big concern when data collection is performed at regional level as well is the quality of data delivered by the regional offices.

One good practical method is to introduce the compulsory preparation of data quality reports at regional level, and then incorporate them in a national quality report.

Recommendation n. 8: Introduce a system of data quality reporting to be prepared at regional level, and the aggregated at central office level.

From this point of view, it is recommended to follow the UNECE (2020) guide, section 340:

“340. To establish trust in poverty measurement and prevent misguided policies, Statistical Offices have to regularly assess and continuously improve the quality of their processes and accuracy of their data. Quality reports which describe the quality criteria and explain any instances in which these criteria could not be met, or statistical concepts could not be correctly applied will not only facilitate the correct interpretation but can also provide the basis for future improvements”.

4. Data disaggregation on poverty measures

a. Description

The UNECE (2020) Poverty Measurement: Guide to Data Disaggregation, which follows the previous Guide on Poverty Measurement (UNECE, 2017) intended to provide further guidance to consolidate the measurement of poverty as well as inspiration for the practice of statistical offices. In particular, this requires that all population groups are counted, since people living in poverty are increasingly missed by household surveys, particularly if they belong to ethnic or other minorities whose life circumstances differ from the general population.

The intention of the this forth and short section is to encourage the BNS to implement as much as possible, the many recommendations available in the UNECE (2020) guide.

b. Analysis and findings

On light of the rich information reported in the UNECE (2020) guide, here we simply suggest taking into account the following disaggregation variables for all members of a household;

- Sex (target group of women and girls);
- Age (target groups of children, youth and older people);
- Disability status (target group of persons with disabilities);
- Migratory status (target group of migrant population);
- Ethnicity (target groups of ethnic minorities);

In addition, variables referring to socioeconomic and geographic strata may be considered as follows:

- Household type (characteristics of household composition);
- Educational attainment level (characteristics of qualification and social status); • Employment status (characteristics of labour force participation);
- Tenure status of the household (characteristics of an arrangement of occupancy of housing unit by a private household);
- Receipt of social transfers (characteristics of income composition);
- Degree of urbanisation (characteristics related to urban/rural composition).

These could lead to the final recommendation of this report:

Recommendation n. 9: In the disaggregation of data on poverty measurement, it is highly recommended – when possible – to follow the target variables proposed by the UNECE (2020) guide.

In particular, Chapter 2 on "Standard Core Variables for Disaggregation" is surely the best practice to follow by the BNS.

5. Conclusions and final recommendations

This report has been prepared under a consultancy conducted under the Poverty measurement stream of the United Nations Development Account 13th tranche project "Strengthening social protection for pandemic responses".

As well stated in the introduction, its main scope is to provide technical assistance to implement the recommendations related to assessing and improving survey methods of the UNECE (2020) Poverty measurement: Guide to Data Disaggregation.

In particular, the report objectives are focused in the improvement of the representativeness under conditions of increasing non-response, and in the advance the production of disaggregations of relevance to poverty and inequality, including in producing the corresponding SDGs indicators.

All the improvements proposed by this report and addressed to the BNS have been also translated in a series of recommendations.

For sake of practical implementation, these nine main recommendations are listed as follows:

Recommendation n. 1. It is recommended to adopt the new Deff = 1.24 and Deff = 1.89 for urban and rural areas respectively, for the calculation of the optimal sample sizes, and their allocation over regions in the future HBS sampling strategy.

Recommendation n. 2. As an example, it is recommended to explore the Eurostat web survey on the similar "Distribution of population by number of years spent in poverty within a four-year period".

Recommendation n. 3: It is recommended to transform the current design of the Household Budget Survey to one similar to the 4-year rotating panel of EU-SILC.

Recommendation n.4: Use larger areas for non-response adjustment. This will reduce the presence of extreme weights in some specific PSU and will reduce the current Kish effect (effect of weights), which determine larger variability in the estimates (standard errors).

Recommendation 5: Use "hot-deck" imputation methods for correcting for item non-response.

Recommendation 6: Perform calibration or post-stratification.

Recommendation 7: Perform trimming on calibrated weights in order to avoid large increase in the variance or standard errors of final estimates.

Recommendation n. 8: Introduce a system of data quality reporting to be prepared at regional level, and the aggregated at central office level.

Recommendation n. 9: In the disaggregation of data on poverty measurement, it is highly recommended – when possible – to follow the target variables proposed by the UNECE (2020) guide.

6. References

Eurostat (2020), METHODOLOGICAL GUIDELINES AND DESCRIPTION OF EU-SILC TARGET VARIABLES 2019 operation (Version February 2020), Luxembourg: Publications Office of the European Union.

Hasanov R. (2022), Living standard measurement survey: Household sampling methodology, translated from Russian.

Hasanov R., Hasanova L. (2020), Study of the practices of national statistical offices in adapting to the COVID-19 crisis situation their household surveys that are used for poverty measurement, Report to UNECE.

Hasanov R., Hasanova L. (2021), Analysis of household survey and poverty measurement in the Republic of Kazakhstan in the context of a pandemic, Report to UNECE.

Husmanns R., Mehran F., Verma V. (1990), Surveys of the Economically Active Population, Employment, Unemployment and Underemployment, International Labour Organisation, Geneva.

Kish L. (1965), *Survey Sampling*. Wiley.

ONS (2021), Data collection changes due to the pandemic and their impact on estimating personal wellbeing; UK: Office for National Statistics.

Raghunathan, T. E., Lepkowski, J. L., Van Hoewyk, J. H., Solenberger, P. W. (2001). A multivariate technique for imputing the missing values using a sequence regression models. In: Survey Methodology, 27: 85-95.

Ravallion M., Jalan J. (2000), Is transient poverty different? Evidence for rural China, Journal of Development Studies, 36(6), pp. 82-99.

UNECE (2017). Guide on Poverty Measurement. New York and Geneva: United Nations. Available from: https://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20174.pdf

Verma V. (1981), Assessment of errors in household surveys, *Bulletin of the International Statistical Institute*, **49(2)**, pp. 905-919.

Verma V., Betti G. (2006), EU Statistics on Income and Living Conditions (EU-SILC): Choosing the survey structure and sample design, *Statistics in Transition*, **7(5)**, pp. 935-970.

Verma V., Betti G. (2011), Taylor linearization sampling errors and design effects for poverty measures and other complex statistics, *Journal of Applied Statistics*, **38(8)**, pp. 1549-1576.

Verma V., Betti G., Gagliardi F. (2010), An assessment of survey errors in EU-SILC, Eurostat Methodologies and Working papers, Luxembourg: Publications Office of the European Union.

7. Annex

Table A.1 Sample size distribution in HBS, 2016

Ministry of National Economy of the Republic of Kazakhstan Committee on Statistics

Sample size

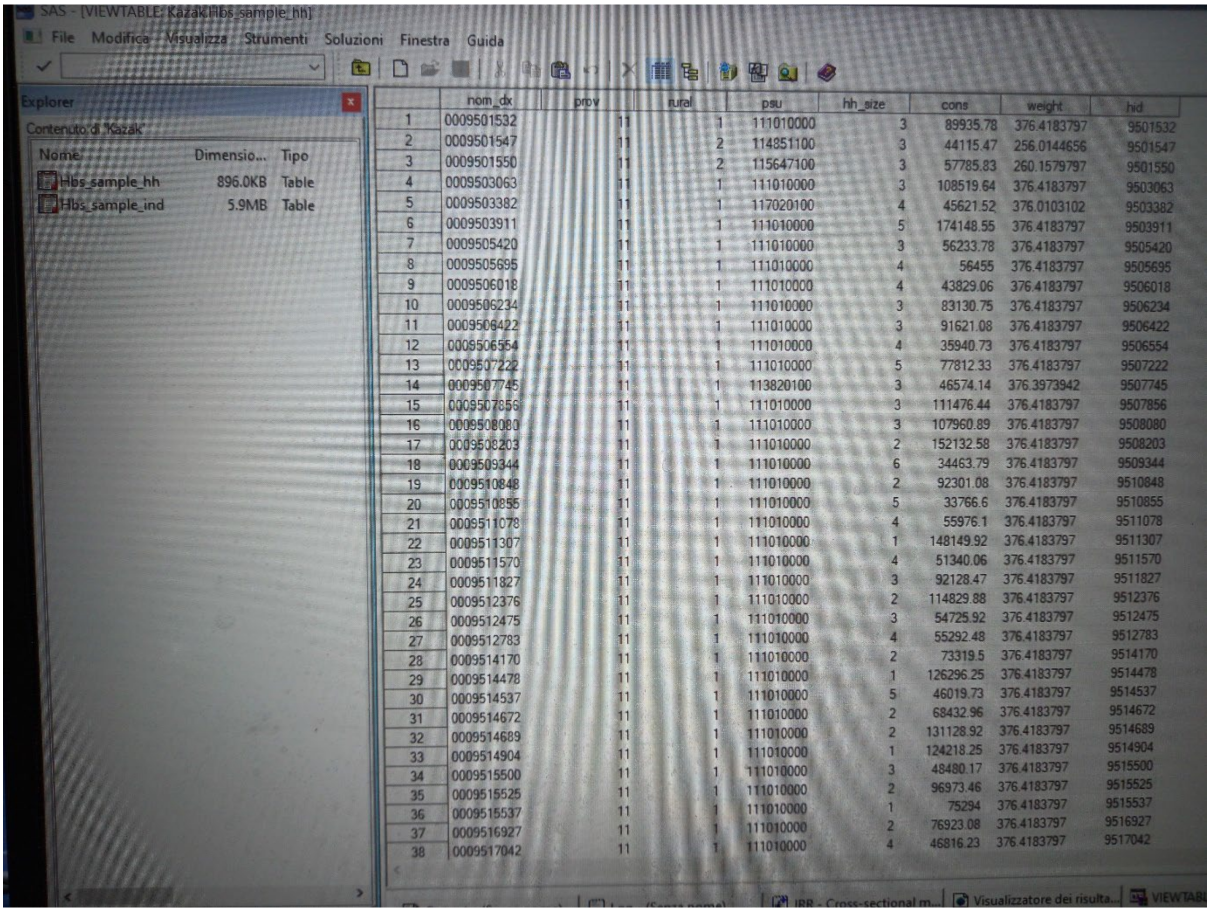
№	regions	№ of interviewers	including		№ of households	including	
			urban	rural		urban	rural
1	Akmolinskaya	28	12	16	840	360	480
2	Aktubinskaya	28	12	16	840	360	480
3	Almatinskaya	24	8	16	720	240	480
4	Atyrauskaya	18	10	8	540	300	240
5	East Kazakhstan	30	14	16	900	420	480
6	Zhambylskaya	23	9	14	690	270	420
7	West Kazakhstan	22	8	14	660	240	420
8	Karagandynskaya	32	20	12	960	600	360
9	Kostanayskaya	27	12	15	810	360	450
10	Kyzylordinskaya	20	8	12	600	240	360
11	Mangystauskaya	20	12	8	600	360	240
12	Pavlodarskaya	28	12	16	840	360	480
13	North Kazakhstan	22	9	13	660	270	390
14	South Kazakhstan	26	10	16	780	300	480
15	Almaty city	30	30	0	900	900	0
16	Astana city	22	22	0	660	660	0
17	**Total**	400	208	192	12000	6240	5760

Source: Ministry of National Economy of the Republic of Kazakhstan Committee of Statistics, presented on November 20-24, 2017 Rome, Italy

Table A.2 Distribution of population by number of years spent in poverty within a four-year period, EU-SILC 2017-2020

Country	1 year	2 years	3 years	4 years	Never
Belgium	8	6.5	4.9	6.4	74.1
Bulgaria	9	7.3	6.3	12.6	64.8
Czechia	6.7	3.4	2.7	2.5	84.7
Denmark	6.5	5.8	5.6	3.2	78.9
Estonia	11.1	8	6.4	9.4	65.1
Greece	13.9	7.2	7.2	7.6	64.2
Spain	8.3	6.3	9	11.4	65
France	10.4	4.9	5.4	5.9	73.5
Croatia	7.7	6.1	4.6	9.9	71.7
Cyprus	7.8	5.8	4.6	6.1	75.8
Latvia	9.6	7.8	5.6	8.6	68.5
Lithuania	8.2	6.6	7	13.1	65.2
Luxembourg	14.8	5	5.4	3.5	71.3
Hungary	16.9	9.6	3.1	2.4	67.9
Malta	10.1	7.2	4.8	7	70.9
Netherlands	6	3.1	4	7.3	79.5
Austria	9.6	4.8	3.3	5.5	76.8
Poland	11.6	6.4	4.8	4.8	72.3
Portugal	10.1	4.7	6	5.9	73.3
Romania	7.3	4.5	4.8	13.8	69.5
Slovenia	6.1	3.3	4.1	4.5	82
Slovakia	7.5	5.6	2	4.1	80.9
Finland	5.4	2.6	4.2	3.6	84.3
Sweden	5.2	3.7	3.3	3.2	84.6
Norway	8.9	5.3	3.8	3.3	78.7
Switzerland	12	5.4	5.3	4.5	72.8
Montenegro	11.3	7.7	9.4	10.9	60.7
Albania	11.7	8.1	9.2	8.3	62.7
Serbia	11.8	7.9	7.8	9.6	62.9
Turkey	11.8	7.8	6.8	9.6	63.9

Figure A.1 SAS data set for estimating the design effect (Deff) in HBS



	nom_dx	prov	rural	psu	hh_size	cons	weight	hid
1	0009501532	11	1	111010000	3	89935.78	376.4183797	9501532
2	0009501547	11	2	114851100	3	44115.47	256.0144656	9501547
3	0009501550	11	2	115647100	3	57785.83	260.1579797	9501550
4	0009503063	11	1	111010000	3	108519.64	376.4183797	9503063
5	0009503382	11	1	117020100	4	45621.52	376.0103102	9503382
6	0009503911	11	1	111010000	5	174148.55	376.4183797	9503911
7	0009505420	11	1	111010000	3	56233.78	376.4183797	9505420
8	0009505695	11	1	111010000	4	56455	376.4183797	9505695
9	0009506018	11	1	111010000	4	43829.06	376.4183797	9506018
10	0009506234	11	1	111010000	3	83130.75	376.4183797	9506234
11	0009506422	11	1	111010000	3	91621.08	376.4183797	9506422
12	0009506554	11	1	111010000	4	35940.73	376.4183797	9506554
13	0009507222	11	1	111010000	5	77812.33	376.4183797	9507222
14	0009507745	11	1	113820100	3	46574.14	376.3973942	9507745
15	0009507856	11	1	111010000	3	111476.44	376.4183797	9507856
16	0009508080	11	1	111010000	3	107960.89	376.4183797	9508080
17	0009508203	11	1	111010000	2	152132.58	376.4183797	9508203
18	0009509344	11	1	111010000	6	34463.79	376.4183797	9509344
19	0009510848	11	1	111010000	2	92301.08	376.4183797	9510848
20	0009510855	11	1	111010000	5	33766.6	376.4183797	9510855
21	0009511078	11	1	111010000	4	55976.1	376.4183797	9511078
22	0009511307	11	1	111010000	1	148149.92	376.4183797	9511307
23	0009511570	11	1	111010000	4	51340.06	376.4183797	9511570
24	0009511827	11	1	111010000	3	92128.47	376.4183797	9511827
25	0009512376	11	1	111010000	2	114829.88	376.4183797	9512376
26	0009512475	11	1	111010000	3	54725.92	376.4183797	9512475
27	0009512783	11	1	111010000	4	55292.48	376.4183797	9512783
28	0009514170	11	1	111010000	2	73319.5	376.4183797	9514170
29	0009514478	11	1	111010000	1	126296.25	376.4183797	9514478
30	0009514537	11	1	111010000	5	46019.73	376.4183797	9514537
31	0009514672	11	1	111010000	2	68432.96	376.4183797	9514672
32	0009514689	11	1	111010000	2	131128.92	376.4183797	9514689
33	0009514904	11	1	111010000	1	124218.25	376.4183797	9514904
34	0009515500	11	1	111010000	3	48480.17	376.4183797	9515500
35	0009515525	11	1	111010000	2	96973.46	376.4183797	9515525
36	0009515537	11	1	111010000	1	75294	376.4183797	9515537
37	0009516927	11	1	111010000	2	76923.08	376.4183797	9516927
38	0009517042	11	1	111010000	4	46816.23	376.4183797	9517042