



UNECE

**UNITED NATIONS ECONOMIC COMMISSION FOR
EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS
Workshop on Harmonization of Poverty Statistics
to Measure SDG 1 and 10**



**UNIVERSITÀ
DI SIENA**
1240

Improving survey methods in Kazakhstan

Gianni Betti

UNECE Consultant, University of Siena, Italy, gianni.betti@unisi.it

Geneva, December the 7th, 2022



Table of contents

- **Introduction**
- **Analysis of the sampling methods currently employed by the Bureau of National Statistics of Kazakhstan**
- **Addressing the quality of data: issues of non-response and coverage errors**
- **Data disaggregation on poverty measures**
- **Options for longitudinal survey 2023 - 2030**
- **Key messages for other NSOs**
- **References**



Introduction

This presentation summarise a report prepared under a consultancy conducted under the Poverty measurement stream of the United Nations Development Account 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), which was the main topic of the UNECE workshop of December 2019.



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

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)

At the **first stage**, the general population is stratified according to the regional distribution, including the distribution into urban and rural areas.

In the **second stage**, a certain number of households are then selected from each sampled PSU.



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

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.

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.

Another interesting aspect of the sampling design undertaken by the BNS for the HBS is the longitudinal dimension.



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

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. These have been estimated on the basis of Verma and Betti (2011) method.

Recommendation n. 2. It is suggested to better analyse such rich longitudinal data set by studying 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).

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).



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

Non-response

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.



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

The non-response adjustment is performed at PSU level, where PSUs are areas with small size in Kazakhstan

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 (1988) effect (effect of weights), which determine larger variability in the estimates (standard errors).

Recommendation n. 6: Perform calibration or post-stratification for facing coverage errors.



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

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.



Data disaggregation on poverty measures

- 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);



Data disaggregation on poverty measures

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).



Longitudinal survey 2023 – 2030 – option 1

Option 1 represents a variant of the EU-SILC 4-year rotating panel survey adapted to Kazakhstan peculiarities

	2023	2024	2025	2026	2027	2028	2029	2030
25%	1	2	3	4	1	2
25%	4	1	2	3	4	1	2	...
25%	3	4	1	2	3	4	1	2
25%	2	3	4	1	2	3	4	1



Longitudinal survey 2023 – 2030 – option 2

With the following option 2 there is the risk that percentage of families (in red) can be observed for more than four year could reach even 50%

Схематично виглядит наступним образом:

2023	2024	2025	2026	2027	2028	2029
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9
10	10	10	10	10	10	10
11	11	11	11	11	11	11
12	12	12	12	12	12	12
percentage of updates:	25%	25%	25%	25%	50%	



Key messages for other NSOs - 1

- 1 - Document coverage, precision and any departure from international standards in quality reports to be issued with each data release.
- 2 - Strictly adhere to probability sampling and refrain from substitution.
- 3 - Adjust sample design and allocation for the required disaggregation. As a minimum the sample size for each group should never be below 50 units. Ideally estimation domains will be used as strata so that their size may be controlled.



Key messages for other NSOs - 2

- 4 - Ensure a sufficient number of contact attempts (e.g.), use group translations, and interviewer's addresses. Field guidelines to be further elaborated.
- 5 - Compensate for non-response and random variation by adjustments to external control distributions (calibration). Ensure that these adjustments (g-weights) remain within reasonable limits. (e.g. 1/3 to 3).
- 6 - Ensure that vulnerable participate in the design of the study and are protected from any harm and can benefit from its results.



Key messages for other NSOs - 3

7 - Minimize the amount of missing information. Because poverty status is usually derived from several variables that have to be observed simultaneously it is essential that group representation is preserved by adequate and transparent imputation rules.

8 - Carefully assess the variance of all key estimates. Publish coefficients of variation in quality reports so that precision can be compared to other surveys and over time. As a minimum present an estimates of the design effect of the poverty rate.

9 - Establish filtering rules for publications that eliminate unreliable estimates. Never publish results based on less than 20 observations or with a correspondingly high coefficient of variation.



References

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

Kish, L. (1988), Multipurpose sample designs. Survey Methodology, 14, 19-32.

UNECE (2020), Poverty Measurement: Guide to Data Disaggregation, Geneva.

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.



UNIVERSITÀ
DI SIENA
1240

**THANK YOU
FOR YOUR
ATTENTION !!!**