rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2
2020

In this issue:

Improving quality and containing respondents' burden in socio-demographic surveys: the case of intra-household relationships
Carolina Facioni, Luciano Fanfoni, Martina Lo Conte, Stefania Macchia, Paolo Piergentili, Luciana Quattrociocchi, Marco Scuderi

Implementing ICD-11 for mortality statistics: translation of decision tables embedded in the automated coding system Iris
Chiara Orsi, Daniele De Rocchi, Mihai Horia Popescu, Friedrich Heuser, Stefanie Weber, Luisa Frova, Vincenzo Della Mea, Francesco Grippo

Optimal Design of a Master Sample: the Case of the Household Surveys in the Republic of Moldova
Giulio Barcaroli, Loredana Di Consiglio, Alessio Guandalini, Marco Dionisio Terribili

Istat
Istituto Nazionale
di Statistica

rivista di statistica ufficiale

REVIEW OF OFFICIAL STATISTICS

n. 1-2
2020

In this issue:

Istat

rivista di statistica ufficiale
n. 1-2/2020

Four-monthly Journal: registered at the Court of Rome, Italy (N. 339/2007 of 19th July 2007).

e-ISSN 1972-4829
p-ISSN 1828-1982

The Scientific Committee, the Editorial Board and the authors would like to thank the anonymous reviewers (at least two for each article, on a voluntary basis and free of charge, with a double-anonymised approach) for their comments and suggestions, which enhanced the quality of this issue of the Rivista di statistica ufficiale.

Editorial Preface

This double issue of the *Rivista di statistica ufficiale* publishes three extended scientific articles dealing with topics related to quality improvement and development of the use of standards in the production of official statistics, focussing on the context of international cooperation projects and sharing of good practices.

The first article (by Carolina Facioni, Luciano Fanfoni, Martina Lo Conte, Stefania Macchia, Paolo

Piergentili, Luciana Quattrociocchi and Marco Scuderi) illustrates the main results of an experimental survey carried out by the Italian National Institute of Statistics - Istat.

It was aimed at testing a new approach for collecting data on socio-demographic variables, in particular on household relationships based on a grid proposed by Eurostat.

Cognitive interviewing was used to investigate both on possible critical issues and on respondents' perception in terms of burden. With the purpose of modernising social surveys and the Population Census, as well as in order to take into account the recent regulation changes that introduced same-sex couples in Italy, this new approach meets the dual need of containing the respondents' burden and offering higher quality statistical information.

The tested method, indeed, modifies completely the traditional way of collecting information about household connections: this is no longer detected with regard to a sole Reference Person, but for each member in relation to all the others.

The results obtained confirm the feasibility of adopting this new approach for all socio-demographic surveys and represent a significant contribution to the reduction of some uncertainties in identifying family nuclei, without neither having a relevant effect on the interview length, nor increasing the respondents' burden.

The authors of the second article (Chiara Orsi, Daniele De Rocchi, Mihai Horia Popescu, Friedrich Heuser, Stefanie Weber, Luisa Frova, Vincenzo Della Mea and Francesco Grippo) instead describe how the causes-of-death coding system *Iris* is adapted in order to face the passage from 10^{th} to 11^{th} revision of the International Classification of Diseases – ICD.

Iris contains a knowledge base composed by a large number of rules consisting in relations between ICD-10 codes. Due to the adoption of the new revision, since 2019 a pilot project has been carried out in order to evaluate the feasibility of the translation of each rule within *Iris* into ICD-11.

The main findings of this project suggest on the one hand that part of these rules can be automatically translated; on the other hand, that most of the rules need to be revised manually, identifying those which have to be prioritised in this translation process, due to their relevance for coding real data.

Finally, the third article (by Giulio Barcaroli, Loredana Di Consiglio, Alessio Guandalini and Marco Dionisio Terribili) shows the results of the international cooperation activities carried out from December 2019 to June 2020, in the framework of a project of technical assistance to support the National Bureau of Statistics – NBS of the Republic of Moldova.

More specifically, these activities concerned all the phases related to the design of the Master Sample for its application to four surveys to be carried out during the period 2021-2025: Labour Force Survey; Household Budget Survey; Energy Consumption Survey; Domestic Violence Survey.

This commitment had to take into account the constraints of the current organisation of the data collection network within the NBS, and involved the use of the methodologies implemented in three generalised software (*ReGenesees*, *R2BEAT* and *FS4*) developed and adopted by the Italian National Institute of Statistics - Istat.

Patrizia Cacioli
Editor

Nadia Mignolli
Coordinator of the Editorial board

Improving quality and containing respondents' burden in socio-demographic surveys: the case of intra-household relationships

Carolina Facioni, Luciano Fanfoni, Martina Lo Conte, Stefania Macchia, Paolo Piergentili, Luciana Quattrociocchi, Marco Scuderi [1]

Abstract

The paper shows the main results of an experimental survey carried out by the Italian National Institute of Statistics - Istat. The aim was testing a new approach for collecting data on socio-demographic variables, particularly on household relationships, managed by a household grid, as proposed by Eurostat.
Cognitive interviews were used to investigate on possible critical issues and on respondents' perception in terms of burden. Following the attempt to modernise social surveys and Population Census, as well as the recent regulation changes introducing same-sex civil unions in Italy (same-sex couples), this new approach meets the dual need to contain the respondents' burden, and to offer higher quality statistical information. The tested approach completely reverses the traditional way of collecting intra-household relationships: this information is no longer asked in relation to a unique Reference Person, but for each family member in relation to all the others. The results of the test confirmed the feasibility of adopting this new approach for all socio-demographic surveys. As a matter of fact, even if some aspects need to be improved, it was showed that it contributes to reduce the ambiguity in identifying family nuclei, without having a relevant effect on the interview length and with no increase of the perceived respondents' burden.

Keywords: Cognitive interviewing, family relationships, household grid, socio-demographic surveys.

1 Carolina Facioni (facioni@istat.it); Luciano Fanfoni (fanfoni@istat.it); Martina Lo Conte (loconte@istat.it); Stefania Macchia (macchia@istat.it); Paolo Piergentili (piergentili@istat.it); Luciana Quattrociocchi (quattroc@istat.it); Marco Scuderi (scuderi@istat.it), Italian National Institute of Statistics - Istat.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[2]

Socio-demographic surveys play a leading role in documenting relevant family transformations taking place in Italy. Policy makers need analysis on family relationships when planning and implementing economic and social policy interventions. The nature of these relationships can be jurisprudential, as well as economic, biological, social, emotional, or simply residential, and can include one or more of these aspects. In order to satisfy this fact-finding need, the Italian National Institute of Statistics – Istat carries out a wide range of socio-demographic surveys, the largest and best known being the Population Census.

Data collection on family composition generally precedes the sections of the questionnaires dealing with the survey subject. These opening questions represent therefore an extremely delicate moment, both for winning the possible fear of invasion of privacy and for guaranteeing the reliability of the information collected.

The current official socio-demographic surveys, being designed at different times and with different needs, are characterised by a lack of harmonisation in the way of collecting data and in the classifications used for the kinship variable, thus creating great difficulty for anyone wanting to use, analyse and compare data on family composition from these surveys.

Furthermore, up to now in socio-demographic surveys, family relationships have been observed with respect to a Reference Person (RP), usually an adult family member or the holder of the family form in the Demographic Registers. Despite the 23 categories of the relationship classification (RP23), identifying all bilateral relationships between family members is not always possible when families have more than one nucleus. Some family structures may therefore be incorrectly registered. In Italy, for Eurostat regulated surveys, this has brought about a need to use additional questions in some surveys, such as the Labour Force Survey (LFS), to retrace all family types required on a supra-national level.

2 This article is the result of the commitment of all authors: Section 1 by Luciana Quattrociocchi; Section 2 by Stefania Macchia (excluding 2.1.1 by Luciano Fanfoni); Paragraph 3.1 by Martina Lo Conte; Paragraph 3.2 by Paolo Piergentili and Marco Scuderi; Paragraph 3.3. by Stefania Macchia; Paragraph 3.4 by Carolina Facioni; Paragraph 4 is a collaboration of all the authors.
The authors thank Gabriella Catapano, Annagrazia Melatti and Simona Rosati, who carried out the interviews and provided valuable comments for the optimisation of the tested procedure.

Following the attempt of modernise social surveys and Population Census, as well as the recent regulation changes introducing same-sex civil unions in Italy (same-sex couples), Istat has given a high priority to the standardisation of variables concerning kinship and family composition for all surveys integrated in the Master Sample[3]. In this context, it was thought to experiment the introduction of the Household Grid (HHG) proposed by Eurostat (European Commission, 2009). Such approach for collecting data on intra-household relationships is very different from the traditional one: information for all family members is no longer asked in relation to a RP, but for each single component in relation to all the others. This allows to use a more agile classification (with only 16 categories instead of 23) and also to simplify item descriptions, as reported below, in order to eliminate possible ambiguities impacting on the family structure identification. The Household Grid approach can respond, therefore, to the needs not only to facilitate survey harmonisation, but also to contain respondents' burden and to produce a high quality statistical information.

The aim of the test presented here was to assess the feasibility of this new approach for the Italian socio-demographic surveys. This paper describes the methods adopted (Section 2) and the main results of the experiment, evaluated, as to the respondents' burden, in terms of interview length and of easiness of compilation, and, from the quality perspective, in terms of the correct identification of household structures (Section 3). Finally, Section 4 summarises some conclusions and perspectives concerning the feasibility of the adoption of this new approach.

3 Master Sample integrates a set of balanced and coordinated sample surveys (Continuous census, Labour force survey, Aspects of daily life survey, Eu-Silc) in the context of the Population Census and Social Surveys Integrated System (CSSIS).

2. Methods

The test was based on a small experimental survey aimed at collecting, through the new approach, the main socio-demographic information of each household member and all the intra-household kinship relations (variables usually collected in social surveys within the 'General Household Information section').

The interviews were followed by retrospective cognitive questions to investigate on possible critical issues and on respondents' perception in terms of burden. Cognitive interviewing has emerged as one of the most prominent methods for identifying and correcting problems with survey questions. This method is used to investigate the response process, so as to determine whether the questions are generating the information that the authors intend (Sudman *et al.,* 1996). Such interviews could consist of respondents' elaborations regarding how they constructed their answers, explanations on what they interpret the question to mean, reports on any difficulties they had answering and anything else that sheds light on the circumstances that their answers were based upon (Beatty and Willis, 2007). Interviewers are generally involved: they can have a minimal intervening role in a 'thinking aloud process', or can interact asking direct questions according to a scripted protocol. This second type of practice was applied, as it is detailed in Paragraph 2.2.

Three data collection techniques were used in the experiment: PAPI (Paper and Pencil Interviewing) administered by an interviewer, CATI (Computer Assisted Telephone Interviewing) and CAWI (Computer Assisted Web Interviewing), in order to replicate the strategies adopted for the main Italian socio-demographic surveys.

Also for the self-administered interviews (CAWI), the presence of a non-participating observer, documenting uncertainties or problems in answering, was considered useful. Therefore, CAWI respondents were asked to fill in the questionnaire in a centralised location, the Istat CAI (Computer Assisted Interviewing) laboratory at the presence of an interviewer.

Generally, cognitive interviewing samples are not designed to be representative of any population. However, it is recommended to take into account demographic variety of respondents and to include people relevant to the topic of the questionnaire being tested (Willis, 1994, 2005). In other words, whatever topic the questions focus on, the sample should cover a variety of situations relevant to that topic (Beatty and Willis, 2007).

For the sample selection, therefore, two constraints were taken into account: firstly, to cover different types of households mainly in terms of household size; secondly to contain costs in terms of time and resources.

Given these restrictions, the test was run completely in-house, interviewing Istat colleagues characterised by different household types. To avoid bias due to the working experience on this topic, only employees working in the General Directorate and in the IT Department – and therefore not having experience on socio-demographic surveys – were considered.

The sample selection was based on household size, gender, age and education: 129 individuals were extracted from the total employees considered (338), with the aim to obtain about 100 interviews, homogeneously divided among the three techniques.

2.1 The questionnaire design

The questionnaire consisted of three sections: the first one, about the number of household members, was followed by a section collecting some demographic information on each member, such as gender, date of birth, marital status, *etc.* (General Household Information, GHI). These variables are a subset of those considered in the GHI section of most socio-demographic surveys. The choice of such variables was made on the assumption that moving some questions (for instance, place of birth, citizenship or labour status) in the individual questionnaires could be more efficient.

The third and final section concerned the intra-household relationships, for each single member in relation to all the others, collected through the household grid (HHG), which was the core of the test.

Figures 1 and 2 show the paper version of the GHI and the HHG sections.

Figure 1 - General Household Information Section (GHI)

Order number of household members	A First Name *Write in capital letters*	B Surname *Write in capital letters*	D Gender 1=M 2=F	E Date of birth *dd/mm/yyyy*	F Age in completed years *(to be filled in if you do not remember the date of birth)*	G Marital status *(only if age ?14 years)*	H Year of marriage or of civil union *(only if married or in civil union)*	I Marital status before last marriage /civil union *(only if married or in civil union)*
*********	*****************	*************	*******	***************	**************	*******	**************	**************
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Source: Istat, Questionnaire of the 2018 Experimental survey on household grid

Figure 2 - Household grid with intra-household relationships (HHG)

<table>
<tr><th rowspan="2">Order number of household members</th><th rowspan="2">Name or initials
Write in capital letters</th><th colspan="10">Name or initials</th></tr>
<tr><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td>**********</td><td>**************</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>1</td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>2</td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>3</td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>4</td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>5</td><td></td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>6</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>7</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>8</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td><td>|_|_|</td></tr>
<tr><td>9</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>|_|_|</td></tr>
<tr><td>10</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="12">2</td></tr>
</table>

Source: Istat, Questionnaire of the 2018 Experimental survey on household grid

In practice, in the household grid only the colored information needs to be asked (Figure 2); once the components' names are known, the responses have to be written row by row: Who is *'component 1'* in relation to *'component 2'*?, Who is *'component 1'* in relation to *'component 3*?', and so on.

In this way, even though the number of questions is higher compared to the traditional approach, the respondent's burden is supposed to be lower, since the number of items is smaller respect to the previous RP23 classification and the corresponding descriptions are simpler (Figure 3).

Figure 3 - Intra-household relationships items used in the test

INTRA-HOUSEHOLD RELATIONSHIPS	
01 Husband/Wife	09 Stepbrother/Sister (with both different parents)
02 Partner in civil union	10 Son/Daughter-in-law
03 Partner/Cohabitée	11 Brother/Sister-in-law
04 Son/Daughter	12 Father/Mother-in-law
05 Stepson/daughter	13 Grandparent
06 Parent	14 Grandchild
07 Stepparent	15 Other relative (not included in the list)
08 Brother/Sister	16 Other non-relative

Source: Istat, Questionnaire of the 2018 Experimental survey on household grid

Using the CATI and CAWI electronic questionnaire, the interview becomes even easier, since each information requested corresponds to a question and the grid is transformed into a series of questions customised with each member's name (Figure 4).

Figure 4 - Series of questions for intra-household relationships in electronic questionnaire

The relationships among the family members are requested in the following section.

Who is *&Name 1* for *&Nome 2?* ☐ (A box is opened with the list of items for intra-household relationships)

and *&Name 1* for a *&Name 3?* ☐
and *&Name 1* for *&Name n?* ☐

Let's talk about *&Name 2*

Who is *&Name 2* for *&Nome 3* ☐
and &*Name 2* for *&Name n* ☐

Let's talk about *&Name 3*

…..

Source: Istat, Questionnaire of the 2018 Experimental survey on household grid

2.1.1 The electronic questionnaire

The IT tool used for the implementation of the electronic questionnaire is LimeSurvey, an open source software. The same electronic questionnaire was used for all the techniques (CAWI self-administered, CATI for telephone interviews and CADE, Computer Assisted Data Entry, for the PAPI questionnaires). It was structured in three sections:

- the first one asks for the number of 'household members', after providing the necessary definitions;
- the second one regards the loop of questions to collect the information of the GHI section;
- the last one contains the series of questions reproducing the household grid. All the questions are customised with the components' names already collected in the previous section.

A set of rules have been implemented to prevent from non-response and consistency errors, similar to those used in socio-demographic surveys that adopt computer-assisted techniques. In details, consistency rules have been managed with two different approaches:

- for the consistency between HHG relationships and the information collected in the GHI section, the list of intra-household relationships displayed for each component was subject to the answers given in the GHI section. The following examples may help clarify:
  - ✓ if the age difference between the two persons involved in the relationship is less than 14, the item 'mother/father' was not displayed;
  - ✓ if two persons have not declared to be married in GHI Section, the item 'wife/husband' was not displayed.
- for the consistency among other data regarding different components, error messages were displayed after giving the 'inconsistent' response (for example, in case two members declared to be married with the same person, or when members with the same gender (same-sex couples) declared to be 'mother/father' of the same person).

2.2 The cognitive test

After the interview, a set of cognitive questions were asked by interviewers. A probing-based paradigm was used, instead of the think-aloud one.

In general, both paradigms aim at generating verbal information that is usually unseen in a survey interview, in order to evaluate how well the questions are meeting their objectives (Beatty and Willis, 2007). Literature shows that there are advantages and disadvantages for both approaches. On one hand, an interviewer probing could introduce bias into the data collection process and can create artificiality (Conrad, Blair and Tracy, 2000). On the other hand, probing does not interfere with the actual process of responding (since it intervenes after the questionnaire but still capturing information stored in short-term memory), while thinking aloud might, since participants have to provide verbal information during the response process. This can also increase the effort spent on creating a response which has an unknown impact on the real answer (Willis 1994). However, the probing, when used, should involve only a few questions per interview (Oksenberg, Cannell and Kalton, 1991).

Another important factor to be decided in a cognitive test is whether it should be standardised or determined by the interviewer judgement, and to what extent (Willis, 2005; Presser *et al.*, 2004).

The cognitive questionnaire used for the test was standardised and administered by the interviewer. It only had a few questions, starting with the easiness/difficulty in identifying the pertaining intra-household kinships and, in case of difficulty, for which relationship this had been encountered. Moreover, it was asked whether it had been necessary to read again or ask the interviewer to repeat some concepts and, if so, which ones. A deepening was carried out to check whether the options ‘Other relatives’ and ‘Other non-relatives’ had been selected correctly. Finally, the respondents were asked to assign a score from 1 to 10 to judge easiness/directness in giving the answer, with the aim of getting an overall feedback on the household grid approach. Lastly, respondents’ suggestions/proposals were also recorded.

3. Results

3.1 The survey results

The test was run between May 29 and June 5, 2018. A first letter was sent by email to 338 Istat employees of the General Directorate and the IT Department in order to illustrate the purpose of the experiment and to inform that a sample of them would be selected. Then, the 129 sampled colleagues were contacted by telephone to fix an appointment for the interview and to let them know the data collection mode they would be interviewed with.

Participation rates were very high, due to a good cooperation of the colleagues. The total response rate was 86%: CAWI interviews registered the most positive result with almost 91% of complete interviews, while the lowest rate was found for the PAPI mode (77.5 %) (Table 1).

Table 1 - Contacts results by data collection technique

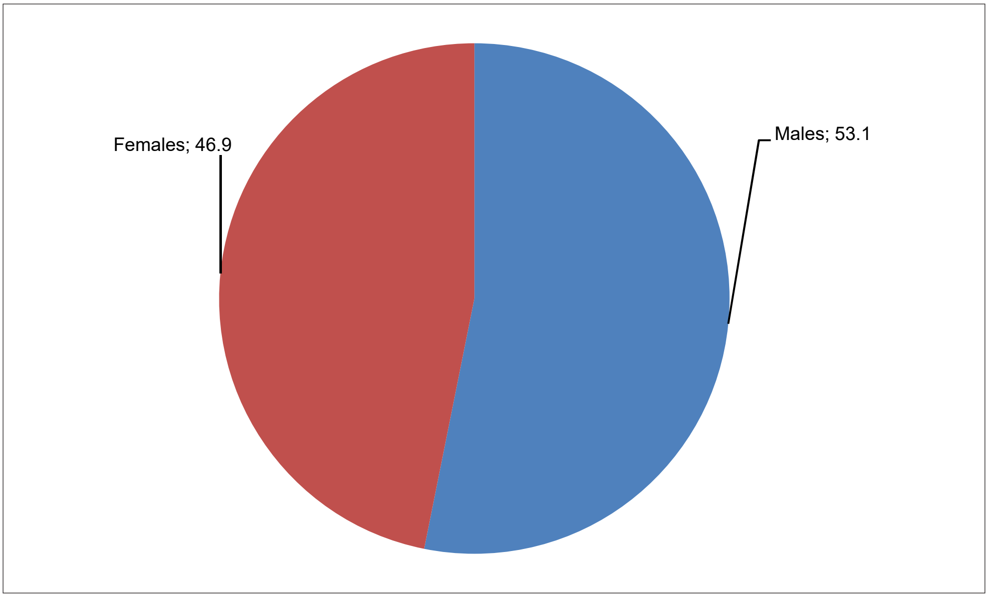
CONTACTS RESULTS	Data collection technique			Total
	CATI	CAWI	PAPI	
Complete	41	39	31	111
Refusal	2	2	3	7
No answer	3	1	5	9
Appointment	0	1	1	2
Total	46	43	40	129
RATES				
Response	89.1	90.7	77.5	86.0
Refusal	4.3	4.7	7.5	5.4
Other no answer	6.5	4.7	15.0	8.5

Source: Istat, Experimental survey on household grid - 2018

Respondents were equally distributed by gender, with a slight prevalence of males (47% of women and 53% of men). Almost half of respondents (45%) were aged between 40 and 49, about one out of five was over 60 and only 7% under 40 (Figures 5 and 6).

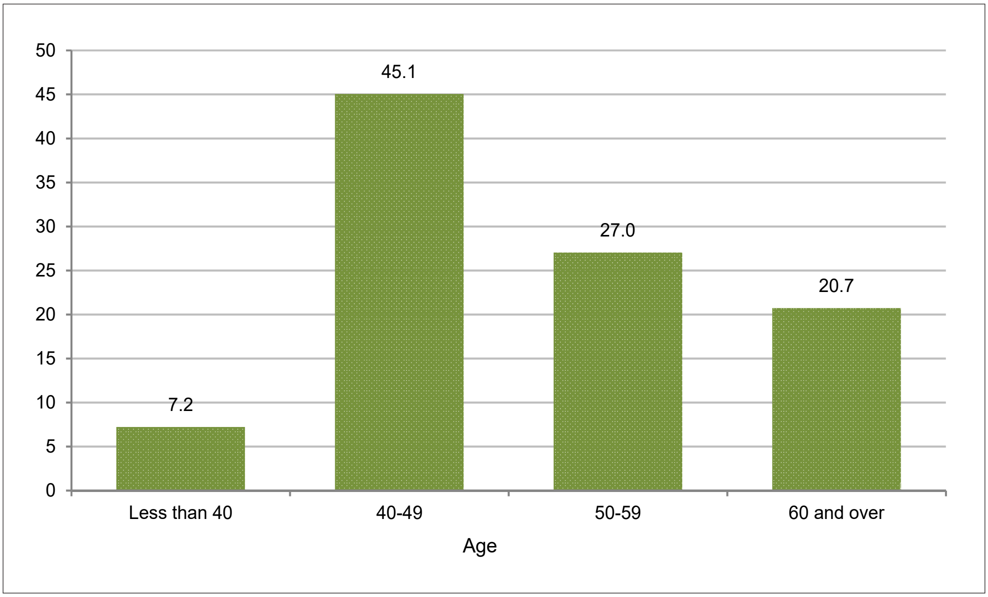
As for marital status, the prevalence of respondents was married (with 64%), followed by 14.4% singles (Figure 7). None of the interviewees was 'in civil union' or 'previously in civil union'.

Figure 5 - Respondents by gender



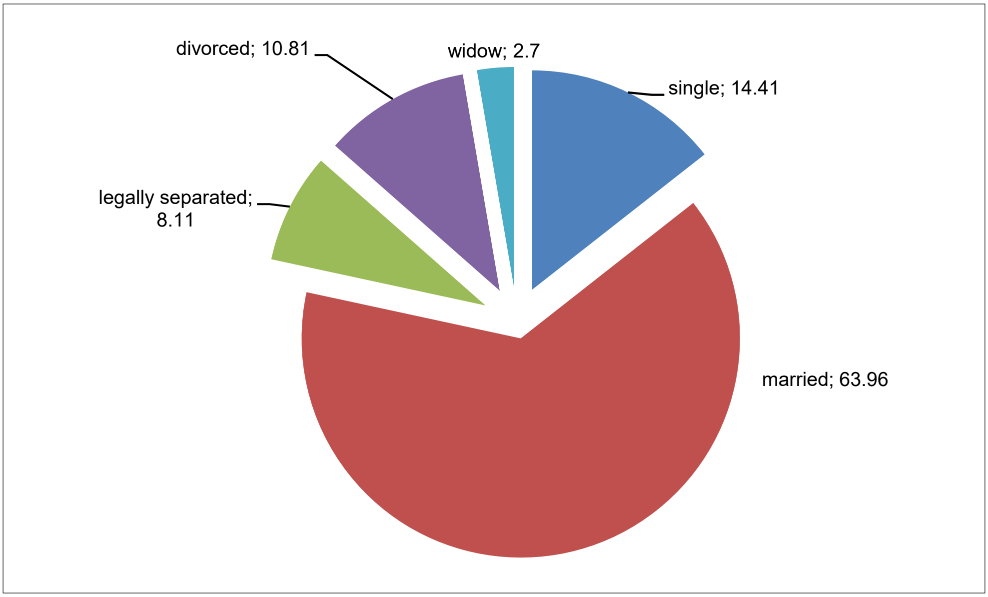
Source: Istat, Experimental survey on household grid - 2018

Figure 6 - Respondents by age



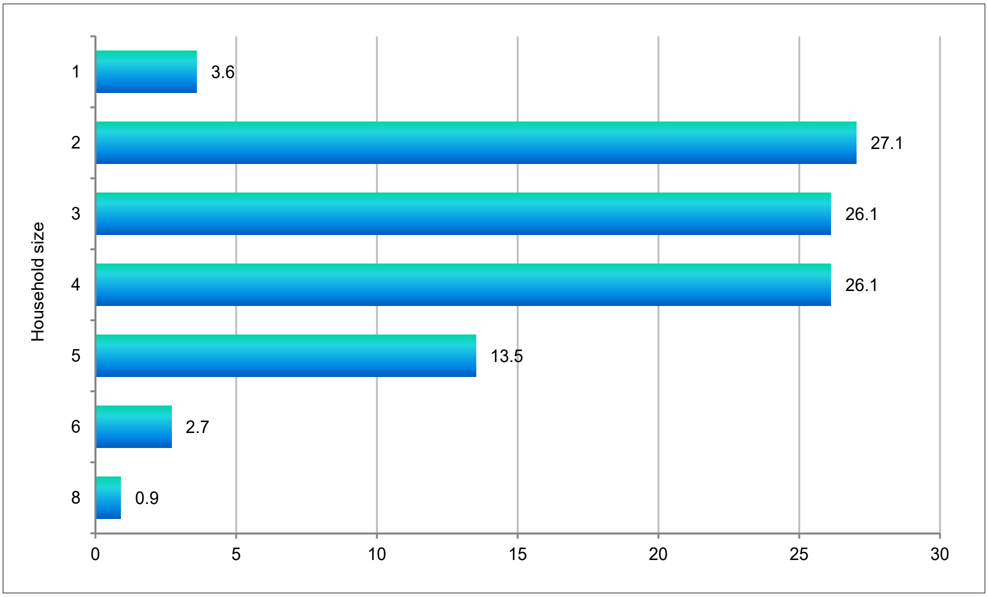
Source: Istat, Experimental survey on household grid - 2018

Figure 7 - Respondents by marital status



Source: Istat, Experimental survey on household grid - 2018

Figure 8 - Respondents by household size



Source: Istat, Experimental survey on household grid - 2018

Table 2 - Intra-household relationships (number of times that items were selected)

Intra-household relationship	N.	%
01 Husband/Wife	75	14.6
02 Partner in civil union	0	0.0
03 Partner/Cohabitée	20	3.9
04 Son/Daughter	18	3.5
05 Stepson/daughter	8	1.6
06 Parent	254	49.3
07 Stepparent	8	1.6
08 Brother/Sister	92	17.9
09 Stepbrother/Sister (with both different parents)	0	0.0
10 Son/Daughter-in-law	10	1.9
11 Brother/Sister-in-law	5	1.0
12 Father/Mother-in-law	5	1.0
13 Grandparent	0	0.0
14 Grandchild	6	1.2
15 Other relative (not included in the list)	8	1.6
16 Other non-relative	6	1.2
TOTAL	**515**	**100.0**

Source: Istat, Experimental survey on household grid - 2018

The interview length represents a direct measure of response burden (Bradburn, 1978; Frankel and Sharp, 1981; Sharp and Frankel, 1983) and it is therefore a good indicator of the applicability of the HHG approach.

Table 3 shows the average interview length, distinguishing for the 3 sections only for CATI and CAWI (recorded by Limesurvey, which allows automatic registration of the compilation times).

Table 3 - Interview length per technique and per section (minutes and seconds)

			CATI		
Interview lenght	**N.**	**Mean**	**Dev std**	**Min**	**Max**
TOTAL	41	4.10	2.17	0.50	10.23
Section N. of components	41	0.35	0.36	0.8	3.37
GHI Section	41	2.34	1.10	0.37	5.31
HHG Section	41	1.1	1.6	0.0	5.42
			CAWI		
Interview length	**N.**	**Mean**	**Dev std**	**Min**	**Max**
TOTAL	**39**	**5.11**	**2.56**	**1.32**	**15.59**
Section N. of components	39	0.39	0.26	0.11	1.58
GHI Section	39	3.8	1.52	0.50	9.43
HHG Section	39	1.25	1.8	0.12	5.34
			PAPI		
Interview length	**N**	**Mean**	**Dev std**	**Min**	**Max**
TOTAL	31	5.16	1.40	2.0	9.0
Section N. of components	31	-	-	-	-
GHI Section	31	-	-	-	-
HHG Section	31	-	-	-	-

Source: Istat, Experimental survey on household grid - 2018

Overall, the compilation took about 4-5 minutes. The shortest interview was obtained through the CATI mode (on average 4'10''), followed by CAWI (5' 11'') and PAPI (5' 16'').

Using CATI, the first section (number of members) lasted on average 35'', the GHI Section 2'34''and completing the household grid involved about 1 minute. On the other hand, the CAWI self-compilation took on average 3'8'' for the GHI section and 1'25'' for the household grid.

Note that the compilation time for the first section depends mainly on the understanding of the explanations provided for concepts such as the household and who to include as members.

A factor that greatly affected the interview length was the household size. Clearly, the compilation time grew along with the number of members, due to the increase of the information requested.

Table 4 - Interview length (minutes and seconds) per technique, section and number of components (in brackets the number of households interviewed)

Data collection technique	Number of components	Interview Lenght			
		TOTAL	Number of compontents Section	GHI Section	HHG Section
CATI	1	1.19 (3)	0.25 (3)	0.54 (3)	0.0 (3)
	2	2.40 (11)	0.40 (11)	1.41 (11)	0.19 (11)
	3	3.23 (11)	0.24 (11)	2.17 (11)	0.42 (11)
	4	6.1 (12)	0.44 (12)	3.35 (12)	1.42 (12)
	5	6.54 (4)	0.32 (4)	3.55 (4)	2.26 (4)
CAWI	2	3.27 (11)	0.49 (11)	2.7 (11)	0.32 (11)
	3	4.3 (9)	0.40 (9)	2.26 (9)	0.56 (9)
	4	5.26 (10)	0.28 (10)	3.23 (10)	1.35 (10)
	5	7.11 (7)	0.37 (7)	4.16 (7)	2.17 (7)
	6	7.32 (1)	0.56 (1)	3.32 (1)	3.43 (1)
	8	15.59 (1)	0.43 (1)	9.43 (1)	5.34 (1)
PAPI	1	5.0 (1)	-	-	-
	2	4.8 (8)	-	-	-
	3	4.40 (9)	-	-	-
	4	6.0 (7)	-	-	-
	5	6.0 (4)	-	-	-
	6	8,3 (2)	-	-	-

Source: Istat, Experimental survey on household grid - 2018

In addition, results show that the largest households (2-8 members) were assigned to CAWI[4], while those interviewed through CATI or PAPI had a maximum size of 5 and 6 respectively (Table 4).

On account of this, in order to control for a possible household size effect on the interview length, we compared households with the same size. For instance, focussing on 4 people families, it can be observed that CAWI reported the lowest total length (5'26''), while CATI and PAPI lasted on average 6'. Compilation times for the section on members and their relationships are quite similar both for CAWI and CATI (slightly lower for the former).

3.2 The identification of 'nuclei' in households

As already mentioned, when families have more than one nucleus, it is very important to identify all the bilateral relationships between family members (couple and parent-children relationships). The analysis of data showed that the HHG allows to solve possible ambiguities, which would remain with the old approach, due to the higher complexity of asking kinships in relation to one Reference Person or to the lack of some information. For example: if in a household there are more than two adults and children with similar age, the edit/imputation phase bears the risk of assigning the parent-children relation to the wrong people. This would not cause important problems from a statistical point of view, since households with more than a nucleus are not frequent in Italy, but it would give a picture not corresponding to the real situation. The test demonstrated that the availability of information on the relationship of each member with respect to all others guarantees the correct identification of all family nuclei.

3.3 The cognitive test results

Cognitive interviewers were managed through a structured questionnaire containing the questions described in paragraph 2.2. The analysis of responses[5] showed an overall positive judgement in terms of easiness/directness in giving the answers for all data collection techniques.

4 It is worth mentioning that the long time needed for the 8-member household to complete the CAWI questionnaire (16 minutes) was also due to the software (Limesurvey), which doesn't perform efficiently when complex rules have to be checked for many members.

5 Respondents were asked to assign a score from 1 to 10 to indicate the easiness/directness in giving the answers.

Table 5 - Judgement on the easiness/directness in giving the answers (absolute numbers in brackets)

Score	Data collection technique			Total
	PAPI	CATI	CAWI	
6	3.2 (1)	0.0	2.6 (1)	1.8 (2)
7	0.0	0.0	7.7 (3)	2.7 (3)
8	3.2 (1)	0.0	10.3 (4)	4.5 (5)
9	12.9 (4)	4.9 (2)	10.3 (4)	9.0 (10)
10	80.7 (25)	95.1 (39)	69.1 (27)	82.0 (91)
Total	**100 (31)**	**100 (41)**	**100 (39)**	**100 (111)**

Source: Istat, Experimental survey on household grid - 2018

Table 5 shows that 82% of respondents assigned the maximum score, while the lowest score assigned (6) was provided by only 2 cases. Considering these 2 respondents, one of them simply pointed out that the interview was boring, while the other expressed difficulty in understanding the logical link between ‘Marital status’ and intra-household relationship. In fact, he had a partner but was not married, so he expected to find a corresponding item in ‘Marital status’, while in Italy de facto families are not recognised by law. This problem emerged also for other respondents, but it does not regard the household grid approach.

Among the 3 respondents who assigned the score 7, only one provided comments, complaining that the software system responded slowly. His family was the largest one, so this was caused by Limesurvey’s already mentioned limits.

Concerning the easiness/difficulty in identifying the pertaining intra-household relationship (see Table 6), 90% of respondents said that the task was ‘Very easy for all the relationships’. Some difficulties were reported for:

- adopted/temporarily assigned children;
- ‘foster parent’: the respondent said it would be easier to reverse the perspective, declaring that the boy was the child of his partner;
- the already mentioned logical link between ‘Marital status’ and intra-household relationship.

Respondents who declared ‘Quite easy for all relationships’, suggested to specify that item ‘09 - brother/sister’ also includes siblings with only one parent in common.

In sum, problems encountered by respondents do not depend on the household grid approach, but on comprehension and completeness of definitions.

Table 6 - Easiness/Difficulty in identifying the pertaining intra-household relationship

In general, identifying the intra-household relationship was:	Data collection technique			Total
	PAPI	CATI	CAWI	
1 – Very easy for all relationships	93.5 (29)	92.7 (38)	84.6 (33)	90.1 (100)
2 – Quite easy for all relationships	0	4.9 (2)	10.3 (4)	5.4 (6)
3 – Easy for some relationships, difficult for others	6.5 (2)	2.4 (1)	5.1 (2)	4.5 (5)
4 – Quite difficult for all relationships	0	0	0	0
5 – Very difficult for all relationships	0	0	0	0
Total	**100.0 (31)**	**100.0 (41)**	**100.0 (39)**	**100.0 (111)**

Source: Istat, Experimental survey on household grid - 2018

The need to read again or to ask the interviewer to repeat some concepts emerged in only 11 cases for the following aspects:

- 'brother/sister in law' definition;
- relationships among siblings;
- the inability to find a suitable response item: in a few cases the pertaining relationship category was not selected because it was not displayed as a result of a wrong answer given in the GHI section (see Paragraph 2.1.1).

In perspective, le latter problem could be solved by showing on the upper section of each screenshot a synthetic family scheme, with name, gender and marital status of each household member. In this way, respondents can easily see whether they gave some wrong answers in the GHI section that may have caused the errors.

Finally, cognitive questions regarding the use of the options 'Other relatives' and 'Other non-relatives' showed that they were always selected correctly, except from one case when 'grandchild' was associated to the item '15 - Other parents' instead of the correct one.

3.4 The role of the interviewers

The interviewers played a key role in the success of this survey. Due to its characteristics, this experience can be considered a very particular methodological experimentation that touches on many topics of the theory and practice of social research (Marradi, 2007; Corbetta, 1999; Campelli, 1996; Statera, 1995; Agnoli, 1994). Regardless of the great interest of the survey's object itself, a further reason of methodological interest lies in the specificity of the context of the interviews (both the interviewers and the interviewees selected in the same professional context) and also in the unexpected reaction of many interviewees, which can be configured as a sort of serendipity (Merton & Barber, 1992). Indeed, it is a rare event that in an experimental survey both the interviewers and the interviewees are selected in the same professional context.

To better understand the tasks performed by the interviewers, we can analyse the whole process as organised in four steps. The interviewers had to:

a) follow an in-depth training on the questionnaire, including simulated interviews on possible criticalities related to the interview and specifically to the HHG;

b) contact the people to be interviewed;

c) carry out the interviews;

d) give their personal feedback on the experience.

During the in-depth training, the interviewers attended some meetings, in which the many activities to be carried out were described. The training was focussed on the possible criticalities related both to the questionnaire and to the recoding of the family relationships among the household members. In this regard, it is worth emphasising that the three different techniques that had to be tested (PAPI, CATI, and CAWI) involved a different kind of relationship with the respondents, and consequently different kind of work for the interviewers. In PAPI and CATI interviews, the interviewers had to submit the questions to the interviewees. They had to fill in the questionnaire themselves, codifying the kinship relations, as declared by the respondents. In the CAWI interview, they had to observe (in a non-participant mode) the respondent's behaviour during the compilation of the questionnaire. For all the three techniques, the interviewers had to provide support – if requested – and ask the questions of the cognitive test. After the interview, they implemented the data file, adding their own observations.

In order to be skilled to cope with possible difficulties, very demanding interviews were simulated during the training phase. They were conducted on a set of hypothetical families, characterised by particularly complex bonds. These simulations were carried out with all the data collection techniques to be used in the real test. Following the training, the interviewers contributed to the improvement of some aspects related to the wording of some questions (Pitrone, 2005), which regarded, for example, the need to insert some further information in the PAPI questionnaire, useful to make it easier to carry out the interviews correctly.

As already said, interviewers and interviewees had been chosen in the same work context. This could create, from a methodological point of view, some doubts whether colleagues would agree to be interviewed more easily than in real surveys. Surprisingly, it was not as easy as expected. In fact, some colleagues expressed the same doubts as expressed by individuals during official surveys, *e.g.* concerns their privacy (Facioni, 2017). Furthermore, this research experience confirms the power of the *situational context*, from the point of view of the phenomenological theories (Garfinkel, 1967). For example, the whole group of interviewers had the feeling that the colleagues, during the interview, had changed their attitude towards them. They were not colleagues, but only individuals in an interview context. They returned to being colleagues only at the end of the cognitive test. This real serendipity represents a further element of interest in this research, which was carried out for a totally different purpose.

The interviewers debriefing highlighted that:

- the flow of the interviews was very simple to manage. Only for paper questionnaires it was suggested to make it clearer that HHG questions had to be asked row by row (and not column by column);
- respondents often requested clarifications on some general concepts not related to the management of the HHG (definitions of 'household', of usually 'resident household member', *etc.*)

Finally, interviewers observed the habit to consider 'relatives' also members that are not proper 'relatives' (for instance 'child of cousin'), which is probably due to a cultural sentiment. Anyway, from the statistical point of view, such an error of classification would not have an impact on the identification of the 'nuclei'.

4. Conclusions and perspectives

The main doubts about the household grid approach regarded the time to collect the information and the perceived respondent burden caused by the repetition of questions.

The test results showed that the interview length was contained for all the techniques (about 5 minutes, with the majority of time spent for GHI Section); responding was also considered easy and not burdening.

Concerning quality, the system of checking rules implemented for CATI and CAWI electronic questionnaire did not cause problems and guaranteed the correctness of data. Paper forms also were error-free. The only suggestion with regard to CATI and CAWI modes was to show a synthetic family scheme on the upper section of each screenshot, so that respondents can easily see whether they gave some wrong answers in the previous questions.

In addition, the test results confirmed the reduction of ambiguity in identifying family nuclei, thanks to the higher precision in the identification of intra-household relationships given by the household grid.

However, some aspects need to be improved; in particular, all the information addressed to respondents to explain some concepts (like the definitions of 'household' or 'usually resident household member', *etc.*) and the specifications to give correct responses (which members should be included or excluded, *etc.*) should be managed more efficiently.

The good results of the test demonstrate that this approach could be adopted for all the socio-demographic surveys, this way guaranteeing the homogeneity of survey questionnaires and comparability of data as well as a better accuracy in identifying household structures. This solution would be particularly important in the actual context of the Master Sample.

References

Agnoli, M.S. 1994. *Concetti e pratica nella ricerca sociale*. Milano, Italy: Franco Angeli.

Beatty, P.C., and G.B. Willis. 2007. "Research synthesis: the practice of cognitive interviewing". *Public Opinion Quarterly*, Volume 71, Issue 2: 287-311.

Bradburn, N.M. 1978. "Respondent Burden". Paper presented at the 138th *Annual Meetings of the American Statistical Association*, San Diego, CA, U.S.

Campelli, E. 1996. *Il metodo e il suo contrario*. Milano, Italy: Franco Angeli.

Conrad, F., J. Blair, and E. Tracy. 2000. "Verbal reports are data. A Theoretical Approach to Cognitive Interviews". In *Proceedings of the 1999 Federal Committee on Statistical Methodology Research Conference*. Washington, D.C., U.S.: Office of Management and Budget.

Corbetta, P. 1999. *Metodologia e tecniche della ricerca sociale*. Bologna, Italy: Il Mulino.

European Commission. 2009. "Social Variables Standardization". Item 5.3, *Meeting of the European Directors of Social Statistics*. Luxembourg, 23rd - 24th February 2016.

Facioni, C. 2017. "Le opinioni, via web, dei cittadini intervistati in Istat". In Romano, M.C. (a cura di). "L'utilizzo della tecnica CAWI nelle indagini su individui e famiglie": 153-159. *Letture Statistiche – Metodi*. Roma, Italy: Istat. https://www.istat.it/it/archivio/203729.

Frankel, J., and L.M. Sharp. 1981. "Measurement of Response Burden". *Statistical Reporter*, N. 81-4: 105-111.

Garfinkel, H. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ, U.S.: Prentice-Hall.

Marradi, A. 2007. *Metodologia delle scienze sociali*. Bologna, Italy: Il Mulino.

Merton, R.K., and E.G. Barber. 1992. *The Travels and Adventures of Serendipity: A Study in Historical Semantics and the Sociology of Science*. Princeton, NJ, U.S.: Princeton University Press.

Oksenberg, L., C.F. Cannell, and G. Kalton. 1991. "New Strategies for Pretesting Survey Questions". *Journal of Official Statistics – JOS*, Volume 7, N. 3: 349-365.

Pitrone, M.C. 2005. *Il sondaggio*. Milano, Italy: Franco Angeli.

Presser, S., J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer. 2004. *Methods for testing and evaluating survey questionnaire*. Hoboken, NJ, U.S.: John Wiley & Sons, *Wiley Series in Survey Metodology*.

Sharp, L.M., and J. Frankel. 1983. "Respondent Burden: A Test of Some Common Assumptions". *Public Opinion Quarterly*, Volume 47, Issue 1: 36-53.

Statera, G. 1995. *Logica dell'indagine scientifico-sociale*. Milano, Italy: Franco Angeli.

Sudman, S., N.M. Bradburn, and N. Schwartz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco, CA, U.S.: Jossey - Bass Publishers; Hoboken, NJ, U.S.: John Wiley & Sons, *Wiley Series in Research Methods for the Social Sciences*.

Willis, G.B. 2005. *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Thousand Oaks, CA, U.S.: SAGE Publications.

Willis, G.B. 1994. "Cognitive Interviewing and Questionnaire Design: A Training Manual". *Working Papers Series*, N. 7. Atlanta, GA, U.S.: U.S. Department of Health and Human Services, Centers of Disease, Control and Prevention – CDC, National Center for Health Statistics.

Implementing ICD-11 for mortality statistics: translation of decision tables embedded in the automated coding system Iris

Chiara Orsi [1], Daniele De Rocchi [2], Mihai Horia Popescu [2], Friedrich Heuser [3], Stefanie Weber [3], Luisa Frova [1], Vincenzo Della Mea [2], Francesco Grippo [1]

Abstract

Iris is a system for causes-of-death coding, based on the International Classification of Diseases, 10th revision (ICD-10). As coding rules, Iris uses a knowledge base consisting in a large number of relations between ICD-10 codes. In 2019, WHO approved a new revision of the ICD, the ICD-11, which greatly differs from the previous. In order to let Iris code with this new version, each single rule contained in the knowledge base need to be translated into ICD-11.
This article describes the findings of a pilot project carried out in order to evaluate the feasibility of this translation. The project highlighted that part of the rules can be automatically translated using the mapping tools provided by WHO between ICD-10 and ICD-11.
However, most of the rules need to be manually revised. The pilot project identified the rules that need to be prioritised in the translation process since they are very frequently used in real data coding.

Keywords: International Classification of Diseases - ICD, automated coding, Iris, decision tables, causes of death, mortality statistics, underlying cause.

1 Italian National Institute of Statistics - Istat, Roma, Italy.
2 University of Udine, Italy.
3 Federal Institute of Drugs and Medical Devices - BfArM, Cologne, Germany.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Background[4]

Mortality statistics are based on causes of death reported by physicians on death certificates. These data are coded according to the provisions of the International Classification of Diseases (ICD) of the World Health Organization (WHO), which allows to make these statistics comparable at an international level. The ICD is periodically revised to take into account medical, epidemiological, technical development. Currently, the 10th revision (ICD-10) is in use worldwide (WHO, 2019*a*), but in May 2019 the 72nd World Health Assembly approved the 11th revision (ICD-11) which will become the new standard (WHO, 2019*b*).

In order to guarantee standardisation in data collection, ICD recommends a standard format of the death certificate, which comprises two parts. In Part 1 the sequence of events leading to death should be reported. This part includes different lines: in the first one the immediate (terminal) cause should be reported and in the lowest the condition that started the sequence leading to death. In the other filled lines, intermediate causes should be reported. Conditions reported on a line of Part 1 should be caused by conditions reported in the line below. Part 2 includes other relevant conditions contributing to death but not part of the sequences reported in Part 1.

For cause-of-death data production, an ICD code is attributed to each medical condition. The whole set of ICD codes of each certificate is referred to as "multiple cause". From this set, one single code is selected and used for international comparisons, the so-called underlying cause of death (UC), defined by WHO as the "disease or condition that initiated the train of events leading directly to death or the circumstances of the accident or violence

4 This work is carried out within the project for the evaluation of efforts needed for ICD-11 implementation into Iris, promoted by the Iris Institute (www.iris-institute.org), and financed by the Australian Bureau of Statistics – ABS.
Authors thank all the countries that cooperated by sharing anonymous data on multiple cause of death: Italian National Institute of Statistics - Istat; Statistics South Africa – Stats SA; Instituto Nacional de Estadística – INE, Spain; Ministry of Health/General Direction of Health Information/Mexican; Collaborating Centres for the WHO Family of International Classifications, WHO-FIC; CC (CEMECE); Hungarian Central Statistical Office – KSH; Office for National Statistics – ONS, UK; Centers for Disease Control and Prevention – CDC, U.S.A.
All authors contributed to the writing and revision of the text. In particular: Daniele De Rocchi, Francesco Grippo, Chiara Orsi and Luisa Frova contributed to the development of methodology and analysis of data concerning the measure of frequent due to in real data and provided analysis for ICD-10-ICD-11 translation.
Mihai Horia Popescu and Vincenzo Della Mea contributed to the translation from ICD-10 to ICD-11 and to the development of IT tools. Francesco Grippo and Vincenzo Della Mea are co-last authors.

causing the fatal injury" (WHO, 2019*a*). The ICD provides a system of rules that allows selecting the UC taking into account all codes in the certificate.

The multiple cause coding and the selection of the UC are generally performed with the use of automated coding systems, in order to increase comparability and quality of data. Automated coding systems greatly increase the comparability of death statistics, since they reduce the variability due to the different interpretation of rules by manual coders. However these systems cannot code automatically all certificates, so a certain rate of manual intervention is unavoidable.

The main automated coding system used at an international level is Iris (www.iris-institute.org), which is based on the international death certificate and coding rules provided by ICD-10. The core component of Iris are the "decision tables" (Iris Institute, 2019). Decision tables are a knowledge base for the correct application of the UC selection process and consist in a list of all plausible relations between all ICD codes. Once a code is attributed to each condition, the selection process is entirely based on ICD codes.

In summary, the process of UC selection can be seen as an algorithm in whose nodes it is necessary to assess relations between codes. Decision tables list all possible relations between codes and are used as a reference for this evaluation. Relations in decision tables are called rules. There are different types of rules, the most used is DUETO, which corresponds to causal relation, *i.e.* it indicates which codes can be considered a plausible consequence of another given code.

The decision tables were first developed by the NCHS (US National Center for Health Statistics) for the ACME automated coding system (Israel, 1990; CDC, 2015). Successively they have been embedded in Iris and updated on the basis of the recommendations of the Mortality Reference Group, which operates in the network of the World Health Organization Collaborating Centres for the Family of International Classifications - WHO-FIC (Navarra *et al.*, 2020, and 2016).

1.1 Objectives

With its adoption by the 72nd World Health Assembly in May 2019, the 11th revision of ICD (ICD-11) (WHO, 2019*b*) will become the new standard for mortality coding. In order to produce mortality statistics with ICD-11, a transition of Iris to this new classification is needed. A pilot project has been carried out to assess feasibility of this transition, in particular the translation of decision tables from ICD-10 to ICD-11 (Della Mea *et al.*, 2019).

Focussing on the rule type DUETO, the objective of this paper is to describe methods identified in the pilot project for the translation of rules contained in decision tables and to quantify the number of them that can be automatically translated. Provided that there are rules that need to be manually translated, the pilot project also aimed at identifying the rules mostly used for real data coding in order to prioritise these during the translation to ICD11. Additionally, it was investigated if there are relations frequently reported by physicians on certificates, but not included in decision tables.

In a future project, the methods identified in the pilot project and described in the paper would be applied to decision tables, with the aim of providing the effective decision tables translation in ICD-11.

2. Materials and methods

For this analysis, the 2019 decision tables were considered, since they include the most recent ICD-10 updates.

2.1 DUETO rule and other rules

When a certifier reports two conditions on different lines of Part 1, he implicitly indicates that there is a causal relation between them, *i.e.* the condition reported in the line above is "due to" the condition in the line below. In general, given a condition A, all conditions reported in lines above can be considered due to A. On the other hand, A can be considered due to conditions reported below. Nevertheless, sometimes certifiers report conditions in a causal order by mistake, or because they report several conditions and, by chance, they happen to be positioned in different lines.

DUETO rules in decision tables can be defined as follows: code A is DUETO code B if B is an acceptable cause of A according to ICD provisions; A is called *codeDef* and B *subcodeDef*. This rule type is used in several steps of the UC selection algorithm, when it is necessary to identify the causal sequence leading to death reported in Part 1.

There is a causal relation between two codes A and B (B is cause of A) if A is reported on a higher line of Part 1 of the certificate respect to B and A can be "due to" B, *i.e.* the decision tables contain the rule "B DUETO A". Decision tables, hence, contain a list of all relations between ICD codes of the type "B DUETO A" considered acceptable by the ICD provisions. For example, the rule "C79.9 DUETO C34.9" (C79.9 is the ICD-10 code for metastasis, C34.9 is the code for lung cancer) is included in the decision tables, as the causal relation "metastasis due to lung cancer" is plausible. On the other hand, the rule "C34.9 DUETO J18.9" (J18.9 is the code for pneumonia) is not included in the decision tables, as the causal relation "lung cancer due to pneumonia" is not plausible, as a cancer cannot be caused from other diseases. Therefore, even if a physician reports the code C34.9 on a higher line of the certificate respect to J18.9, the causal relation between these two codes is not taken into account from the UC selection algorithm. The idea behind decision tables is that there are some causal relations that cannot be considered acceptable (for example cancer due to other diseases), all the others are considered plausible and therefore accepted if reported by physicians.

This paper focusses on DUETO rules, which are the most represented in decision tables (more than 20,000,000) however other rule types are included in the decision tables. They are listed in Table 1, which shows also their number. Besides *codeDef*, *subcodeDef*, and rule type, decision tables contain other variables that increase the complexity of the system, but this additional information is not described in the present paper.

It is important to remark that DUETO rules apply only to Part 1 of the certificates, as the causal sequence leading to death should be reported in this part. Other rule types take into account also relations with codes reported in Part 2, but they are not treated in this paper.

Table 1 - List of the rule types included in decision tables, with the steps of the UC selection algorithm in which are used and the number of rules

Rule	Description	Step of UC selection	Number of rules (a)
DUETO	Due to	SP3-SP5	20,433,525
DS	Direct Sequel	SP6	2,026,631
DSC	Direct Sequel with Combination		17,257
IDDC	Ill-defined in Due to with Combination	SP7	2,250
IDMC	Ill-defined with mention with combination		127
LDC	Linkage in Due to with Combination	M1	50,682
LDP	Linkage in Due to with Preference		6,194
LMC	Linkage with Mention with Combination		31,608
LMP	Linkage with Mention with Preference		36,697
SDC	Specificity in Due to with Combination	M2	5,504
SMC	Specificity with Mention with Combination		1,513
SMP	Specificity with Mention with Preference		46,830

Source: Our processing
(a) 2019 edition.

2.2 Translation method

Along with ICD-11, the WHO releases the mapping table between ICD-10 and ICD-11, *i.e.* a table that reports, for each ICD-10 code, the corresponding ICD-11 code/codes. Mappings can be classified based on cardinality as follows (the cardinality is indicated with axb, where "a" is the cardinality of ICD-10 codes and "b" is the cardinality of ICD-11 codes):

- 1x1 (equivalent, ≡), the ICD-10 code is translated in one ICD-11 code;

- 1xn (⊒), the ICD-10 code is translated in more ICD-11 codes;
- nx1 (⊑), more ICD-10 codes are translated in the same ICD-11 code;
- nxn, the structure of the classification changes.

From former evaluations of the transition between previous revisions of ICD (in particular from ICD-9 to ICD-10) (Anderson *et al.*, 2001), it is known that the transition to a new revision will have an impact and the mapping will not be enough to completely automatise the transition. The mapping table allows translating single codes. However, when rules are considered, we cannot translate codes separately, but we need to interpret the relation from a logical point of view and consider the impact that the different mapping cardinality of the two codes involved has on the translation process.

In order to identify DUETO rules that can be automatically translated, the first step is to verify if a mapping exists for both *codeDef* and *subcodeDef*. Mapping could be missing since the knowledge on diseases evolves: some ICD-10 codes may have no mapping to the ICD-11 since the concept is no longer used, and some ICD-11 codes may have no mapping from the ICD-10 since they are new.

The second step is, for DUETO rules with existing mapping for both *codeDef* and *subcodeDef*, to assess if the rule can be automatically translated. A rule can be automatically translated if both *codeDef* and *subcodeDef* can be automatically translated; this is established on a logical point of view, taking into account the mapping cardinality of codes.

Based on the analysis results, a prototype of a web tool for translation was developed to help and guide the experts in the translation process.

2.3 Analysis of due to in real data

As seen, in decision tables there are several millions of rules involving all codes of ICD. Nevertheless, not all of them are used during coding, since many refer to rare causes of death not (or very rarely) reported on death certificates. Therefore, we developed a method for understanding which rules are the most relevant and more frequently applied in data coding.

For this purpose, we analysed multiple cause data referring to years 2016-2018 from different countries (table 4). All countries collected data using the international certificate and Iris as coding tool. This resulted in the same data format for all countries, containing all codes representing the medical conditions reported on the death certificate, and the position of each code on the certificate. Overall, 4,812,100 certificates were analysed. For the analysis, only Part 1 of certificates was taken into account, as DUETO rules refer only to this part. Nearly all death certificates collected (4,811,844 out of 4,812,100) contained at least one code in Part 1.

First, for each pair of codes A and B reported on the certificates, we calculated the frequency of certificates reporting A as “due to” B, *i.e.* A on a higher line respect to B.

Successively, provided that the death certificates could contain two given conditions reported in both directions (A “due to” B as well as B “due to” A) we developed a method for the identification of recurrent causal patterns in multiple cause-of-death data.

We applied the following two steps.

1. Analysis of association between codes. Two different codes (A and B) may appear jointly on the same certificate by chance depending on their frequency in the total sample of certificates. We tested the null hypothesis by a X^2 test, comparing the observed frequency of certificates showing the joint presence of codes A and B with the expected one. This step allowed identifying codes positively associated, *i.e.* reported on certificates more than expected.
2. Analysis of DUETO relations. For codes positively associated, we wanted to understand if there is a preferred direction of the DUETO relation. For this step, we used only cases in which the two codes are reported in different lines (sometimes codes can be placed on the same line). If codes were randomly reported on certificates’ lines, we would expect half of cases in which code A is DUETO B and half the contrary (null hypothesis – expected frequency). With a X^2 test we compared this expected frequency with the observed one and we identified DUETO relations reported more than expected.

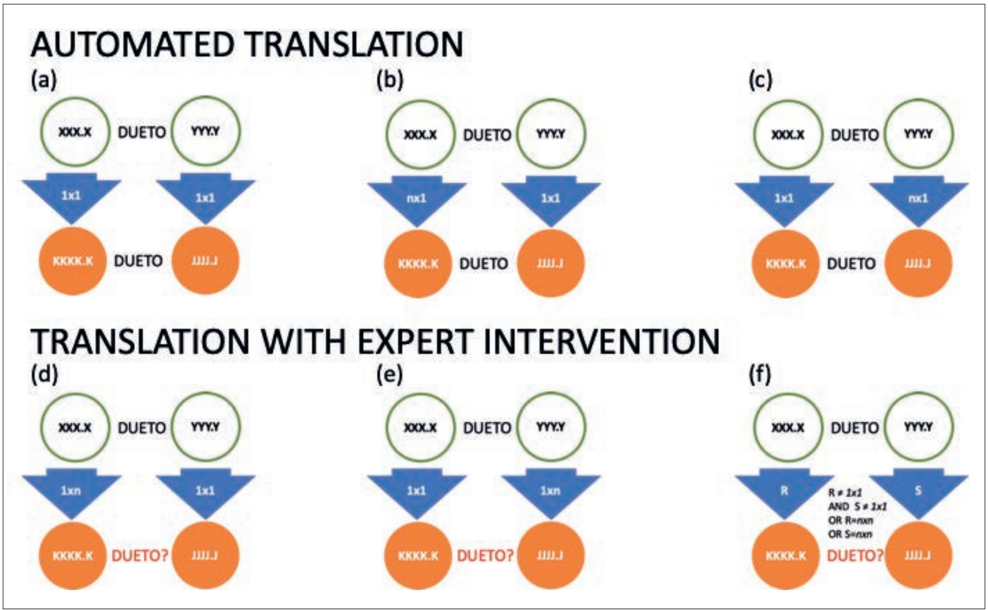
For DUETO relations reported more than expect, we checked if the corresponding DUETO rule is included in the decision tables.

3. Results

3.1 Translation

Rules in which both codeDef and subcodeDef have equivalent mapping (1x1) can be always automatically translated. On the other side, if codeDef and/or subcodeDef have a mapping with cardinality nxn, translation will always be manual. The other cases provide some chance of automated translation. Figure 1 shows all possible cases. In the high part of the Figure, cases that can always be automatically translated (a) and those that only in some cases require a manual revision (b and c) are shown. For example, rules included in case (b) can be automatically translated only if all the n codeDef have the DUETO rule with the subCodedef in ICD-10. If at least one of the n codeDef does not have the DUETO rule with the subCodedef in ICD-10 the corresponding rule in ICD-11 need a manual revision. The same applies for case (c). The low part of the Figure shows cases that always require a manual revision (d, e, f).

Figure 1 - Summary of translation cases



Source: Our processing

Table 2 shows examples of cases (a) (b) and (c) of Figure 1. Table 3 shows examples of cases (d) and (e). In these cases, ICD-11 classifies with a finer detail the concept coded by ICD-10. An expert should confirm if it is possible to extend the relation, valid for a broader concept in ICD-10, to all the detailed concepts in ICD-11.

Table 2 - Examples of cases (a), (b), and (c) in Figure 1

Case	ICD-10 rule	Mappings involved	ICD-11 rule
(a)	A01.0 (Typhoid fever) DUETO C33 (Malignant neoplasm of trachea)	A01.0 ≡ 1A07 (Typhoid fever) C33 ≡ 2C24 (Malignant neoplasms of trachea)	The ICD-11 rule 1A07 DUETO 2C24 can be automatically included in ICD-11 decision tables
(b)	F03 (Unspecified dementia) DUETO R54 (Senility)	F03 ⊑ 6D8Z (Dementia, unknown or unspecified cause) R54 ≡ MG2A (Old age) *But also:* *F00 (Dementia in Alzheimer disease) ⊑ 6D8Z* *F01 (Vascular dementia) ⊑ 6D8Z*	The ICD-11 rule 6D8Z DUETO MG2A must be revised as the ICD-10 rule F01 DUETO R54 is included in ICD-10 decision tables, but the rule F00 DUETO R54 is not included in ICD-10 decision tables
(c)	J96.9 (Respiratory failure, unspecified) DUETO F03	J96.9 ≡ CB41.2 (Respiratory failure, unspecified as acute or chronic) F03 ⊑ 6D8Z *But also:* *F00 ⊑ 6D8Z* *F01 ⊑ 6D8Z*	The ICD-11 rule CB41.2 DUETO 6D8Z can be automatically included in ICD-11 decision tables as all the rules J96.9 DUETO F00, J96.9 DUETO F01, J96.9 DUETO F03 are included in ICD-10 decision tables

Source: Our processing

Table 3 - Examples of cases (d) and (e) in Figure 1

Case	ICD-10 rule	Mappings	ICD-11 rule
(d)	I46.9 (Cardiac arrest, unspecified DUETO R26.3 (Immobility)	I46.9 ⊒ MC82 (Cardiac arrest) I46.9 ⊒ MC82.0 (Ventricular tachycardia and fibrillation cardiac arrest) I46.9 ⊒ MC82.1 (Bradycardic cardiac arrest) I46.9 ⊒ MC82.2 (Asystolic cardiac arrest) I46.9 ⊒ MC82.3 (Cardiac arrest with pulseless electrical activity) R26.3 (Immobility) ≡ MB44.3 (Immobility)	An expert should check if all types of cardiac arrest can be due to immobility
(e)	I27.9 (Pulmonary heart disease, unspecified) DUETO B44.1 (Other pulmonary aspergillosis)	I27.9 ≡ BB0Z (Pulmonary heart disease or diseases of pulmonary circulation, unspecified) B44.1 ⊒ CA82.4 (Aspergillus-induced allergic or hypersensitivity conditions) B44.1 ⊒ 1F20.12 Chronic pulmonary aspergillosis)	An expert should check if Pulmonary heart disease can be due to both CA82.4 and 1F20.12 can cause.

Source: Our processing

Examples are not provided for mixed situations involving nxn cardinality that are very complex. Further details on the translation methodology can be found in the pilot project report (Iris Institute, 2019).

By exploring the distribution of mapping types in the coding rules, we established that the number of DUETO rules needing manual intervention varies between 3 and 6 million out of about 20 million.

For the prototype tool implementation, we choose a web-based model where experts can work collaboratively from different locations and that can assure the consistency of results. The tool identifies rules that require an expert intervention. In these cases, the tool provides the list of mappings for all the codes involved in the rule and, if possible, a proposal of the rule translation. Where possible, rules with the same *subcodeDef* are grouped to ease the translation and facilitate maintenance over time.

3.2 Due to in real data

Table 4 shows the results of some descriptive analyses of real certificates. The total number of different codes reported on certificates varies from 1,405 in United Kingdom to 5,553 in United States; considering all countries, 6,786 different codes were found on certificates. The total number of different codes reported in Part 1 (therefore considered for the present analysis) varies from 1,102 in United Kingdom to 5,008 in United States; considering all countries, 6,292 different codes were found in Part 1. The overall average number of codes per certificate is 3.2; it varies from 1.7 in South Africa to 4.6 in Hungary. The overall average number of codes in Part 1 per certificate is 2.4; it varies from 1.6 in United Kingdom to 3.4 in Italy and Hungary.

Table 4 - Death certificates analysed

Country	Number of cases	Number of different codes reported	Number of different codes reported in Part 1	Average number of codes per death certificate	Average number of codes in Part 1 per death certificate	Source
Italy (IT)	618,083	4,029	3,576	4.4	3.4	Italian National Institute of Statistics;
South Africa (ZA)	473,938	3,17	3,169	1.7	1.7	Downloaded from Statistics South Africa – Stats SA website;
Spain (ES)	424,523	3,577	3,326	3.7	3.1	Instituto Nacional de Estadística – INE;
Mexico (MX)	307,433	2,539	2,375	2.9	2.4	Ministry of Health/General Direction of Health Information/Mexican WHO-FIC CC (CEMECE);
Hungary (HU)	131,668	3,204	2,775	4.6	3.4	Hungarian Central Statistical Office – KSH;
United Kingdom (UK)	36,421	1,405	1,102	2.7	1.6	Office for National Statistics – ONS;
United States (US)	2,820,034	5,553	5,008	3.2	2.2	Downloaded from Centers for Disease Control and Prevention – CDC website
All countries	4,812,100	6,786	6,292	3.2	2.4	

Source: Our processing

Table 5 shows the results of descriptive analyses about DUETO relations. Almost 464,000 different DUETO relations are reported on death certificates analysed. Of them, more than 50,000 are observed more than expected. The agreement with decision tables is high: 78% of the relations reported more than expected are included in the tables.

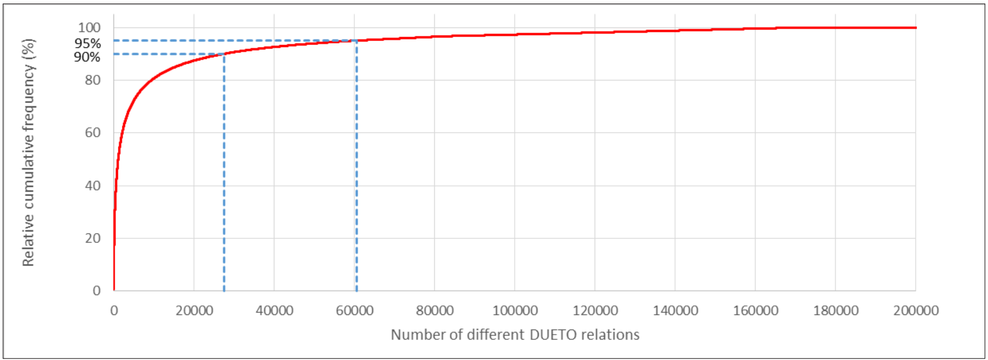
Table 5 - Results of due to analysis

	N	%
Number of different ordered pairs found in death certificates	463,939	
Of which		
Pairs reported in a given causal order (due to) more than expected (X^2 test, p<0,05)	51,059	100.0
Of which:		
In agreement with decision tables	39,819	78.0
In disagreement with decision tables	11,240	22.0

Source: Our processing

Figure 2 shows the cumulative frequency for DUETO relations ordered by decreasing frequency on certificates. On the x-axis the number of different DUETO relations reported on certificates, starting from the most frequent, is shown. On the y-axis the percentage of DUETO relations reported is shown. For example, from the Figure, we can see that 27,000 and 60,000 different DUETO relations represent respectively 90% and 95% of all DUETO relations reported on certificates. The curve allows estimating the percentage of completeness of translation we can reach starting the translation from the most frequent relation. The curve shows that if the first 27,000 most frequent DUETO are translated, about 90% of all DUETO reported on certificates is translated. If the first 60,000 most frequent DUETO are translated, it is possible to reach 95% of completeness.

Figure 2 - Cumulative frequency curve for DUETO relations ordered by decreasing frequency



Source: Pooled cause-of-death statistics Italian National Institute of Statistics - Istat; Statistics South Africa – Stats SA; Instituto Nacional de Estadística – INE, Spain; Ministry of Health/General Direction of Health Information/ Mexican; Collaborating Centres for the WHO Family of International Classifications, WHO-FIC; CC (CEMECE); Hungarian Central Statistical Office – KSH; Office for National Statistics – ONS, UK; Centers for Disease Control and Prevention – CDC, U.S.A.

Figure 3 shows DUETO relations reported on real certificates more than expected, distinguishing between those included in decision tables (left part, graph A) and those not included in decision tables (right part, graph B). On the x-axis the *codeDef* are reported and on the y-axis the *subcodeDef*. Even if it is not possible to interpret the meaning of each point in the Figure, this representation allows making general observations and considerations on specific groups of relations.

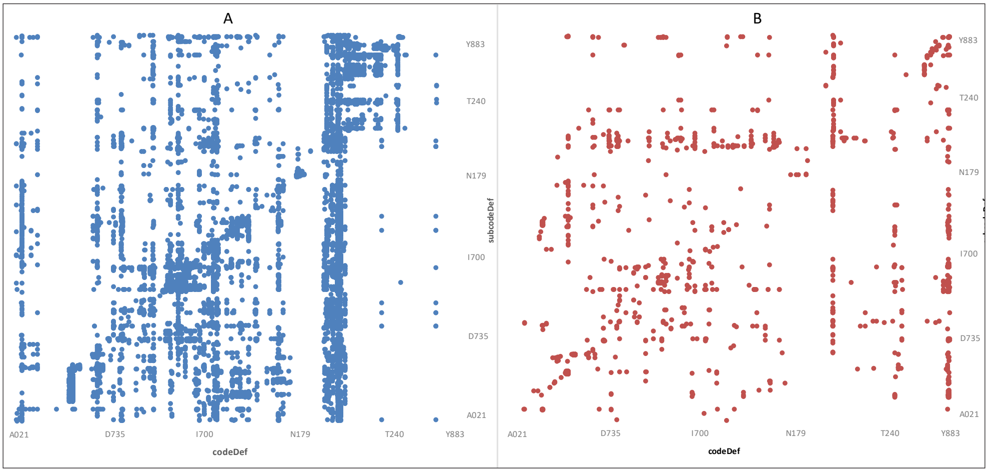
Among relations not included in the tables, it is possible to highlight different situations:

1. medically, the causal relation may exist but the classification explicitly provides not to accept the due to, for instance cancers due to some risk factors or viral diseases; such provisions of ICD are established for public health reasons and for making clear and comparable the counts of some specific diseases;
2. wrong reporting by certifiers such as:
    a. well defined diseases reported as due to symptoms or ill-defined conditions, for example stomach cancer reported due to gastritis
    b. chronological order preferred over causal order such as chronic obstructive pulmonary disease due to hypertension
    c. different clinical stages, for example neoplasm of unspecified behaviour causing malignant neoplasm;
4. diseases due to a very similar disease (diagonal in the graph).

We can also note that DUETO relations involving symptoms and signs are often not accepted by decision tables although they are frequently reported in death certificates. A frequent case is senility (R54) reported as due to many other conditions; this indicates that the mention of senility on the certificates should be seen as a synonym of "general frailty" and should be accepted as due to other conditions. Another frequent case is hemorrhage (R58) due to injuries and external causes.

Moreover, relations involving complications of medical and surgical care are reported, but these conditions are not included in tables as they are complicate cases that Iris cannot completely solve automatically.

Figure 3 - DUETO relations found in death certificates included in current decision tables (A) and not included (B)



Source: Pooled cause-of-death statistics Italian National Institute of Statistics - Istat; Statistics South Africa – Stats SA; Instituto Nacional de Estadística – INE, Spain; Ministry of Health/General Direction of Health Information/ Mexican; Collaborating Centres for the WHO Family of International Classifications, WHO-FIC; CC (CEMECE); Hungarian Central Statistical Office – KSH; Office for National Statistics – ONS, UK; Centers for Disease Control and Prevention – CDC, U.S.A.

4. Discussion and conclusion

Some results of the pilot project have been presented in a previous paper, mostly focussed on describing the logical knowledge needed to distinguish rules that can and cannot be automatically translated (Della Mea *et al.*, 2020). The results of the analyses showed in the present paper allow giving an idea of the efforts needed for the translation of decision tables in ICD-11.

The results show that the big majority of DUETO rules can be translated automatically, but the number of rules that require an expert intervention is anyway high. Nevertheless, the translation could in a first moment focus on the most frequent relations.

In the future project, a more precise estimation of the efforts needed should be carried out, using the methods identified in the pilot project and described in the present paper. Moreover, results from the two analyses can be combined in order to make decisions on how many and which rules should be translated. An evaluation of cost-effectiveness of different possible choices can also be made.

The results of the analyses on DUETO relations frequently reported by physicians but not included in decision tables could be useful to revise the tables and solve some issues, if necessary, during the translation process. Nevertheless, it should be taken into account that for some causes of death, such as external causes and injuries, tables might be not complete since Iris does not manage these cases. This incompleteness can be the origin of the differences between reported data and relations in the tables.

This transition is expected to need a big effort, also in terms of human resources, and to have a big impact on the system transition and use. Moreover, revising ICD and changing coding rules have a big impact on the comparability of cause-specific mortality statistics over the time. However, periodic revision of the ICD is essential to stay abreast of advances in medical science and changes in medical terminology (Boerma *et al.*, 2016; WHO, 2019*c*). Institutionally, revision of the ICD requires an enormous investment of national resources to revise software, training, publications, edit procedures, *etc.* (Anderson *et al.*, 2001). For the Iris transition to ICD-11, classification and coding experts are needed. To support their work, formal procedures are needed to ensure the correctness of the transition and validation of the system.

The main limitation of this research is that only DUETO rules are considered. For the complete translation of decision tables the same analyses need to be applied also to the other rule types, which take into account also Part 2 of certificates.

In conclusion, the presented methods seem suitable for supporting the process of transition of Iris from ICD-10 to ICD-11, however it further needs expert validation to correctly estimate the workload needed and to be applied to the other rule types.

References

Anderson, R.N., A.M. Miniño, D.L. Hoyert, and H.M. Rosenberg. 2001. "Comparability of Cause of Death Between ICD–9 and ICD–10: Preliminary Estimates". *National Vital Statistics Reports – NVSS*, Volume 49, Number 2: 1-32.

Boerma, T., J. Harrison, R. Jakob, C. Mathers, A. Schmider, and S. Weber. 2016. "Revising the ICD: explaining the WHO approach". *The Lancet*, Volume 388, Issue 10059: 2476-2477.

Centers for Disease Control and Prevention – CDC, National Center for Health Statistics – NCHS, National Vital Statistics System - NVSS. 2015. *About The Mortality Medical Data System*. https://www.cdc.gov/nchs/nvss/mmds/about_mmds.htm.

Della Mea, V., F. Grippo, M.H. Popescu, C. Orsi, and D. De Rocchi. 2019. *Evaluation project for integration of ICD-11 into Iris. Final report*. Cologne, Germany: Federal Institute for Drugs and Medical Devices - BfArM, Iris Institute.

Della Mea, V., M.H. Popescu, F. Grippo, C. Orsi, and F. Heuser. 2020. "Logical Rules and a Preliminary Prototype for Translating Mortality Coding Rules from ICD-10 to ICD-11". *Studies in Health Technology and Informatics*, Volume 270: 297-301.

Federal Institute for Drugs and Medical Devices - BfArM, Iris Institute. 2019. *Information about the coding rule types for mortality with Iris*. Cologne, Germany: BfArM.

Israel, R.A. 1990. "Automation of mortality data coding and processing in the United States of America". *World Health Statics Quarterly*, Volume 43, Issue 4: 259-262.

Navarra, S., M. Cappella, L.A. Johansson, L. Pelikan, F. Heuser, L. Frova, and F. Grippo. 2020. "Decision tables for mortality coding: methods and tools for the management and documentation of changes". *Rivista di statistica ufficiale*, N. 1/2018: 63-91. Roma, Italy: Istat. https://www.istat.it/it/archivio/254423.

Navarra, S., M. Cappella, L.A. Johansson, L. Pelikan, L. Frova, and F. Grippo. 2016. "Decision Table editor: a web application for the management of the international tables for mortality coding". *Istat working papers*, N. 6/2016. Roma, Italy: Istat. https://www.istat.it/it/archivio/184113.

World Health Organization - WHO. 2019*a*. *International Statistical Classification of Diseases and Related Health Problems. 10th Revision. 2019 Edition*. Geneva, Switzerland: WHO. https://icd.who.int/browse10/2019/en#/.

World Health Organization - WHO. 2019*b*. *ICD-11. International Classification of Diseases 11th Revision. The global standard for diagnostic health information*. Geneva, Switzerland: WHO. https://icd.who.int/en.

World Health Organization - WHO. 2019*c*. *ICD-11 Implementation or Transition Guide*. Geneva, Switzerland: WHO. https://icd.who.int/docs/ICD-11%20Implementation%20or%20Transition%20Guide_v105.pdf.

Optimal Design of a Master Sample: the Case of the Household Surveys in the Republic of Moldova

Giulio Barcaroli, Loredana Di Consiglio, Alessio Guandalini, Marco Dionisio Terribili [1]

Abstract

This article describes the results of the activities carried out in the framework of the "Project of Technical Assistance to Support the National Bureau of Statistics (NBS) of the Republic of Moldova", in particular those regarding the Activity A14 (Design of Master Sample): the identification of the household surveys that will make use of the new master sample and, for each of them, a newly optimised sample design; the calculation of the total amount of final sampling units to be selected from the master sample in its whole life span; the selection of Primary Selection Units (PSUs); the definition of two different possible configurations of the master sample, depending on the coordination or not of the PSUs samples of the different surveys; the evaluation of pros and cons of the two solutions, taking into account the constraints derived by the current organisation of the NBS data collection network. The above activities have been carried out making use of the methodologies implemented in three generalised software (ReGenesees, R2BEAT and FS4) developed and used by the Italian National Institute of Statistics - Istat.

Keywords: Statistical cooperation, Master Sample, two-stage sample design, optimal allocation, social surveys, sample coordination, R2BEAT, ReGenesees.

1 Giulio Barcaroli (gbarcaroli@gmail.com); Loredana Di Consiglio (diconsig@istat.it); Alessio Guandalini (alessio.guandalini@istat.it); Marco Dionisio Terribili (terribili@istat.it), Italian National Institute of Statistics.

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics - Istat.

The authors would like to thank the anonymous reviewers for their comments and suggestions, which enhanced the quality of this article.

1. Introduction[2]

Starting from December 2019 to June 2020 several cooperation missions were carried out in the framework of the Project of Technical Assistance to Support the National Bureau of Statistics of the Republic of Moldova, within the Activity A14 (Design of master sample)[3]. Activity A14 includes the following:

1. Identification of household surveys that make use of the master sample; then, determination for each survey, the initial allocation of units planned for each wave of the master sample;
2. Calculation of the total amount of units to be selected from the master sample in its whole life span;
3. Selection of Primary Selection Units (PSUs);
4. Refinement of sample designs for all surveys and selection of samples;
5. Improvement and documentation of the estimation procedures and sampling design;
6. Training on the ReGenesees and its use at the NBS, or other software tools;
7. Assistance to NBS in the production of a complete methodology.

The in-use version of the master sample was designed in 2018 and was intended to work during the years 2019 and 2020. Due to the exhaustion of included PSUs, the new version of the master sample had to be designed in the first half of 2020 and implemented in the second half, ready to be used at the beginning of 2021 for a period of four or five years.

The 2019-2020 master sample was implemented by selecting 150 PSUs in eight different strata, resulting from the cross-product of the four statistical regions (North, Center, South and Chisinau) and the PSUs rural/urban feature. There is an important difference between PSUs in the stratum Chisinau urban and in the other strata:

2 The authors wish to sincerely thank the entire NBS staff and in particular Lilian Galer, Olga Moraru and Galina Ostapenco for their precious collaboration during the activities of the cooperation project, without whom the results obtained would not have been reached.

3 EU funded Project ENI/2019/406-262, “Technical Assistance to Support the National Bureau of Statistics of the Republic of Moldova”.

- in Chisinau urban the PSUs correspond to the EAs in the Population and Housing Census 2014 (PHC 2014), for a total of 1,718 PSUs;
- in the rest of the country the PSUs are an aggregation of EAs from PHC 2014 in order to reach a minimum size of 420 households, for a total of 1,742 PSUs.

The first stage sample size was equal to 150 PSUs, with a simple random sampling in the Chisinau urban stratum, and a PPS (probability proportional to size) sample in the other strata. For each selected PSU a complete enumeration of households was carried out, and a simple random sample of them was drawn to guarantee the interviewers (150, one for each selected PSU) a predetermined number of interviews for the Labour Force Survey and the Household Budget Survey.

In the Chisinau urban stratum the sample PSUs for LFS and HBS do not match, whereas in the rest of the country the same sample PSUs are used both for LFS and for HBS, taking care that the sampled households do not overlap. As for the rotation schemes, to implement the 2-2-2 scheme of the LFS, in the Chisinau urban stratum only, each quarter, one-fourth of the sample -the entering LFS panel- is renewed and new PSUs are used for it; after 6 quarters - when the cycle is completed - each LFS PSU is removed.

Also in the Chisinau urban, PSUs used by HBS are removed once used for a round of the survey. In the other strata, there is no rotation of LFS PSUs. For both, LFS and HBS PSUs are removed once exhausted.

The multivariate optimal allocation carried out in 2018 was based on the following steps:

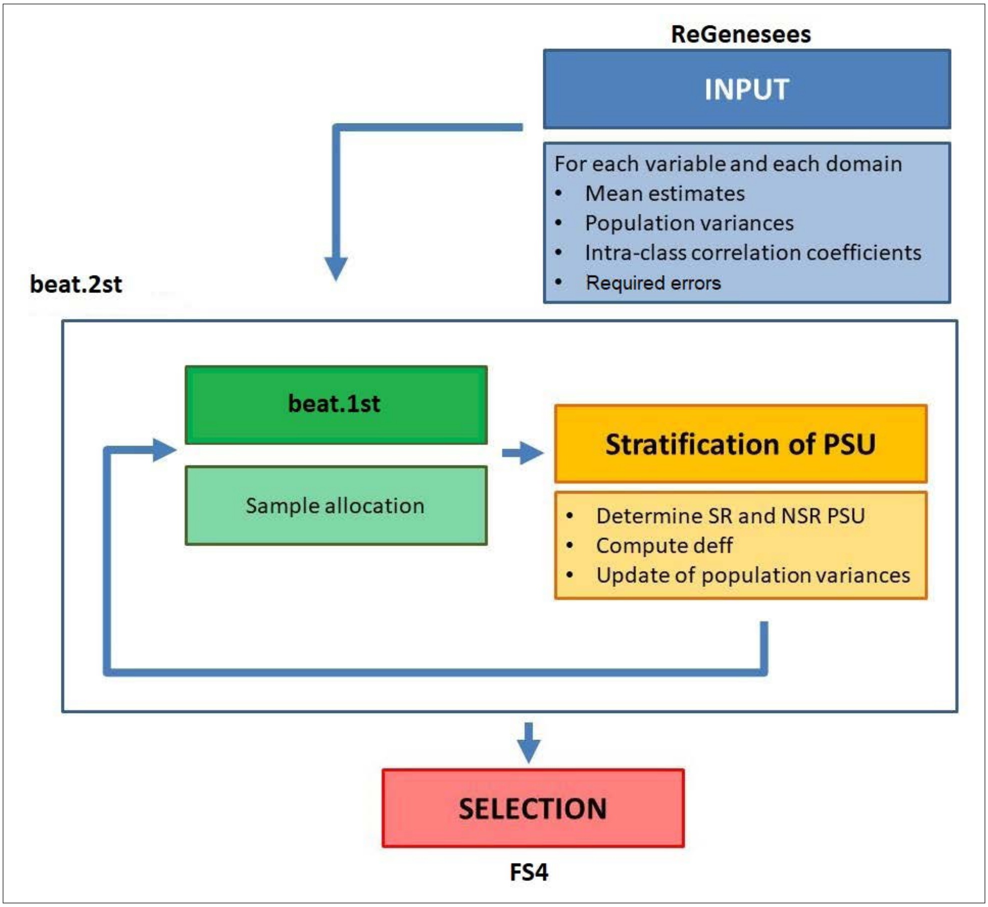
- the allocation of SSUs (households) in the 8 strata was based on the variability in these strata of the LFS and HSB target variables;
- taking the total amount of the PSUs as fixed (determined by the number of interviewers, *i.e.* 150), and setting a minimum number of 30 and 24 households per PSU respectively for LFS and HBS, the allocation of these 150 PSUs was determined by applying the methodology implemented in the R package R2BEAT that will be illustrated in the following sections.

According to the Inception Report of the Project, in December it was decided to perform the new master sample design taking into account not only the aforementioned surveys (LFS and HBS), but all those planned to be carried out during the life span of the MS. Consequently, beyond these two ones, two more were taken into consideration: the Energy Consumption and the Domestic Violence surveys. Differently from LFS and HBS, which are carried out quarterly, the two additional surveys are to be carried only once in a date to be still defined, but included in the interval 2021-2024.

So, to determine the total amount of Secondary Stage Units (SSUs) that are needed to conduct the four surveys and therefore that the new master sample is expected to satisfy over its lifetime, new sample designs for each survey have been produced. For each survey the following steps have been performed (see Figure 1):

1. previously available rounds of the surveys have been taken into consideration;
2. for each round, sampling estimates have been calculated together with the information needed by the next steps, *i.e.* strata variances, design effects and estimator effects: this has been done using the Istat software ReGenesees;
3. taking the output of the previous step as input, optimal allocation in terms of PSUs and SSUs has been obtained by using the R package R2BEAT, developed by Istat (Fasulo *et al.,* 2020);
4. the first stage selection of PSUs has been performed by using the R package FS4 (developed by Istat): this package stratifies further the PSUs in terms of their size and classifies them as Self-Representative (SR, *i.e.* with only one unit, with inclusion probability equal to 1), and Non-Self-Representing (NSR); in the latter strata, PSUs are selected with a PPS scheme.

Figure 1 - General framework of the work performed for each survey



Source: Our processing

The sample designs of the four surveys have been optimised concerning these important criteria:

1. the total amount of final units (households) have been minimised under the precision constraints (expected coefficients of variation not exceeding given values) specified by NBS, but tentatively modified in order to get feasible solutions;
2. in doing that, total non-response rates observed in previous survey rounds have been taken into account and the planned sample sizes have been inflated accordingly (oversampling);

3. First and second stage probability of inclusions are such that the overall inclusion probabilities are almost constant for each unit, resulting in self-weighted samples.

Each selected PSU may or may not be used by more than one survey: in the former case we say there is no coordination of PSUs (the only overlap is by chance), in the latter case we say there is coordination of PSUs.

In our specific case, coordination of samples has been obtained by considering LFS PSUs sample as pivotal, and coordinating with it the HBS sample (and then the Domestic Violence and Energy Consumption surveys), favouring the PSUs with a higher number of households. Anyway, positive coordination has never been applied for PSUs in the Chisinau urban stratum, due to the very low average number of households.

The master sample obtained without coordination is characterised by the highest number of PSUs, both in terms of selected ones and joint use in each quarter. In contrast, the master sample obtained by coordinating PSUs is characterised by the lowest number of initial PSUs, but a higher rate of exhaustion.

This paper is organised as follows: in section 2 an illustration of methodologies employed and of tools used is given; in Section 3 the sample design for each one of the four involved surveys is reported; in Section 4 the methodological approach to coordinate PSUs is described and the results obtained are illustrated; in Section 5 the two scenarios of implementation of the master sample (with and without coordination of PSUs) are illustrated by means of simulating the master sample use in the life span 2021-2025. Some conclusions are reported in the Section 6.

2. Methods and tools

This section briefly describes the main statistical methods that are needed for the determination of a sampling design and the software tools used for their application as illustrated in Figure 1.

2.1 Sampling estimates and sampling errors (ReGenesees)

The first step to study the new allocation for a sample survey is, when previous occasions of the survey are available, to estimate means and variances for its main target variables.

In order to compute sampling estimates and their precision, the R package ReGenesees[4] can be used. The package is the prneoduct of a long term Istat project.

The package enables computation of the calibration estimators and their variance estimation (see Deville & Särndal, 1992; Särndal, 2007) based on the linearisation method (see Woodroof, 1971). The package also contains utilities for calculating population totals, that are used for calibration estimation, or to aggregate strata, in case of unique sample units within strata. It also allows estimating precision measures of complex indicators that can be linearised.

The main functions are the following:

- e.svydesign describes the adopted sampling design;
- e.calibrate computes the calibration weights (Deville and Särndal, 1992);
- svystatTM computes weighted estimates for totals and means using suitable weights depending on the class of design;
- svystat is a general function to compute weighted estimates of totals, ratios, quantiles, *etc.*

See Zardetto (2015) for details on the use of the software. An illustration of its use is given with some detail in the following sections 3.1, 3.2, 3.4, 7 where the previous functions are specified accordingly to the contexts they are applied.

4 The software is not on R-CRAN, but can be downloaded from the Istat website: https://www.istat.it/en/ methods-and-tools/methods-and-it-tools/process/processing-tools/regenesees

2.2 Multivariate allocation in two-stage sampling design (R2BEAT)

Two-stage sampling design with stratification of PSUs are very common for household surveys in Official Statistics. However, their sample allocations are usually a non-easy task, because the household surveys have multi-domains and multi-purpose objectives, so they have to provide accurate estimates for different variables and different domains (*e.g.* geographical areas such as national, regional, *etc.*).

The whole sample size of a survey, n, is often an exogenous detail because it is usually defined by budget and, sometimes, also by logistic constraints.

Instead, the allocation of the sample among the strata ($h = 1, . . . , L$), generally at the lower level at which estimates are required, can be defined in different ways (mainly with uniform, proportional or optimal allocation). In Official Statistics, the optimal allocation is the most widely used method, especially because information on the size of the strata and the variance of target variables in the strata can be obtained. In particular, the variances of the target variables or at least of proxy variables for each stratum can be computed, for instance, from register data or previous occasions of the same survey.

Since August 2019, a package called R2BEAT, developed by Istat, is available on the R-CRAN repository (Fasulo *et al.*, 2020). R2BEAT easily manages all the complexity due to the optimal sample allocation in two-stage sampling design and provides several outputs for evaluating the allocation[5]. Its name stands for R "to" Bethel Extended Allocation for Two-stage. It is an extension of another open-source software called Mauss-R (Multivariate Allocation of Units in Sampling Surveys), implemented by Istat researchers[6]. Mauss-R determines the optimal sample allocation in multivariate and multi-domains estimation, for one-stage stratified samples. It extends the Neyman (1934) - Tschuprow (1923) allocation method to the case of several variables, adopting a generalisation of Bethel's proposal (1989).

5 To complete the suite of tools developed by Istat in order to cover the stratified sample design, we cite SamplingStrata (Barcaroli *et al.*, 2020), which allows to jointly optimise both the stratification of the sampling frame and the allocation, still in the multivariate multidomain case (only for one-stage designs), and MultiWay. Sample. Allocation, that allows determining the optimal sample allocation for multi-way stratified sampling designs and incomplete stratified sampling designs. See https://www.istat.it/en/methods-and-tools/methods-and-it-tools/design/design-tools/ multiwaysampleallocation for more details.

6 https://www.istat.it/en/methods-and-tools/methods-and-it-tools/design/design-tools/mauss-r.

The main idea for optimal allocation is that strata with larger size and larger variability with respect to the target variables need a larger sample size to provide better estimates. Since, in our case, data from previous surveys occasions were available, an optimal allocation could be deployed.

R2BEAT develops Bethel's proposal but, besides, enables the determination of optimal sample allocation for two-stage stratified samples. The R2BEAT version used in this paper includes three functions:

- beat.1st that computes multivariate optimal allocation for different domains in one-stage stratified sample design;
- beat.2st that computes multivariate optimal allocation for different domains in two-stage stratified sample design, considering the design effect;
- beat.cv that, given a multivariate optimal allocation, calculates the coefficient of variation.

In our specific case, because all the considered surveys will implement a two-stage sample design, only the function beat.2st was used. Before using this function, a preliminary step was needed. In fact, it has been necessary to define some parameters for the function (see the package documentation for the input formats)[7]. In particular:

1. population size of each stratum in terms of households or individuals depending on the aim of the survey (N);
2. strata to be censused (CENS);
3. cost of interviews per stratum (COST);
4. number of PSUs to be selected in each stratum (MINCS);
5. average dimension of SSUs in each stratum (DELTA, Δ);
6. minimum number of SSUs to be interviewed in each selected PSU (MINIMUM);
7. mean of each target variables in each strata (M, $\hat{p}$ or $\hat{\bar{x}}$);

7 The latest version available on R-CRAN includes two functions that, starting from survey data, provide the inputs as the other functions require.

8. estimated population standard deviation of each target variable in each stratum (S, $\hat{S}$);
9. intra-class correlation coefficients for each target variable in each stratum (RHO, $\hat{p}$);
10. estimator effect for each target variable in each stratum (EFFST, $\widehat{effst}$).

To better understand the optimal allocation, a few words on points 7-10 are necessary. In particular, the mean can be estimated using the function svystatTM of the package ReGenesees. For the estimated population standard deviation, a distinction between a binary and a quantitative variable is needed. In the case of binary variables, the estimated population standard deviation of variable *i* is equal to

$$\hat{S} = \sqrt{\hat{p}_i \times (1 - \hat{p}_i)} \tag{1}$$

where $\hat{p}_i$ is an estimate of the proportion of individuals with a given characteristic in the population strata, which is the mean of a binary variable. While, in the case of a quantitative variable, it is equal to

$$\hat{S} = \sqrt{\hat{M}_i^2 - \hat{\hat{Y}}_i^2} \tag{2}$$

where $\hat{M}_i^2 = \frac{\sum_{k \in s} y_{ik}^2 \; w_k}{N}$ and $\hat{\hat{Y}}_i = \frac{\sum_{k \in s} y_{ik} \; w_k}{N}$ are the quadratic mean and the arithmetic mean estimated on sample data of the previous occasion(s) of the survey, respectively, with y_{ik} the value of the target variable i observed on the unit *k*, w_k the corresponding sampling weight and *N* the population size.

A crucial statistic for the optimal sample allocation in a two-stage sampling design is the intra-class correlation coefficient, $\hat{p}$. Due to logistics, in household surveys, the sample is not wide-spread in the whole country, but often “clusterised” in PSUs and, then, in SSUs (clusters of family members). If the units of clusters are too similar to each other, it is not efficient to collect too many units from the same cluster. Anyway, sometimes a small loss in efficiency of the estimates is accepted because it is paid off by a gain in the organisation of the survey and, moreover, by cost reduction.

Therefore, to optimise the sample size, it is important to compute the intra-class correlation coefficient, because it provides a measure of data clustering in PSUs and SSUs. In general, if the value of $\hat{p}$ is close to 1 the clustering is high and it is convenient to collect only a few units in the cluster. On the contrary, if $\hat{p}$ is close to 0, the collection of units from the same cluster does not affect the efficiency of the estimates.

To compute $\hat{p}$, another important statistic must be introduced: the design effect (DEFF, *deff*). The design effect measures how much the sampling variance under the adopted sampling design is inflated with respect to a simple random sample (*srs*), with the same sample size. In formula:

$$deff(\hat{\hat{Y}}_i) = \frac{\text{var}(\hat{\hat{Y}}_i)_{des}}{\text{var}(\hat{\hat{Y}}_i)_{srs}} \tag{3}$$

$$= 1 + \hat{\rho}_i \ (b - 1)) \tag{4}$$

where b is the average cluster (*i.e.* PSU) size in terms of the final sampling units and $\hat{p}_i$ the intra-class correlation within the cluster (PSU) for the variable i.

The package R2BEAT takes into account a more general expression of *deff*. This expression refers to a typical situation for household surveys in which PSUs are assigned to Self-Representing (SR) strata, that is they are included for sure in the sample, or to Not-Self-Representing (NSR) strata, where they are selected by chance. In practice, this assignment is usually performed by comparing their measure of size (MOS) with respect to the threshold:

$$\lambda = \frac{\bar{m}\ \Delta}{f} \tag{5}$$

where $\bar{m}$ is the minimum number of SSUs to be interviewed in each selected PSU (MINIMUM), $f = n/N$ is the sampling fraction and Δ (DELTA) is the average dimension of the SSU in terms of elementary survey units. Then DELTA must be set equal to 1 if, for the survey, the selection units are the same as the elementary units (that is, household-household or individuals-individuals), whereas it must be set equal to the average dimension of the households if the elementary units are individuals, while the selection units are the households.

PSUs with MOS exceeding the threshold are identified as SR, while the remaining PSUs are NSR.

Then the extended expression of *deff* is

$$deff(\hat{\hat{Y}}_i) = \frac{N_{SR}^2}{n_{SR}}(1 + (\hat{\rho}_{i,SR}\,(b_{SR} - 1)) + \frac{N_{NSR}^2}{n_{NSR}}(1 + (\hat{\rho}_{i,NSR}\,(b_{NSR} - 1)) \quad (6)$$

where, for *SR* and *NSR* strata,

- N_{SR} and N_{NSR} are the population sizes;
- n_{SR} and n_{NSR} are the sample sizes;
- $\hat{p}_{i,SR}$ and $\hat{p}_{i,NSR}$ the intra-class correlation coefficient for the variable *i*;
- b_{SR} and b_{NSR} are the average PSU size in terms of the final sampling units.

Of course, if there are no *SR* strata the expression (4) recurs. The design effect is equal to 1 under the *srs* design and increases for each additional stage of selection, due to intra-class correlation coefficient which is, usually, positive. It can be computed using the function svystatTM from ReGenesees, setting up the parameter deff=TRUE.

The intra-class correlation coefficient for *NSR* can be derived from the expression of *deff*, that is

$$\hat{\rho}_{i,NSR} = \frac{deff_i - 1}{b_{SNR} - 1}. \quad (7)$$

While it is not necessary to compute the intra-class correlation coefficient for SR strata because just one PSU is selected and the intra-class correlation is 1 by definition.

The last statistic to be defined is the estimator effect (EFFST, *effst*), that is, how much the sampling variance of the applied estimator under the adopted design is inflated or deflated with respect to the HT estimator, on the same sample. The *effst* is equal to

$$effst(\hat{\hat{Y}}_i)_{st} = \frac{\text{var}(\hat{\hat{Y}}_i)_{st}}{\text{var}(\hat{\hat{Y}}_i)_{HT}} \quad (8)$$

It is an optional parameter for R2BEAT, but it useful to take into account, from the allocation phase, of the impact on the estimates of a different estimator other than the Horvitz-Thompson estimator (*HT*). Indeed, the most applied estimator for the households survey is the calibrated estimator that, through the use of auxiliary variables, provides better estimates than HT.

Then, a reduction of the final sample size can be expected when applying the calibration estimator in place of the HT.

The optimal allocation is defined by R2BEAT solving the minimum optimisation problem

$$\begin{cases} C = \min \\ \mathrm{CV}\left(\hat{\hat{Y}}_{i,h}\right) \leq \delta\left(\hat{\hat{Y}}_{i,h}\right) & \begin{matrix} i = 1, \ldots, J \\ h = 1, \ldots, L \end{matrix} \end{cases},$$

where C is the global cost of the survey (if COST is equal to 1 in all the strata, $C = n$) and $\mathrm{CV}\left(\hat{\hat{Y}}_{i,h}\right)$ is the relative error we expect to observe estimating the Y_i variable in the stratum h (expected errors), that must be less than or equal to the precision constraints defined by the user. The solution is obtained with an iterative algorithm. Then, $\hat{S}$ (see, *e.g.*, expression (1) and (2)) is multiplied by the new design effect and the estimator effect and a new allocation is computed. The algorithm stops when the difference between two consecutive iterations is lower than a predefined threshold. At each step, an allocation is provided and the design effect is updated following the expression (6).

The package R2BEAT provides several outputs that help the evaluation of the allocation. For further details see the manual of the package available on-line.

It is important to point out that, as stated at the beginning of this section, the n is usually given. Then, to find the optimal allocation it is necessary to tune the precision constraints until the desired sample size is matched.

2.3 First stage selection (FS4)

FS4 is an open-source generalised software developed by Istat for stratification and selection of the PSUs in a two (or more) stages sampling design[8].

The function carries out the stratification of the PSUs, for each estimation domain, according to a size threshold (5). PSUs with a measure of size exceeding the threshold are identified as Self-Representing units, SR (see also previous section). The remaining Not Self-Representing units PSUs, NSR,

8 https://www.istat.it/en/methods-and-tools/methods-and-it-tools/design/design-tools/fs4.

are ordered by measure of their size and divided into strata whose sizes are approximately equal to the threshold multiplied by the number of PSUs to be selected in each stratum (MINCS). In this way, strata are composed of PSUs having as homogeneous as possible size. Next, the selection of a fixed number of PSUs per stratum is carried out using Sampford's method (unequal probabilities, without replacement, fixed sample size), implemented by the UPsampford function of the R package sampling.

The package FS4, through the function StratSel, stratifies and selects the PSUs to be sampled. Furthermore, the output provides the first order inclusion probability for each PSU, that is, for the first selection stage. The inclusion probability for the second stage can be easily obtained by dividing the number of SSUs to be selected in the PSU by the measure of size of the PSU. Then, the design weights for each unit in the sample are equal to the inverse of the product of the first-order and the second stage inclusion probabilities.

The inclusion probabilities of the first and second stage of the sampling scheme are such that the overall inclusion probabilities are almost constant for each unit in the sample, resulting in a self-weighting samples.

3. Sampling design of selected surveys

3.1 Labour Force Survey

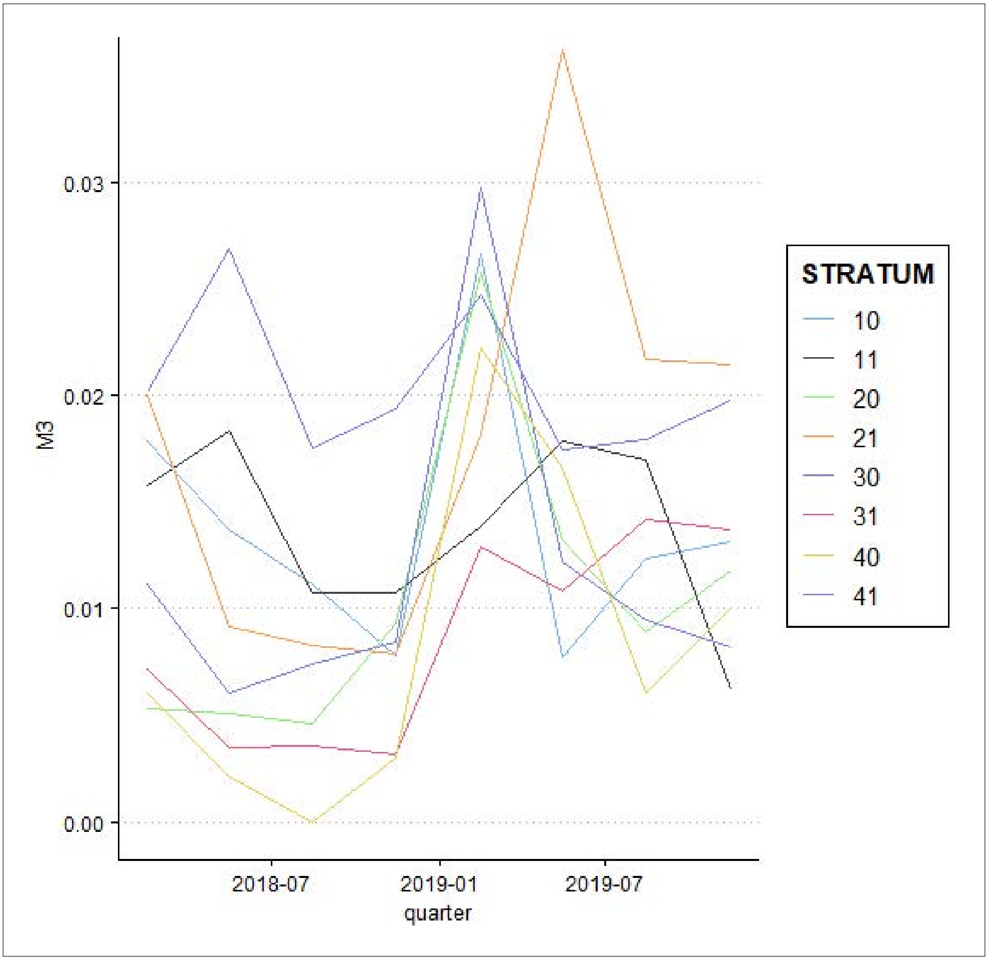
The new allocation of the Labour Force Survey (LFS) was determined on the basis of the main target variables, as established by NBS:

1. non-active (inactive);
2. employed (oc);
3. unemployed (som).

As already described in Section 2.2, the information from the 2018-2019 survey waves is useful for optimising the allocation. In detail, the estimates of the ratios of the target variables: namely non-active, employed, and unemployed population and their sampling variances, as well as the design effects (3) and the estimator effects (8) for the target estimators in each of the 8 available waves, was evaluated to provide the necessary input to R2BEAT, see Section 2.2. The R package Regeneeses was employed for this preliminary activity (Section 2.1).

To determine the sample allocation, different scenarios were analysed with different input information; indeed, having 2 years of LFS survey data available, one can use estimates from the last year only (*i.e.* the most recent information) or otherwise an average of the estimates from all surveys in the period 2018-2019. The latter alternative was chosen because it was more conservative in terms of variance. This was because in 2019 there was an increase in the unemployment rate in some areas, which therefore resulted in a lower variance in the latest period. Figure 2 depicts the unemployment rate for all the strata in the 8 quarters.

Figure 2 - Unemployment rate 18-19



Source: Our processing

Another important feature characterising the different alternative scenarios was the minimum number of interviews in each PSU. This parameter has an impact on the allocation via the design effect (see Equation 4).

Taking into consideration the differences in the rates in the two years, as mentioned above, the constraints on the required coefficients of variation of the estimators are “averaging” the current values. For the determination of sample size and its allocation within strata, the R package R2BEAT (see Section 2.2) was employed.

Let us see more in detail the main steps for determining the key input element for the allocation with R2BEAT. First, it was necessary to estimate the mean (rates) of the target variables, Following Section 2.1 to proceed in evaluating these quantities, we have to define the applied sampling design for the available data in the data frame lfs:

```
library(Regenesees)
des_19 <- e.svydesign(data = lfs, ids = ~ centr + ident_HH,
strata = ~ strat_new1,
weights = ~ w_corr, check.data = TRUE)
```

where ids is the identifier of the selection units, centr the PSU and ident_HH the household identifier, the sampling weights in the function are w_corr, the already corrected weights for non-response. Finally, strata represents the stratification of the sampling design. LFS weights are then calibrated to known totals for regions, urban or rural area, and age-sex groups. The calibration step in ReGenesees is easily carried out with the following function:

```
cal_18 <- e.calibrate(design=des_19, df.population=popfill,
calmodel= ~ reg + urb +gr_vr - 1,
calfun= "linear", bounds =   c(0.1,3))
```

Once the calibrated weights are calculated, the estimates and sampling errors can then be simply evaluated with the following function:

```
est_2018_reg_urb <- svystatTM(cal_18,y=~total+inact+oc+som +activ,
by = ~ reg:urb, estimator="Mean",
vartype=c("se"), deff=TRUE)
```

where by=~reg:urb provides the estimates for the most detailed domain, that is the strata for which the allocation is determined. Estimates can be evaluated at each desired level, specifying the value for by, appropriately.

The previous steps were applied to all 8 available waves and then the results were averaged to provide the input to R2BEAT.

The design effects in (4) of the two-stage design was evaluated on a new stratification composed of three classes:

- rural;

- urban;
- Chisinau.

This stratification was chosen based on the sampling design and the different features of rural and urban areas.

Moreover, only waves relating to 2019 were considered for design effect calculation, since the sampling design in terms of primary sampling units is different from the previous year and assuming the PSUs of the new sample design will have the most recent size.

Function svystat, similarly to svystatTM, evaluates the estimates and the design effects, in this case the option by requires the new specifically defined stratification stra1.

```
deff <- svystat(des_19,
kind ="TM",
estimator="Mean",
y= ~ inact+oc+som+activ,
by=~stra1,
deff=TRUE,
forGVF=TRUE)
```

Similarly, the estimator effect is the ratio between the variance of the calibrator estimator and the variance of the HT estimator under the applied sampling design. This component helps to determine the sampling size, taking into account that a calibration estimator is used instead of the HT estimator. As mentioned in Section 2.2, effst is an optional parameter for the software r2beat that can be used to consider the actual calibration estimator instead of the HT in defining a more efficient allocation.

Similarly to those seen above, the following instructions provide the ratio between $var(CAL)_{des}$ and $var(CAL)_{srs}$ in variable DEFF of the data frame effst:

```
effst <- svystat(cal_18,
kind ="TM",
estimator="Mean",
y= ~ inact+oc+som+activ,
by = ~ stra1,
deff=TRUE,
forGVF=TRUE )
```

The estimator effect can then be obtained by removing the design effect from the previous result.

```
effst$EFFST <- effst$DEFF/deff$DEFF
```

Other relevant quantities needed for allocation are the bNAR, average dimension of the PSU and b_{AR}, the average size of the families.

Finally, the input for the LFS allocation is:

```
strat=data.frame(STRATUM=as.numeric(N[,1]),
N=N[,2],M1=M1[,2],M2=M2[,2],M3=M3[,2])
strat$S1=sqrt(strat$M1*(1-strat$M1))
strat$S2=sqrt(strat$M2*(1-strat$M2))
strat$S3=sqrt(strat$M3*(1-strat$M3))
strat$CENS=0
strat$COST=1
strat$DOM1=1
strat$DOM2=substr(strat$STRATUM,1,1)
strat$DOM3=substr(strat$STRATUM,2,2)
strat$DOM4 <- strat$STRATUM
```

where N is the size of the stratum in terms of individuals, M1, M2, and M3 the average of values over the 8 waves of the three target variables and S1, S2, S3 their population variances (see also (1) for the applied expression in their evaluation), *DOM 1* the national domain, *DOM 2* the region, *DOM 3* the urban-rural classification and *DOM 4* the finer domain, *i.e.* the stratum. The design features are described in the object des:

```
des<-data.frame(STRATUM=des[,"STRATUM"], STRAT_MOS=des[,"PSU_MOS"],
DELTA=des[,"b_ar"],
MINIMUM=c(rep(36,6),rep(36,2)))
```

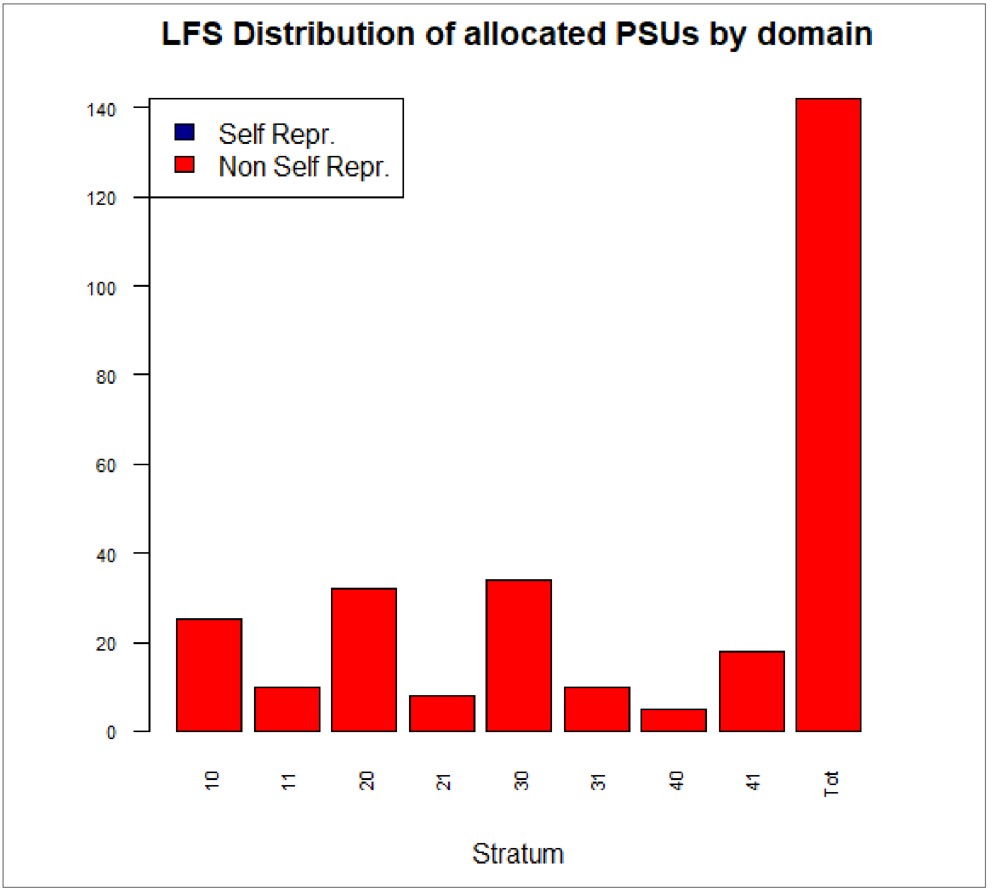
where STRAT_MOS is the PSU size in terms of individuals, DELTA is the average size of the secondary units (households) in terms of individuals, !MINIMUM! is, as recalled above, the min- imum number of households to be interviewed in each PSU. The sample size and its allocation among strata are reported in Table 1. An illustration of PSUs allocation is in Figure 3.

Table 1 - LFS allocation

Region	Urban	EA	Individuals	HH	Oversampling
1	0	25	2282	910	978
1	1	10	836	364	434
2	0	32	2912	1157	1252
2	1	8	653	284	344
3	0	34	3041	1325	1207
3	1	10	816	355	430
4	0	5	444	176	253
4	1	18	1433	646	1006
		142	12417	5101	6022

Source: Our processing

Figure 3 - LFS allocation among strata



Source: Our processing

Table 2 reports the expected errors of the three target variables for the different domains with the proposed allocation.

The corresponding yearly errors can be approximated by multiplying the previous errors by the following quantity:

$$\frac{1}{4}(1 + \frac{3}{4}\rho_1 + \frac{1}{8}\rho_3)$$

Table 2 - LFS Expected CVs

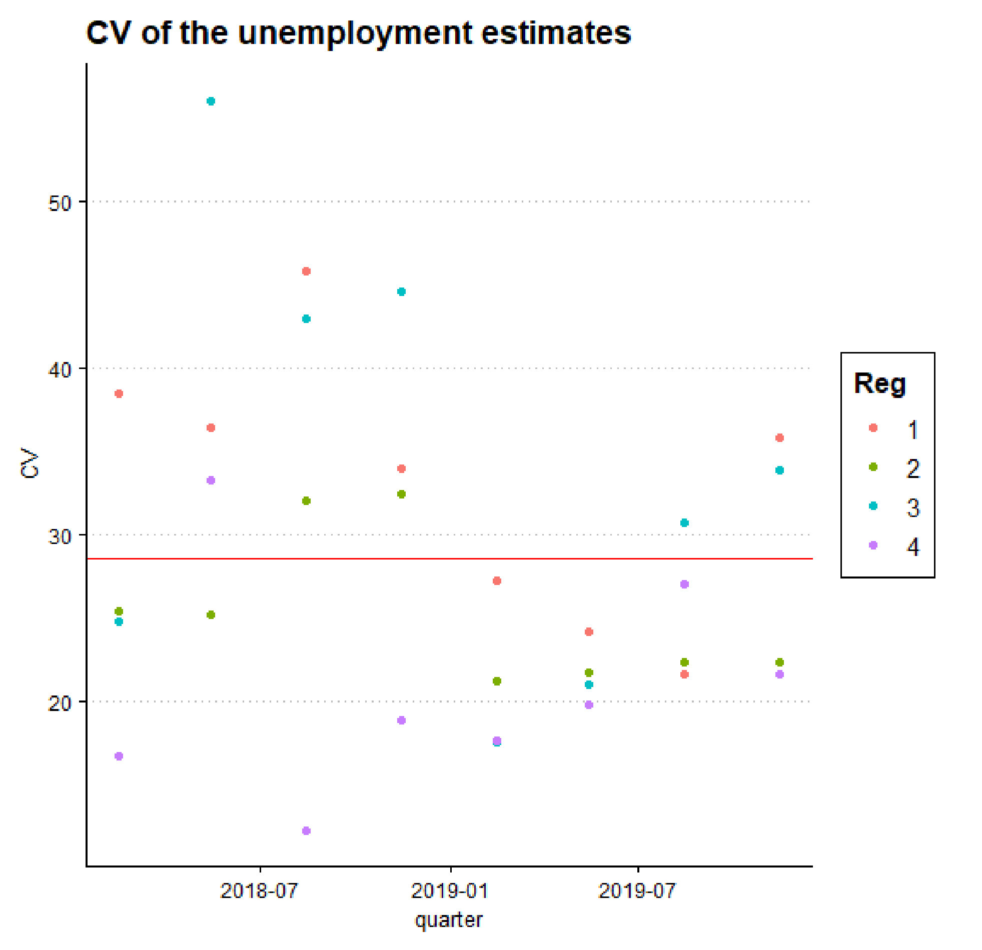
Type	Dom	Active	Empl	Unemp
National	1	0.0127	0.0326	0.1420
Region	1	0.0278	0.0620	0.2795
Region	2	0.0228	0.0745	0.2860
Region	3	0.0206	0.0673	0.2860
Region	4	0.0264	0.0437	0.2571
Urban	1	0.0170	0.0496	0.1965
Urban	0	0.0165	0.0325	0.2013

Source: Our processing

where ρ_1 is the correlation between quarters at one lag and ρ_3 is the correlation between quarters at three lags.

Figure 4 reports the observed errors of the unemployment rate estimates in the 8 waves of 2018/2019, while the horizontal red line represents the input constraints for the CV.

Figure 4 - LFS CV at regional level in 2018/2019 and required errors



Source: Our processing

Note that the new Eurostat regulation on social surveys would require tighter constraints, however the needed sample size would be much higher than the current one and for this reason, it would be difficult to apply.

3.1.1 A Proposal of gradual introduction of the new LFS sample

Instead of replacing the current sample with the new LFS sample all at once, we propose a gradual introduction that takes into account the LFS (2-2-2) panel scheme. In fact, since each quarter a panel (1/4 of the total sample) is replaced, we suggest that the new sample enters only for a portion, as a replacement of the exiting panel.

Let us denote the new panels from the new allocation with letters a-f and the panels of the current running panels with n-q, and with 1-4 the waves, *e.g.* for quarter T1-21 *a*1 is the entering panel selected with the new LFS allocation, whereas *n*4 is panel *n* at its 4th survey occasion, *o*3 panel *o* at its 3rd survey occasion, and so on. Similarly, for the following quarters. At quarter T2 22, all four panels are selected with the new allocation and the transition from the old to the new sample is complete, as follows:

T1	21	a1						n4	o3			r2
T2	21	a2	b1						o4	p3		
T3	21		b2	c1						p4	q3	
T4	21			c2	d1						q4	r3
T1	22	a3			d2	e1						r4
T2	22	a4	b3			e2	f1					

The proposed gradual introduction of the new LFS allocation has the main advantage of keeping the composition of the sample in each quarter in the 4 panels of households at different survey occasion (first, second, third, and fourth interview), thus preventing that for some quarters the composition of the sample is altered.

In fact, if we introduce all the new LFS sample in T1 21, for that quarter it will be composed of all households at their first interviews.

A well-known problem of panel surveys is differences in responses according to the number of occasions the units have been already interviewed, therefore a time-series break might occur if the composition is altered.

On average, this scheme does not require a much higher sample size during the transition between the old and new sample, since the two allocations are similar in terms of total sample size. However, there might be some issues due to different stratification of the current and new allocation and the global allocation can then be different from both. Moreover, as the panels come from different samples, the number of PSUs may be higher depending on how many PSUs should be kept from the previous allocation to maintain the panels on the last survey occasions.

Let us recall the rotation mechanism of PSU in different strata:

a. Chisinau: for the current design there is a rotation of the PSUs in Chisinau. The proposed scheme could be easily applied to this stratum. In fact, in Chisinau, after 6 consecutive quarters (*i.e.* completing the life-span of a panel) the PSU is removed from the survey. Each quarter 1 fourth of all PSUs is renewed when a new panel is entering the sample (*i.e.* 1/4 of the quarter sample), in this case, the new panel (and its corresponding PSUs) can be easily selected according to the new survey introducing the new allocation.

   For Chisinau, the number of allocated PSUs will be moderately larger depending only on the sampling stratification: following the previous scheme the entering panels *a*1, *b*1 and so on, will be selected in PSUs replacing the current ones.

b. Remaining strata - no rotation of PSUs is planned unless the PSU is exhausted. For these strata, extra PSUs would be needed for all the first year to account both for the closing panels of the current survey and the entering panel of the new allocation. To avoid this additional number of PSUs, the new PSU sample could be positively coordinated with the current one. However, as many PSUs are currently already almost fully investigated, this option could be difficult to apply in practice.

Note that the transition by itself does not increase the risk of exhaustion of the current PSUs. The households have been already selected, *e.g.* panel p4 in T3 21 has been already selected in quarter T2 20, similarly, all the panels n-r are to be selected by the end of 20 for the conduction of the current LFS scheme.

Finally, estimates from the two different sampling schemes (the current and the one) running at the same time with different size can be obtained either (a) by combining estimates from the different sources, for example, weighting the estimates with their sampling variance, or (b) by pooling data at the micro-level, for example by simply re-weighting the different sample weights to the known totals.

3.2 Household Budget Survey

Target variables of the Household Budget Survey have been defined by NBS as:

1. total pro-capite expenses (TE);
2. food pro-capite expenses (FE);
3. total pro-capite revenues (TR).

And the related precision constraints:

Domain	CV(TE)	CV(FE)	CV(TR)
National	0.02	0.01	0.04
Regional	0.04	0.02	0.06
Urban/Rural	0.02	0.02	0.04

The 8 rounds of the survey in 2018 and 2019 were available. Due to the high rise in non-response rates occurred in 2019, only this year has been retained, as we wanted to be conservative by considering the worst situation.

For each quarter of 2019, estimates of the three target variables have been calculated by using ReGenesees.

```
library(Regenesees)
des_19rev_2 <- e.svydesign(data = hb19rev_2, ids = ~ cod_ter + id_hh,
strata = ~ strat_new1,
weights = ~ wcor, check.data = TRUE)
```

Note that initial weights are the ones already treated to handle non-response.

Calibration estimates have been obtained by using known totals (total population by region and rural/urban, number of children and number of retired males and females):

```
cal_19rev_2 <- e.calibrate(design=des_19rev_2, df.population= popfill,
calmodel= ~ reg + urb_rur + n_copii + n_pens_f + n_pens_m - 1,
calfun= "linear", bounds =   c(0.001,10))
```

Then, inputs for two-stage allocation:

- the variability of target variables in strata;
- the design effect (*deff*);
- the estimator effect (*effst*);
- the intra-cluster correlation coefficient ($\hat{rho}$)

have been prepared in the same way already described for the Labour Force Survey, with a key difference, due to the different nature of the target variables, *i.e.* categorical in LFS and continuous in HBS: the variance in strata for HBS target variables has been calculated with the method of moments, as reported in (2).

The allocation and PSUs selection have been performed in two separate executions, one regarding strata not including Chisinau (urban and rural), and the other one only for the two strata of Chisinau. The reason for this procedure is because the PSU selection (function StratSel in FS4 package) does not allow to define different minimum numbers of SSUs for each selected PSU, while after different trials we realised it was necessary to increase the number of PSUs only in Chisinau strata by fixing a lower minimum number of SSUs per PSU (12, compared to the value 30 used for other strata). In this way, we were able to increase the number of PSUs and decrease the SSU allocation in Chisinau, otherwise exceedingly high.

The following steps were executed for strata not including Chisinau. First, the optimal allocation:

```
stratif_all <- stratif[stratif$STRATUM!=40 & stratif$STRATUM!=41,]
errors_all <- errors
des_file_all <- des_file[des_file$STRATUM!=40 & des_file$STRATUM!=41,]
des_file_all$MINIMUM <- 30
psu_file_all <- psu_file[psu_file$STRATUM!=40 & psu_file$STRATUM!=41,]
rho_all <- rho[rho$STRATUM != 41,]
effst_all <- effst[effst$STRATUM!=40 & effst$STRATUM!=41,]

alloc <- beat.2st(stratif_all,
errors_all,
des_file_all,
psu_file_all,
rho_all,
deft_start = NULL,
effst_all)
```

and then the selection of PSUs:

```
allocat <- alloc$alloc[-nrow(alloc$alloc),]

sample_2st <- StratSel(dataPop= psu_file_all,
idpsu= ~ PSU_ID,
dom= ~ STRATUM,
final_pop= ~ PSU_MOS,
size= ~ PSU_MOS,
PSUsamplestratum= 3,
min_sample= min,
min_sample_index= FALSE,
dataAll= allocat,
domAll= ~ factor(STRATUM),
f_sample= ~ ALLOC,
planned_min_sample= NULL,
launch= F)
```

Same for Chisinau strata, first the optimal allocation:

```
stratif_4 <- stratif[stratif$STRATUM==40 | stratif$STRATUM==41,]

errors_4 <- errors[2:3,]
errors_4$CV1 <- mean(errors_4$CV1)
errors_4$CV2 <- mean(errors_4$CV2)
errors_4$CV3 <- mean(errors_4$CV3)
errors_4$DOM <- c("DOM1","DOM2")

des_file_4 <- des_file[des_file$STRATUM==40 | des_file$STRATUM==41,]
des_file_4$MINIMUM <- min
psu_file_4 <- psu_file[psu_file$STRATUM==40 | psu_file$STRATUM==41,]
rho_4 <- rho[rho$STRATUM==40 | rho$STRATUM==41,]
effst_4 <- effst[effst$STRATUM==40 | effst$STRATUM==41,]

alloc_4 <- beat.2st(stratif_4,
errors_4,
des_file_4,
psu_file_4,
rho_4,
deft_start = NULL,
effst_4)
```

and then the PSUs selection:

```
allocat <- alloc_4$alloc[-nrow(alloc_4$alloc),]

sample_2st_4 <- StratSel(dataPop= psu_file_4,
idpsu= ~ PSU_ID,
dom= ~ STRATUM,
final_pop= ~ PSU_MOS,
size= ~ PSU_MOS,
PSUsamplestratum= 3,
min_sample= min,
min_sample_index= FALSE,
dataAll= allocat,
domAll= ~ factor(STRATUM),
f_sample= ~ ALLOC,
planned_min_sample= NULL,
launch= F)
```

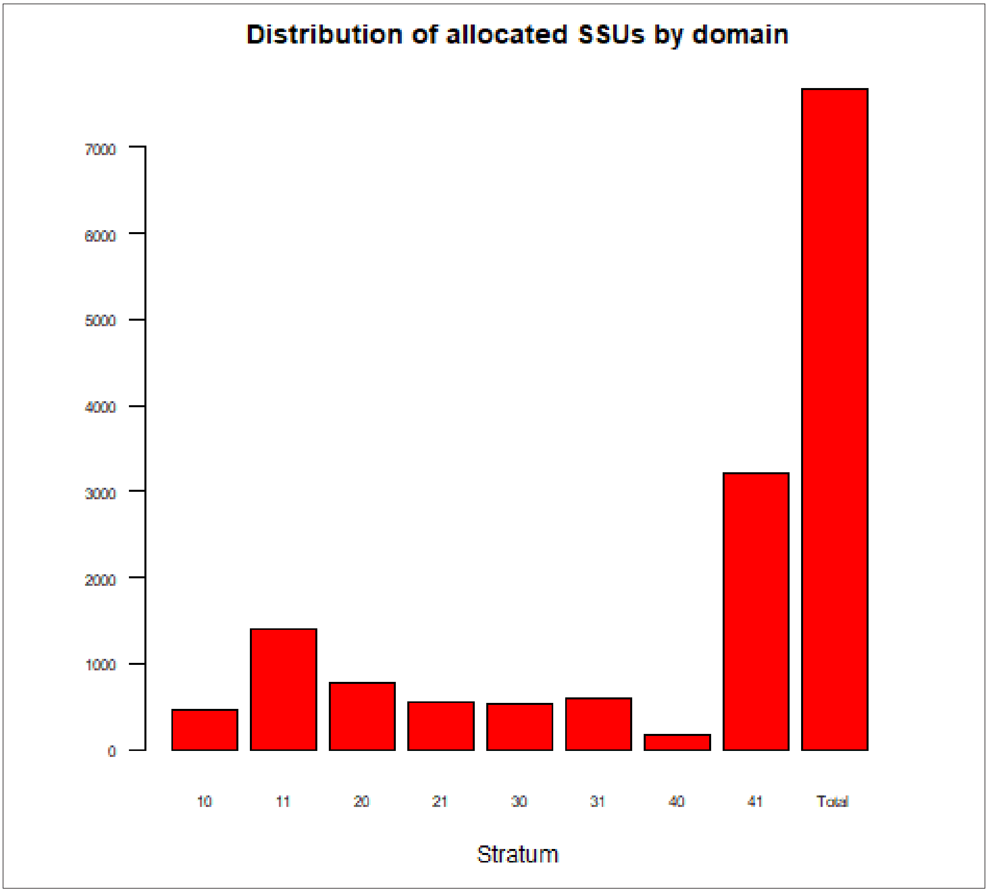
The outputs of the separate executions have been put together, obtaining the overall two- stage sample design reported in Table 3.

Table 3 - HBS allocation

Region	Urban	EA	HH	Oversampling
1	0	12	336	460
1	1	18	579	1396
2	0	15	465	781
2	1	9	297	548
3	0	12	324	518
3	1	12	339	603
4	0	3	48	162
4	1	57	684	3218
Total		138	3072	7886

Source: Our processing

Figure 5 - HBS optimal allocation



Source: Our processing

The oversampling is obtained considering the non-response rate observed in the previous occasion of the survey that are:

Region	Urban	Non-response rate
1	0	0.27
1	1	0.59
2	0	0.40
2	1	0.46
3	0	0.37
3	1	0.44
4	0	0.70
4	1	0.79

The expected errors are reported in Table 4.

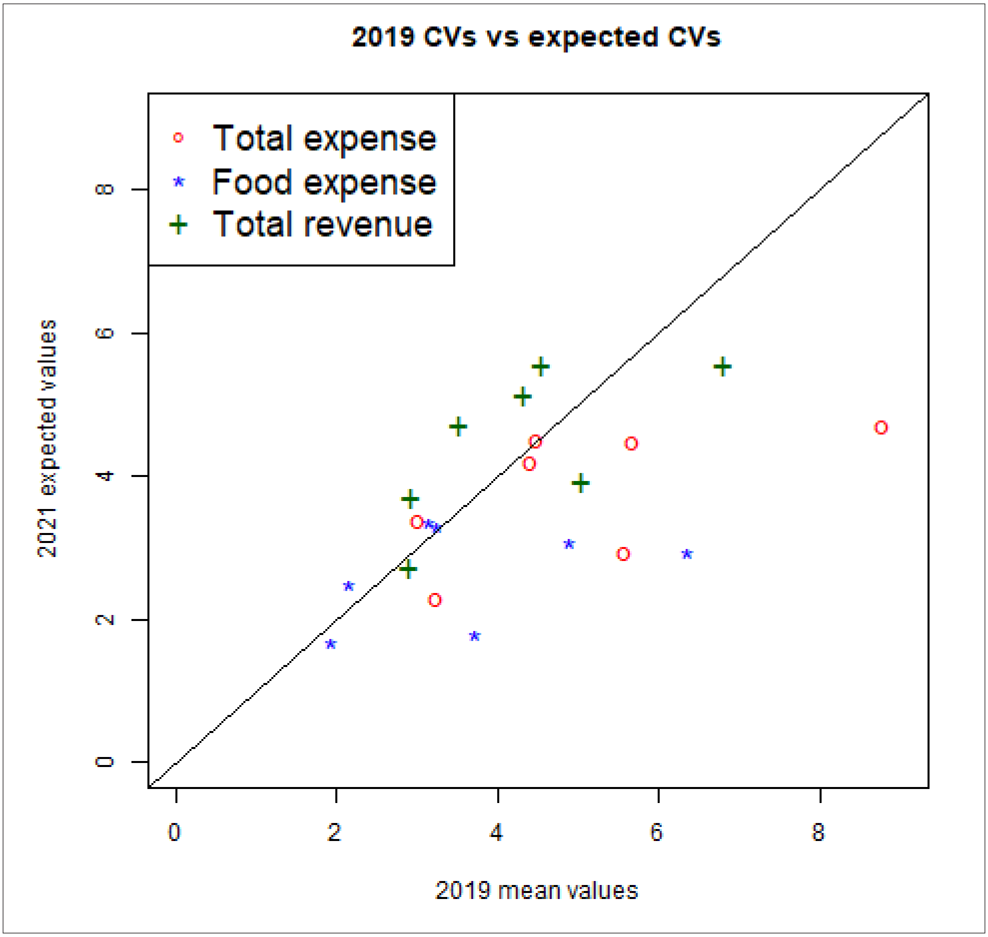
Table 4 - HBS expected CVs

Type	Dom	CV(TE)	CV(FE)	CV(TR)
National	1	0.0228	0.0166	0.0273
Region	1	0.0418	0.0330	0.0556
Region	2	0.0451	0.0334	0.0514
Region	3	0.0446	0.0306	0.0472
Region	4	0.0470	0.0294	0.0554
Urban	1	0.0337	0.0249	0.0394
Urban	0	0.0293	0.0179	0.0370

Source: Our processing

Figure 6 reports how these expected coefficients of variation are with respect to those observed in HBS 2019 rounds.

Figure 6 - HBS 2019 CVs vs expected CVs



Source: Our processing

Analysing this Figure, it can be seen that there is a general gain in terms of precision for variables "Total expense" and "Food expense", while for "Total revenue" in some cases there are gains and losses in others.

3.3 Domestic Violence Survey

The new allocation of the Domestic Violence surveys has been defined on the basis of the three main target variables chosen by the NBS:

- psychological violence,
- physical violence and
- sexual violence.

The input values for the sample allocation assumes rates equal to the estimates obtained in the last survey in 2010.

Similarly to the process for determining the allocations of 3.1 and 3.2, first the target variables estimates, the sampling variances, design effects ((3)) and estimators effects ((8)) are obtained by means of the software ReGenesees (see Section 2.1) and then the allocation via the software R2BEAT.

In particular, the relevant features of the previous occasion of the Domestic Violence survey that shall be maintained are:

- the sample size approximately equal to 1,500 interviews and PSUs equal to 150,
- for each sample SSU (household) only a female within is selected,
- the calibration is obtained on the number of females for national level, region, urban/rural and age-sex classes.

Given these features, the MINIMUM parameter (minimum number of interviews per PSU) has been set approximately as the one of the last survey, *i.e.* 10 households/interviews per PSU. Note also that in this case the number of final units is equivalent to the number of SSUs (households).

The resulting allocation by domain is reported in Table 5, while Figure 7 illustrates the PSUs allocation in strata.

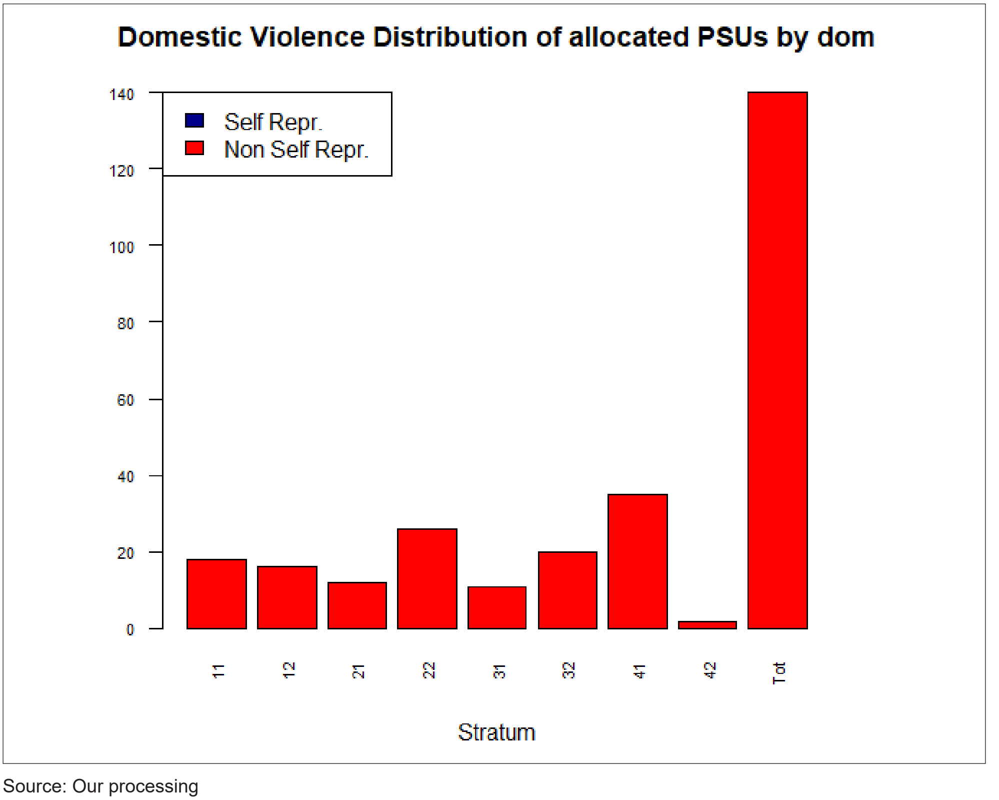
Table 5 - Domestic Violence survey allocation

Region	Urban	EA	Individuals	HH	Oversampling
1	0	18	180	180	237
1	1	20	199	199	311
2	0	28	280	280	368
2	1	14	134	134	209
3	0	22	221	221	291
3	1	12	121	121	189
4	0	2	20	20	26
4	1	38	380	380	594
Total		154	1535	1535	2225

Source: Our processing

To guarantee the required sample size, an oversampling is obtained on the basis of the response rate of the 2012 survey, that is: 0.76 and 0.64 in urban and rural areas, respectively.

Figure 7 - HBS 2019 CVs vs expected CVs



Source: Our processing

Finally, Table 6 reports the expected CV for the three target variables.

Table 6 - Domestic Violence survey expected CVs

Type	Dom	Ps-vio	Ph-vio	S-vio
National	1	0.0319	0.0471	0.0718
Region	1	0.0640	0.0889	0.1503
Region	2	0.0566	0.0792	0.1251
Region	3	0.0763	0.0992	0.1599
Region	4	0.0593	0.1223	0.1422
Urban	1	0.0373	0.0659	0.1078
Urban	0	0.0522	0.0661	0.0946

Source: Our processing

3.4 Energy Consumption Survey

Target variables of the Energy Consumption Survey have been defined as:

1. Coal, consumption (kg);
2. Natural gas, consumption (m^3);
3. Liquefied (petroleum) gases, consumption[9] (l);
4. Lighters and pellets, consumption (kg);
5. Firewood, consumption (m^3);

and the related precision constraints are reported in Table 7.

Table 7 - Energy Consumption, precision constraints CVs

Domain	Precision constraints, CV					
	Coal	Natural gas	Liquefied petrolium gas	Lighters and pellets	Firewood	Electricity
National	0.0872	0.0200	0.0600	0.9999	0.0700	0.0600
Regional	0.4022	0.1069	0.2667	0.9999	0.1731	0.0629
Urban/Rural	0.1635	0.0689	0.1271	0.9999	0.1144	0.0408

Source: Our processing

The data of the last occasion of the survey were available.

The estimates of the target variables have been calculated by using ReGenesees, as follows:

9 Because of its low level, it is used just as control variable.

```
library(Regenesees)
desHT <- e.svydesign(data=ene, ids=~ centr + id_cce,
strata=~strat,
weights=~w_cor,
fpc = NULL, self.rep.str = NULL,
check.data = TRUE)
```

Note that initial weights are the ones already treated to handle non-response.

Calibration estimates have been obtained by using known totals (consumption of natural gas (m^3) [ccf2321] and consumption of electricity (kWh) [ccf23131] by region and coal procurement at national level [ccf23101]):

Because the calibration was already performed by NBS, the function ext. calibrated is used:

```
des <- ext.calibrated(data=ene, ids=~ centr + id_cce,
strata=~strat,
weights=~w_cor,
fpc = NULL, self.rep.str = NULL,
check.data = TRUE,
weights.cal=~wfinal,
calmodel=~region:(ccf2321 + ccf23131) + ccf23101 - 1)
```

Then, inputs for two-stage allocation have been prepared:

- the variability of target variables in strata;
- the design effect (*deff*);
- the estimator effect (*effst*);
- the intra-cluster correlation coefficient ($\hat{p}$).

as already described.

Finally, the optimal allocation was obtained, using the function beat.2st of R2BEAT:

```
alloc <- beat.2st(stratif,
errors,
des_file,
psu_file,
rho,
deft_start = NULL,
effst,
epsilon1 = 5,
mmdiff_deft = 1,maxi = 15,
epsilon = 10^(-11), minnumstrat = 2, maxiter = 200, maxiter1 = 25)
```

and setting MINIMUM=36.

The selection of the PSUs was performed through the function StraSel of FS4:

```
sample_2st <- StratSel(dataPop= psu_file,
idpsu= ~ PSU_ID,
dom= ~ STRATUM,
final_pop= ~ PSU_MOS,
size= ~ PSU_MOS,
PSUsamplestratum= 1,
min_sample= 36,
min_sample_index= FALSE,
dataAll= allocat,
domAll= ~ STRATUM,
f_sample= ~ ALLOC,
planned_min_sample= NULL,
launch= F)
```

and setting MINCS=1.

An overview of the allocation of the two-stage sample design is provided in Table 8.

The oversampling was obtained considering the non-response rates observed in the previous occasion of the survey and shown in Table 9.

Table 8 - Energy Consumption allocation

Region	Urban	EA	HH	Oversampling
1	0	26	928	1006
1	1	11	396	515
2	0	13	462	501
2	1	3	106	143
3	0	12	447	495
3	1	4	141	169
4	0	2	55	64
4	1	27	955	1441
		98	3490	4334

Source: Our processing

Table 9 - Energy Consumption, non-response rate observed in the previous occasion of the survey

Region	Urban	Non-response rate
1	0	0.08
1	1	0.23
2	0	0.08
2	1	0.26
3	0	0.09
3	1	0.16
4	0	0.14
4	1	0.36

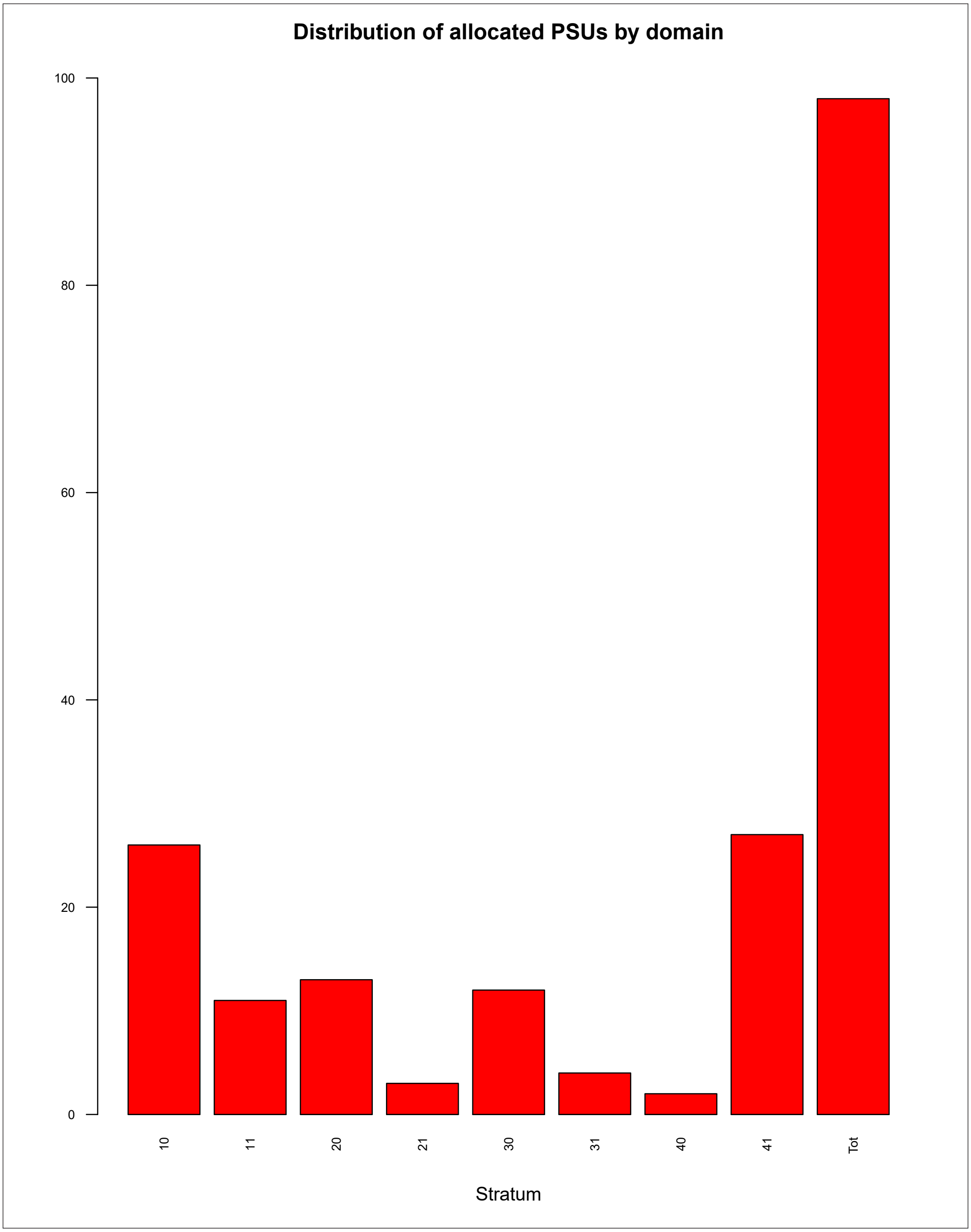
Source: Our processing

Table 10 - Energy Consumption, expected CVs

Type	Dom	Coal	Natural gas	Liquefied petrolium gas	*Lighters and pellets*	Firewood	Electricity
national	1	0.0872	0.0356	0.0354	0.1977	0.0354	0.0263
regional	1	0.0946	0.0797	0.0443	0.3650	0.0452	0.0329
regional	2	0.2553	0.1069	0.0718	1.0533	0.0680	0.0565
regional	3	0.2004	0.0635	0.0663	0.2651	0.0578	0.0629
regional	4	0.4022	0.0487	0.2667	0.5734	0.1731	0.0532
urban	1	0.1004	0.0689	0.0360	0.2307	0.0368	0.0333
urban	0	0.1635	0.0413	0.1271	0.3775	0.1144	0.0408

Source: Our processing

Figure 8 - Optimal allocation of Energy Consumption survey, PSUs by stratum

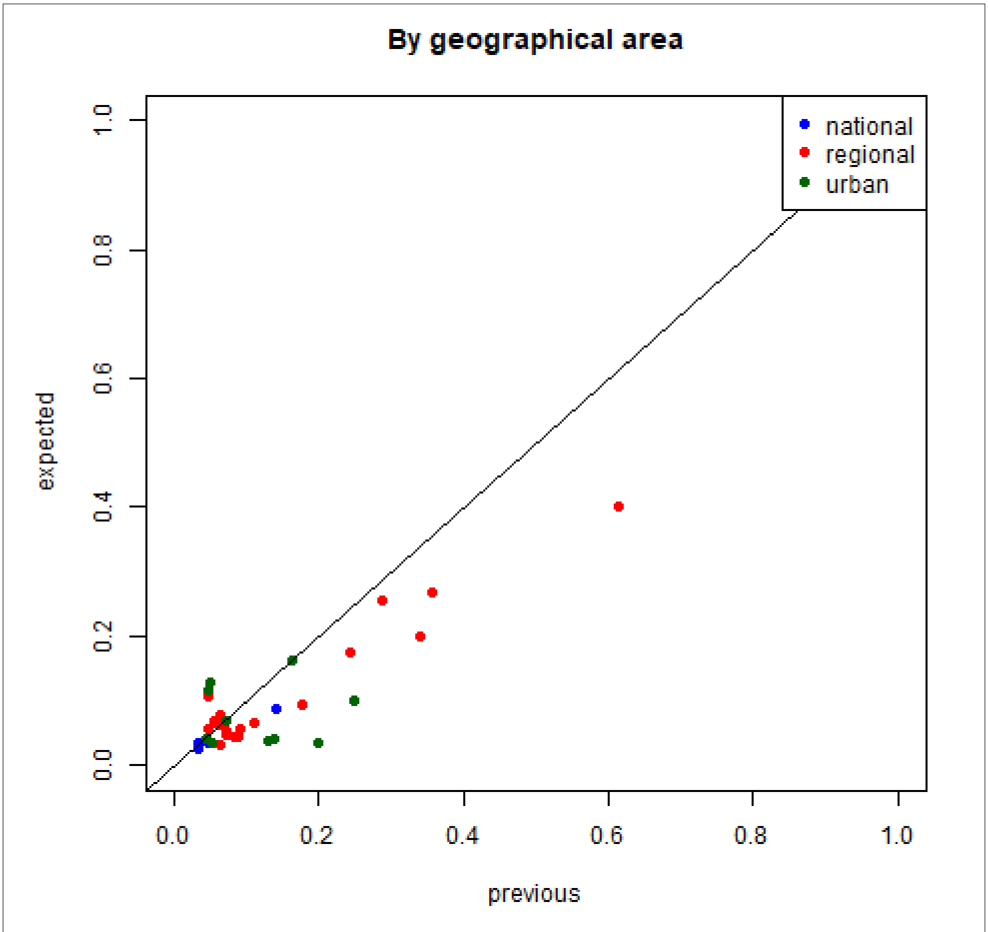


Source: Our processing

The expected errors for the Energy Consumption survey are in Table 10.

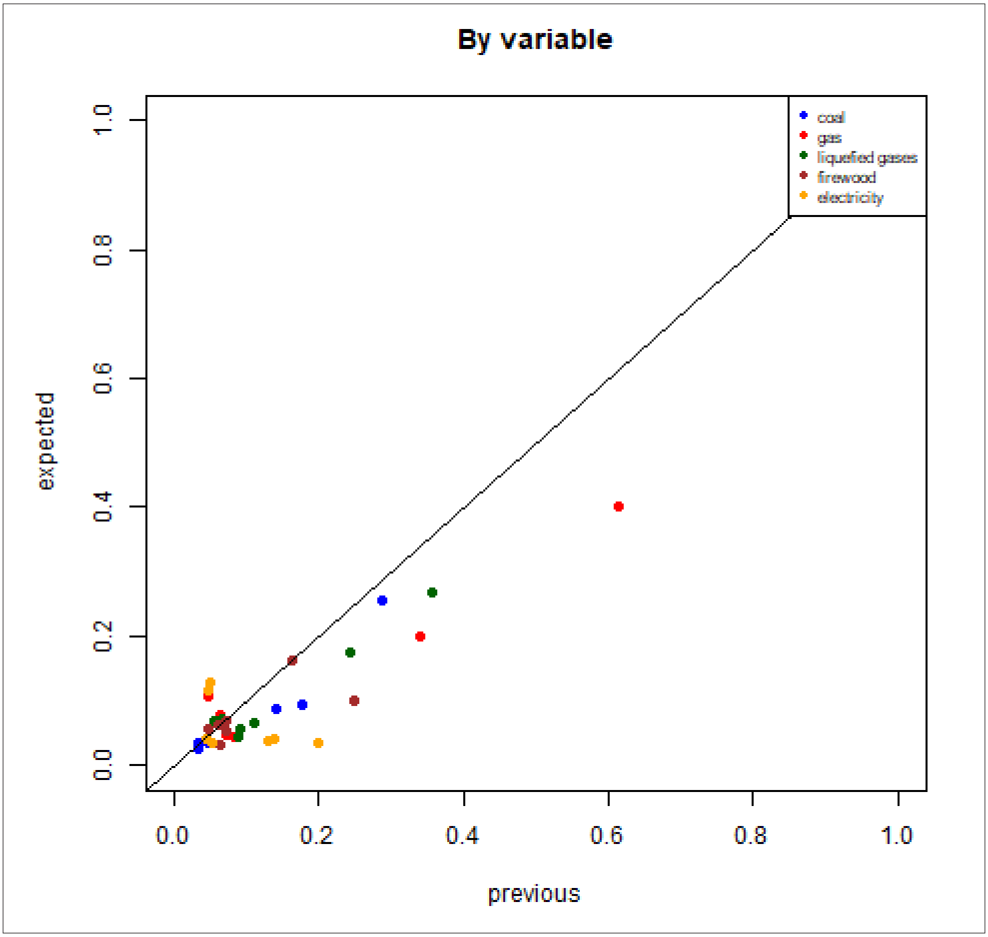
How these expected coefficients of variation are concerning those observed in previous occasions? The comparison with respect to interest variables and geographical area is reported in Figures 9 and 10, respectively.

Figure 9 - Energy Consumption previous CVs vs expected CVs, by geographical areas



Source: Our processing

Figure 10 - Energy Consumption previous CVs vs expected CVs, by interest variables



Source: Our processing

Analysing these figures, it can be seen that a general gain in terms of precision for variables is expected using the proposed allocation.

4. Survey coordination

The structure of the master sample has been built through the coordination of the four illustrated surveys.

In the literature, several sample coordination methods have been developed to obtain a maximal or minimal overlapping between samples drawn at different occasions (see, *e.g.* Matei and Skinner, 2009 and reference therein).

It is possible to distinguish, with respect to a survey taken as reference (in our case LFS), between negative and positive coordination. In the former, the aim is that PSUs with a high probability to be included in the sample of the reference survey have a low probability to be included in the sample of the other survey. The overlapping between the two samples is expected to be minimum and the sample of the two surveys is expected to be spread among different PSUs. On the contrary, for positive coordination, PSUs with a high probability to be included in the sample of the reference survey have a high probability to be included in the sample of the other survey. The overlapping between the two samples is expected to be maximum and the samples of the two surveys are expected to be concentrated in fewer PSUs.

It is important to point out that the two samples are still probabilistic. The choice between negative and positive coordination depends on the need of spreading as widely as possible the sample across PSUs to avoid exhausting households in PSUs and reduce the design effects for the master sample estimates (that is, the estimates obtained joining the same variables collected in the four surveys), and the need to concentrate the interviews due to the constraints established by the organisation of the data collection network. In our case, the main aim was to coordinate different surveys to minimise as much as possible the number of PSUs of the master sample, to match the number of available interviewers.

In the following sections, the scenario of no coordination and the scenario of maximum coordination are illustrated in detail.

4.1 No-coordination

The no coordination scenario was obtained by drawing independently the PSUs for the four surveys. The results we obtained represent the upper bound in terms of PSUs included in the master sample. In the following tables the number of PSUs for the surveys taken two by two, in case of no coordination, are reported:

	LFS		
HBS	0	1	
0	3421	138	3559
1	134	4	138
	3555	142	3697

	LFS		
Violence	0	1	
0	3414	129	3543
1	141	13	154
	3559	142	3697

	HBS		
Violence	0	1	
0	3414	129	3543
1	145	9	154
	3559	138	3697

	LFS		
Energy	0	1	
0	3465	134	3599
1	90	8	98
	3555	142	3697

	HBS		
Energy	0	1	
0	3467	132	3599
1	92	6	98
	3559	138	3697

All the details on this scenario are in Section 5.1.1.

4.2 Maximum coordination

Positive sample coordination has been obtained by considering the LFS PSUs sample a pivot, and coordinating the HBS, Domestic Violence and Energy Consumption samples with it, favouring the PSUs with a higher number of households. Anyway, positive coordination has never been applied for PSUs in the Chisinau urban stratum, due to the very low average number of households.

To obtain maximum coordination, a deterministic approach has been used. The surveys (HBS, Domestic Violence, Energy Consumption) were considered separately. Looking at each stratum, the number of PSUs already selected for the LFS sample and those required for each survey are counted. Three were the possible situations for the sample of PSUs of the coordinated surveys:

1. if the needed number of PSUs for one of the other surveys was equal to the number of PSUs selected for LFS, then the latter were all included;
2. if the needed number of PSU for one of the other surveys was lower than the number of PSUs selected for LFS, then only the LFS PSUs with higher MOS were included;
3. if the needed number of PSU for one of the other surveys was higher than the number of PSUs selected for LFS, then the PSUs selected for LFS were supplemented with the PSUs originally selected for the surveys in case of independent selection.

In the following tables the number of PSUs for the surveys taken two by two, in the case of maximum coordination, are reported:

	LFS		
HBS	0	1	
0	3487	72	3559
1	68	70	138
	3555	142	3697

	LFS		
Violence	0	1	
0	3473	70	3543
1	82	72	154
	3569	142	3697

	HBS		
Violence	0	1	
0	3452	91	3543
1	107	47	154
	3559	138	3697

	LFS		
Energy	0	1	
0	3506	93	3599
1	49	49	98
	3555	142	3697

	HBS		
Energy	0	1	
0	3511	88	3599
1	48	50	98
	3559	138	3697

All the details on this scenario are in Section 5.1.2.

5. Master sample

Two alternative versions of the master sample have been evaluated, depending on the sample coordination of the PSUs for the four different surveys:

- master sample with no coordination;
- master sample with maximum coordination of PSUs.

The master sample with no coordination is characterised by a higher number of PSUs, both in terms of the number of selected ones and in terms of their joint use in each quarter. On the contrary, the master sample obtained by positive sample coordination of the PSUs is characterised by a lower number of initial PSUs, with a higher rate of exhaustion. For both scenarios, a simulation has been carried out to evaluate the exhaustion rate of the PSUs and the need for renewal.

The advantage of the no-coordination solution is that it requires almost no renewal of PSUs outside Chisinau urban. The drawback is in the high number of PSUs simultaneously used in each quarter (270). Conversely, the maximum-coordination scenario requires an additional (limited) renewal of exhausted PSUs out of Chisinau urban (19), but the number of PSUs to be accessed each quarter is lower: 210 instead of 270.

5.1 Master sample simulation of use

The use of the master sample during its life span (from 2021 to 2025, for a total of 20 quarters) has been simulated. The simulation was carried out differently for "Chisinau urban" and the remaining strata.

In other than "Chisinau urban" strata:

1. in quarter t=1 (first quarter of 2021) the number of households to be interviewed are assigned to all LFS and HBS PSUs, the relative PSUs' household counter is consequently updated;
2. in quarter t+1, an attempt is made to assign the corresponding number of LFS or HBS households to be interviewed; it is successful only if not yet interviewed households are sufficient to meet the required number of households:

   - if it is successful, the respective household counter is increased by the new amount;
   - otherwise, the PSU is declared as "exhausted" and has to be renewed: the household counter is set to zero and is valorised by adding the current number of LFS and/or HBS households to be interviewed;
3. step 2 is repeated until the end of the period (quarter t=20);
4. without indicating the quarter, step 1 is performed for PSUs belonging to Energy Consumption and/or Domestic Violence.

In "Chisinau urban" the flow of operations is similar, but with two important differences:

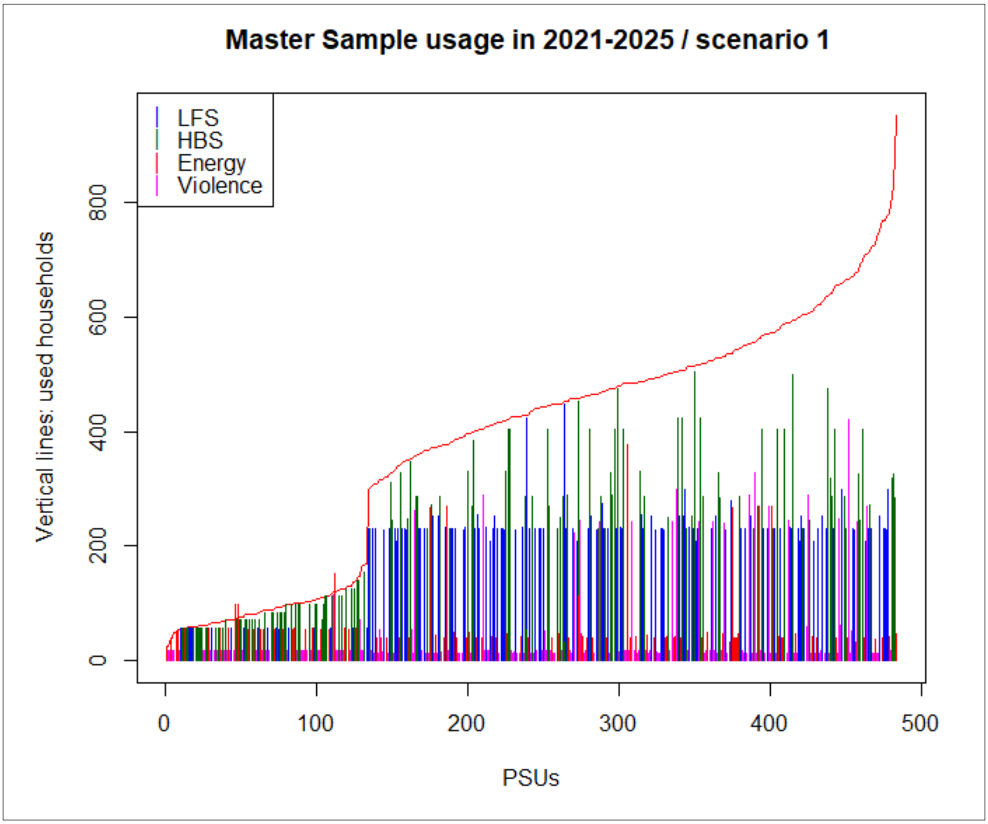
- each PSU can be used by only one survey;
- in the case of PSUs used by LFS, a different exhaustion criterion is adopted: a single LFS PSU is used only for a rotation group, and after 6 consecutive quarters the PSU is removed from the survey.

5.1.1 Simulation of use of the master sample with no coordination of PSUs

In Figure 11 the use of the master sample with no coordination of PSUs is displayed. Alongside the x-axis, PSUs are ordered by size (total number of households), that is indicated by the red line. Each vertical bar indicates the use of households by the four surveys (characterised by different colours). Each time the vertical bar exceeds the red line, then the corresponding PSU has to be substituted. In the no-coordination scenario we have the following situation:

- number of PSUs in master sample: 486 (135 in Chisinau urban)
- number of PSUs contemporarily used in each quarter: 270 (75 in Chisinau urban)
- exhaustion/renewal of available PSUs out of Chisinau urban: 0
- exhaustion/renewal of available PSUs in Chisinau urban: 193

Figure 11 - Simulation of use of the master sample with no coordination of PSUs (scenario 1)



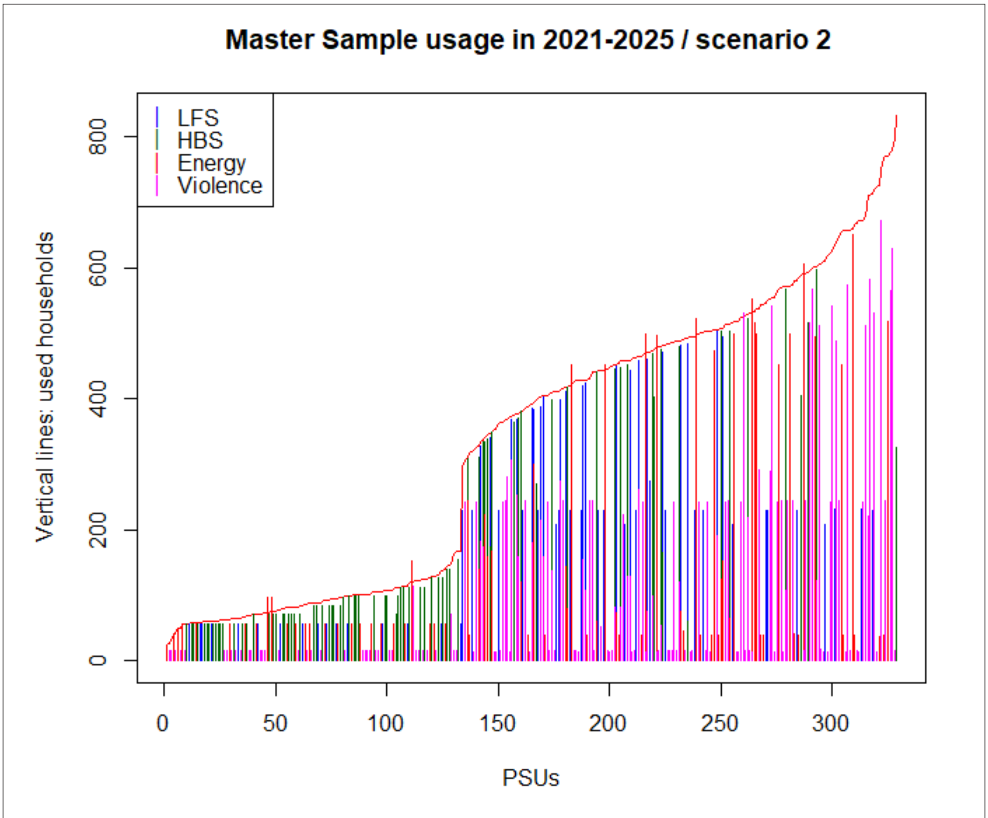
Source: Our processing

5.1.2 Simulation of use of the master sample with coordination of PSUs

In Figure 12 the use of the master sample with maximum coordination of PSUs is displayed. In the maximum coordination scenario we have the following situation:

- Number of PSUs in master sample: 332 (135 in Chisinau urban)
- Number of PSUs contemporarily used in each quarter: 210 (75 in Chisinau urban)
- Exhaustion/renewal of available PSUs out of Chisinau urban: 19
- Exhaustion/renewal of available PSUs in Chisinau urban: 193

Figure 12 - Simulation of use of the master sample with coordination of PSUs (scenario 2)



Source: Our processing

5.2 Selection of additional PSUs to substitute exhausted ones

As illustrated in the simulation, whichever solution is chosen, during the use of the master sample some PSUs will get exhausted, and an additional list of replacement PSUs will be required. It is fundamental to substitute each exhausted PSU with a new one using the following criteria. The new PSU has to be selected.

- In the same stratum utilised at the time of initial selection of the PSUs, *i.e.* in the stratum defined during the execution of the FS4 function StraSel;

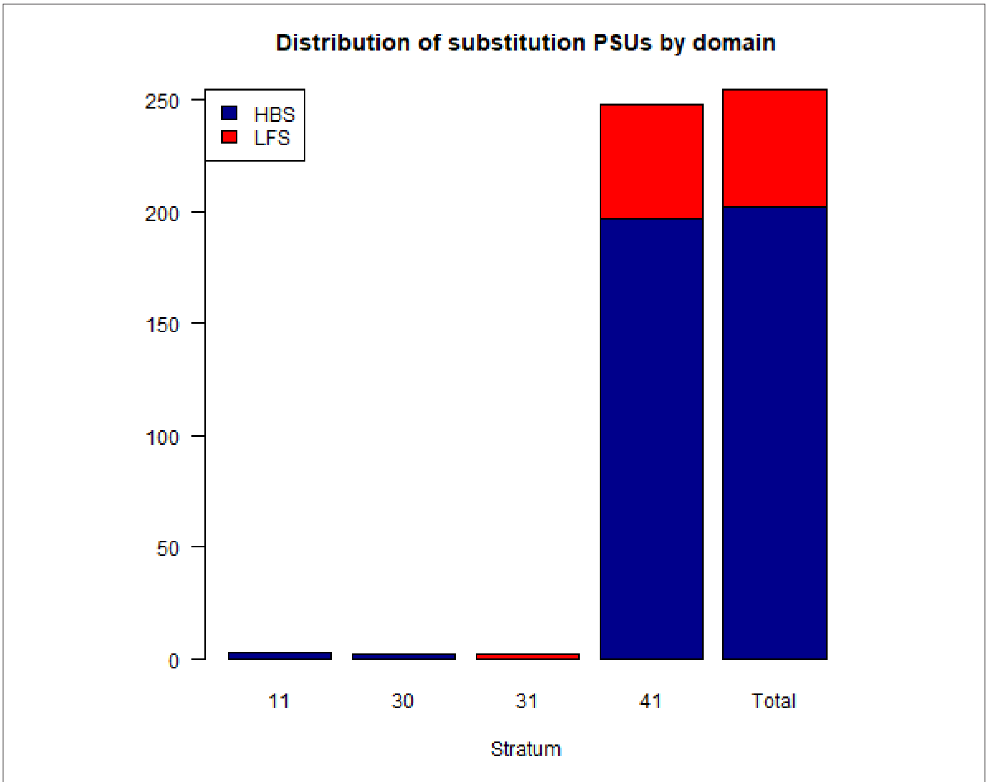
- with approximately the same size of the exhausted one, in order to preserve the first stage inclusion probability.

Having this in mind, two different lists of substitution PSUs have been produced, one for the no-coordination solution and another one for the maximum-coordination solution.

In the case of the no-coordination solution, a list is available with a total of 255 PSU: 202 for HBS and 53 for LFS. In the case of the maximum-coordination solution, the list includes 291 PSU, of which 219 for HBS and 72 for LFS.

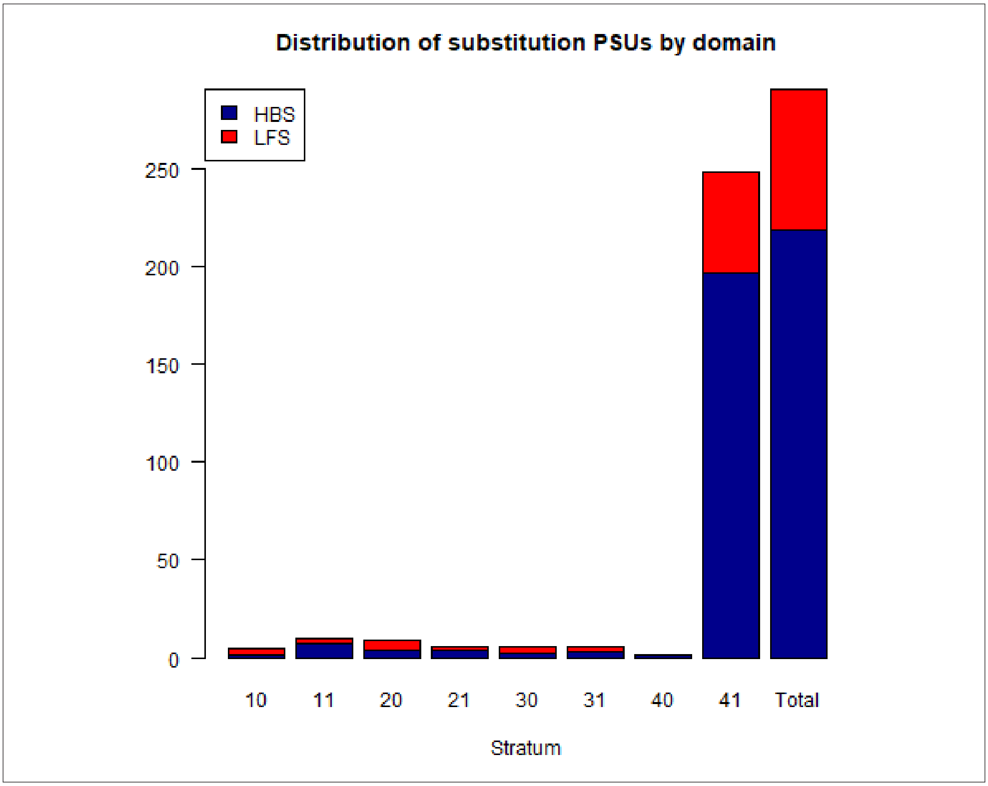
The distribution of additional PSUs in strata under the two scenarios of implementation of the master sample is visualised in Figure 13.

Figure 13 - Distribution of additional PSUs in strata
(a) No-coordination scenario



Source: Our processing

Figure 13 - Distribution of additional PSUs in strata
(b) Maximum-coordination scenario



Source: Our processing

By analysing these distributions, it is clear that there is a high concentration of substitution PSUs in "Chisinau urban", even more evident in the case of the first scenario.

6. Conclusions

Based on the sample designs of the four surveys that will make use of the new master sample during its life span (2021-2025), two possible configurations of the master sample have been defined, depending on the degree of sample coordination of the PSUs required by each survey.

The main advantage of the no-coordination solution is that it requires a very limited renewal of exhausted PSUs outside Chisinau urban. However, its drawback is the high number of PSUs that are simultaneously used in each quarter. Conversely, the maximum-coordination scenario requires a non-negligible renewal of exhausted PSUs out of Chisinau urban, but the number of PSUs to be accessed each quarter is lower.

This last feature (number of PSUs to be accessed each quarter) is crucial, as it has a strong impact on the feasibility and sustainability of the solution. Indeed, the current data collection network organisation is based on the availability of only 150 interviewers, where each interviewer covers only one PSU. If the maximum coordination solution is adopted, there are 135 PSUs out of Chisinau urban, and 75 in Chisinau urban. If the constraint in Chisinau urban is relaxed, and if the number of enumerators in Chisinau urban is increased, then this solution can be considered feasible.

The overall approach followed for the design of the master sample is such that it can be easily generalised in other similar situations. In particular, based on this experience, one of the most burdensome task of the whole procedure, namely the preparation of the input datasets required by the R2BEAT optimisation functions for the two-stage sample design, has been completely automatised in a new release of this package. This new version (1.0.2) also includes the functions of the FS4 package, thus simplifying the overall workflow. A complete example, from the input preparation to the two-stage allocation, ending with the PSU selection, is contained in a dedicated vignette, "*Two-stage sampling design workflow*"[10].

10 https://urlsand.esvalabs.com/?u=https%3A%2F%2Fbarcaroli.github.io%2FR2BEAT%2Farticles%2FR2BEAT_workflow.html&e=3cfb7ead&h=098d22c7&f=y&p=y.

References

Barcaroli, G., M. Ballin, H. Odendaal, D. Pagliuca, E. Willighagen, and D. Zardetto. 2020. "SamplingStrata: Optimal Stratification of Sampling Frames for Multipurpose Sampling Surveys". In *The Comprehensive R Archive Network – CRAN*. Wien, Austria: Institute for Statistics and Mathematics of Wirtschafts Universität. https://cran.r-project.org/package=SamplingStrata.

Bethel, J. 1989. "Sample allocation in multivariate surveys". *Survey methodology*, Volume 15, N. 1: 47-57. Ottawa, Ontario, Canada: Statistics Canada.

Deville, J.-C., and C.-E. Särndal. 1992. "Calibration estimation in survey sampling". *Journal of the American Statistical Association*, Volume 87, N. 418: 376-382.

Fasulo, A., G. Barcaroli, R. Cianchetta, S. Falorsi, A. Guandalini, D. Pagliuca, and M.D. Terribili. 2020. "R2BEAT: Multistage sample design". In *The Comprehensive R Archive Network – CRAN*. Wien, Austria: Institute for Statistics and Mathematics of Wirtschafts Universität. https://cran.r-project.org/package=R2BEAT.

Kish, L. 1995. "Methods for design effects". *Journal of Official Statistics - JOS*, Volume 11, N. 1: 55-77. Stockholm, Sweden: Statistics Sweden.

Matei, A., and C.J. Skinner. 2009. "Optimal sample coordination using controlled selection". *Journal of Statistical Planning and Inference*, Volume 139, N. 9: 3112-3121.

Neyman, J. 1934. "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection". *Journal of the Royal Statistical Society*, Volume 97, N. 4: 558-625.

Tschuprov, A.A. 1923. "On the mathematical expectation of the moments of frequency distributions in the case of correlated observations". *Metron*, 2: 461-493; 646-683.

Särndal, C.-E. 2007. "The calibration approach in survey theory and practice". *Survey methodology*, Volume 33, N. 2: 99-119. Ottawa, Ontario, Canada: Statistics Canada.

Woodruff, R.S. 1971. “A Simple Method for Approximating the Variance of a Complicated Estimate”. *Journal of the American Statistical Association*, Volume 66, N. 334: 411-414.

Zardetto, D. 2015. “ReGenesees: an Advanced R System for Calibration, Estimation and Sampling Error Assessment in Complex Sample Surveys”. *Journal of Official Statistics - JOS*, Volume 31, N. 2: 177-203. Stockholm, Sweden: Statistics Sweden.

The *Rivista di statistica ufficiale* publishes peer-reviewed articles dealing with cross-cutting topics: the measurement and understanding of social, demographic, economic, territorial and environmental subjects; the development of information systems and indicators for decision support; the methodological, technological and institutional issues related to the production process of statistical information, relevant to achieve official statistics purposes.

The *Rivista di statistica ufficiale* aims at promoting synergies and exchanges between and among researchers, stakeholders, policy-makers and other users who refer to official and public statistics at different levels, in order to improve data quality and enhance trust.

The *Rivista di statistica ufficiale* was born in 1992 as a series of monographs titled "*Quaderni di Ricerca Istat*". In 1999 the series was entrusted to an external publisher, changed its name in "*Quaderni di Ricerca - Rivista di Statistica Ufficiale*" and started being published on a four-monthly basis. The current name was assumed from the Issue N. 1/2006, when the Italian National Institute of Statistics – Istat returned to be its publisher.

La Rivista di statistica ufficiale pubblica articoli, valutati da esperti, che trattano argomenti trasversali: la misurazione e la comprensione di temi sociali, demografici, economici, territoriali e ambientali; lo sviluppo di sistemi informativi e di indicatori per il supporto alle decisioni; le questioni metodologiche, tecnologiche e istituzionali relative al processo di produzione dell'informazione statistica, rilevanti per raggiungere gli obiettivi della statistica ufficiale.

La Rivista di statistica ufficiale promuove sinergie e scambi tra ricercatori, stakeholder, policy-maker e altri utenti che fanno riferimento alla statistica ufficiale e pubblica a diversi livelli, al fine di migliorare la qualità dei dati e aumentare la fiducia.

La Rivista di statistica ufficiale nasce nel 1992 come serie di monografie dal titolo "Quaderni di Ricerca Istat". Nel 1999 la collana viene affidata a un editore esterno, cambia nome in "Quaderni di Ricerca - Rivista di Statistica Ufficiale" e diventa quadrimestrale. Il nome attuale è stato scelto a partire dal numero 1/2006, quando l'Istituto Nazionale di Statistica - Istat è tornato a esserne l'editore.

ISSN 1828-1982 € 15,00