UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Editing

7-9 October 2024, Vienna

The editing and imputation process of the 2021 household and nuclei types reconstruction in Italy

Rosa Maria Lipsi, Anna Pezone (Istat – Italian National Institute of Statistics, Italy)

lipsi@istat.it, pezone@istat.it

I. Introduction

1. Six years ago, in October 2018, the Italian National Institute of Statistics (Istat) has launched the Permanent Population and Housing Census (PPHC), by using the integration of information available from local and central administrative sources that acquired from sample surveys carried out in rotation on all Italian municipalities. The administrative sources contributed to the creation of the Base Statistical Register of Individuals (RBI) in which the individuals, residing in the area, can be grouped, with respect to a registry origin code, into households or cohabitations according to their main characteristics.

2. The sample census surveys, List-based survey (L) and Distribution range survey (A), were performed by chosen respectively samples of households or samples of house numbers, extracted from RBI, in the municipalities selected for 2018 and 2019. For the national publication of the data from the first two surveys of the Permanent Census, therefore, macro-level counts were obtained using sampling weights and over- and under-coverage correctors applied to the RBI data.

3. In 2020, due to Covid-19, as it was not possible to conduct the field survey, Istat produced a count of the resident population in Italy by gender, age, citizenship and level of education, by using the Integrated Archive of Usual Resident Population (AIDA) (Istat, 2022), already tested, but until then not yet included in the census data production process. This was only possible thanks to a massive use of all information deriving from the various administrative sources available and from the registers, perfected for quality and timeliness. AIDA has identified two population subgroups that allowed the RBI to be corrected at 31st of December 2020:

1) people habitually residing in Italy with *direct signs of life* of at least one year in the administrative archives, if not present in the RBI as residents at 31st of December 2020, represent the under-coverage of the municipal registers on the same date;

2) people resident in RBI at 31st of December 2020 without *direct and indirect signs of life* in the administrative archives represent the over-coverage of the municipal registers.

RBI corrected with AIDA, for under and over coverage, determined RBI_CENS2020.

4. Immediately after the 2011 census, when Istat began to make the transition from a traditional "door-to-door" Census to the permanent Census, researchers began to think about the use of administrative sources in the permanent Census strategy. In fact, in 2011, despite the traditional census methodology (exhaustive and simultaneous field survey) two administrative sources have been used to guide the survey: the Municipal Registry Lists (LAC) and the Supplementary List from Auxiliary Sources. The first used for the postal sending of the questionnaires to the registered households in the registry office, and the second used to detect the habitually resident population not yet registered in the registry office.

5. Subsequently, each year, the LACs were acquired by the municipalities at 1st of January Year *t* as a reference date. This source contains detailed information on individuals residing in households or cohabitations and it is a valid basis for extracting the sampling units of social surveys and experiments of the permanent census, constituting the main information source for the creation of the RBI too.

6. At the same time, since 2013, the Ministry of the Interior (ANPR, 2024) has started a process of centralizing data on the resident population of each municipality to create the National Register of Resident Population (ANPR). Since 2018, simultaneously to the acquisition of LACs, the municipalities have begun to take over ANPR and Istat has begun to acquire data directly from the Ministry of the Interior through Sogei (Sogei, 2024). Sogei is an IT partner that oversaw the development of the portal and is responsible for sending annually data to Istat for the entire resident population at 1st January Year *t*. On 1st of January 2021, the 90.4 percent of the municipalities (7,147 on 7,903) had taken over, completing the takeover on 1st of January 2022.

7. In 2021, Istat was again able to carry out census surveys (List and Areal), which, integrated with RBI_CENS2021 (referring to 31st of December 2021), produced by using updated and more efficient methodologies than those applied to RBI_CENS2020, represented the information basis for the population count and for the production of census hypercubes, as required by the EU regulation 2017/712 (Eurostat, 2017).

8. One of the mandatory information to produce, at macro-micro level, is official statistics on households and their characteristics. The main problem to solve is the correct identification of household, which is a very complex aggregate to detect, validate and disseminate. The reconstruction of the household in its internal composition is possible through the correction of individual variables taking into account those of the other household members.

9. Our goal is to provide an overview of the whole process to produce statistics on households and their characteristics, focusing on the revision of the overall Editing and Imputation system, involving innovative generalized solution and specific adaptations of the "Families Procedure" (FP) for the reconstruction of the household and nuclei types, usually used for social surveys.

II. Data and methods

A. Data

10. The data used to reconstruct the household and nuclei types are those of RBI_CENS2021 described in the previous paragraph. The Italian population at 31st of December 2021, amounts to 58,678,795 individuals (included 149,059 under-covered people) in 26,206,246 private households distributed in 20 regions (Table 1).

Table 1: Distribution of the Italian Population and Households by regions at 31st of December 2021. Absolute and percentage values.

	Number of Individuals		**Number of Households**	
Regions	**A.V.**	**%**	**A.V.**	**%**
Piemonte	4,218,723	7.2	2,001,951	7.6
Valle d’Aosta	122,547	0.2	60,468	0.2
Lombardia	9,882,579	16.8	4,492,423	17.1
Trentino-Alto Adige/Südtirol	1,061,745	1.8	469,907	1.8
Veneto	4,812,583	8.2	2,109,478	8.0
Friuli Venezia Giulia	1,184,966	2.0	564,743	2.2
Liguria	1,495,874	2.5	760,931	2.9
Emilia Romagna	4,391,763	7.5	2,032,219	7.8
Toscana	3,642,200	6.2	1,662,574	6.3
Umbria	853,493	1.5	383,931	1.5
Marche	1,479,967	2.5	646,864	2.5
Lazio	5,672,202	9.7	2,630,892	10.0
Abruzzo	1,270,858	2.2	558,313	2.1
Molise	290,367	0.5	130,888	0.5
Campania	5,606,656	9.6	2,212,896	8.4
Puglia	3,910,701	6.7	1,635,899	6.2
Basilicata	538,773	0.9	237,160	0.9
Calabria	1,848,679	3.2	808,445	3.1
Sicilia	4,812,598	8.2	2,066,148	7.9
Sardegna	1,581,521	2.7	740,116	2.8
Total	**58,678,795**	**100**	**26,206,246**	**100**

Source: Our elaboration on Istat data

11. The almost total coverage of the ANPR stock data on RBI has allowed enriching the microdata with some high-quality variables present in the administrative source. In particular, the Register, corrected for over and under coverage by integrating the RBI information with the Integrated Archive of Usual Resident Population in Italy (AIDA), has inherited the following variables from ANPR referring to 1st of January 2022: relationship of kinship (30 categories defined by the Ministry of the Interior), marital status and date of marriage or civil union.

12. The resident population by age, sex and marital status (POSAS) at 31st of December Year *t* is another administrative source used as a benchmark for the comparison of the marital status variable between the RBI_CENS2021 data and those published by demographic statistics. One of the difficulties encountered was the lack of information for undercover individuals for whom only family code, gender, age and citizenship were available.

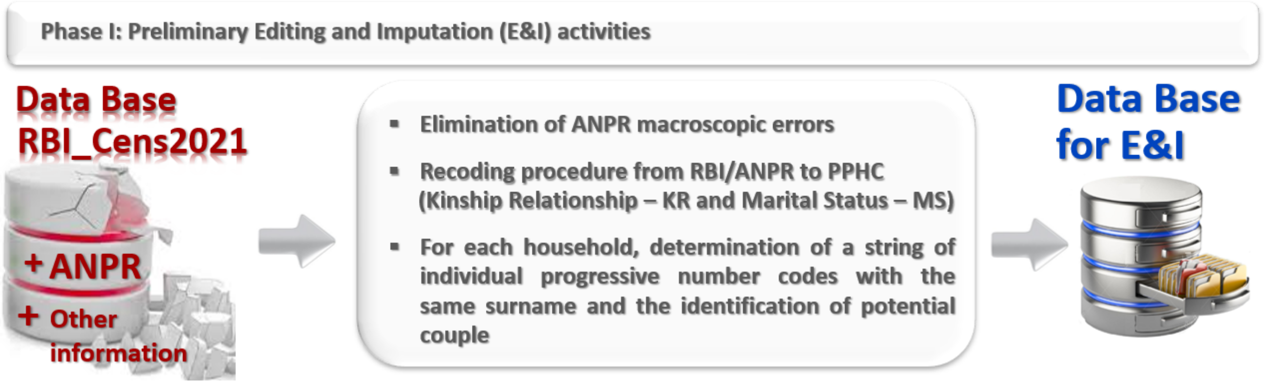
B. Methods

13. For the household and nuclei types reconstruction, the variables in RBI were enriched by those in the administrative source (ANPR) as individual code, household code, age, date of birth, sex, citizenship, relationship with the references person (RP), marital status, date of marriage or civil union and number of members. In addition to these variables, auxiliary ones were calculated, useful to the reconstruction process.

14. The kinship relationship and marital status of ANPR have been reclassified to match the classification used in the questionnaires of the permanent population census surveys.

15. Subsequently it was necessary to carry out some preliminary Editing and Imputation (E&I) activities to verify the validity and correctness of the individual variables (Figure 1).

Figure 1: Phase I of the E&I activities



Phase I: Preliminary Editing and Imputation (E&I) activities

- Elimination of ANPR macroscopic errors
- Recoding procedure from RBI/ANPR to PPHC (Kinship Relationship – KR and Marital Status – MS)
- For each household, determination of a string of individual progressive number codes with the same surname and the identification of potential couple

B.1. Phase I: Preliminary Editing and Imputation (E&I) activities

16. The relationship with the reference person was the subject of an experiment in the first months of 2022 to test the very complex and accurate reclassification methodology for the categories of the variables for which there was no exact correspondence.

17. In particular, taking into account the compatibility rules between components, the analysis of the observed data allowed to define the criteria, with a high degree of reliability, for reclassifying generic categories of ANPR into specific categories of the census questionnaire. For example, category 23 of the kinship relationship (cohabiting with adoption or emotional ties) in ANPR corresponded to two categories of the census questionnaire: 4 (cohabiting in consensual union with the reference person) and 23 (other cohabiting person without being a member of a couple, a relative, or extended family).

18. The reclassification of marital status did not cause particular problems, however the personal data had a lack of structural information due to the registry regulation especially for the foreign population; the registry offices, in fact, do not record the marital status when people cannot produce adequate documentation of the country of origin, therefore, all these individuals married abroad are registered in the registry office with "Other" or "Unknown" marital status (approximately 1.4 million individuals equal to 2.4% of the total

population and 30% of the total foreign population in Italy). To fill the lack of demographic information, it was necessary to proceed with the imputation of this variable, harmonizing it with the distribution of resident population by age, sex and marital status (POSAS).

B.2. Auxiliary variables description

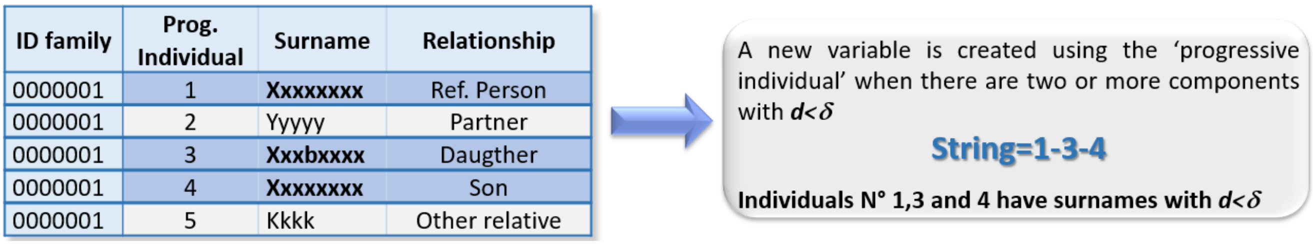
19. One of the auxiliary variables calculated was the string of progressive numbers of individuals with the same surname within each household, respecting the anonymization process by guarantee the privacy of the individual data, according to the General Data Protection Regulation (GDPR – Regulation 2016/679); this string was very useful for validating household members and correcting any anomalies.

20. The process to calculate the string was based on an internal algorithm distance that is a string metric for measuring the difference between two sequences [1]:

$$d\,(\text{Surname}_i\,,\ \text{Surname}_j) < \delta \quad \forall\ i,j=1,2,..,\ \text{n components [1]}$$

21. The distance function d measures the similarity between two strings. It is calculated by using an internal method based on N-gram algorithm and Jaro-Wikcler distance. This function d allows to understand if two strings are the same or sufficiently similar. Smaller distances correspond to more similar strings.
The $0 \leq \delta \leq 1$ (with 0=max similarity, 1=min similarity) value in [1] is the acceptability threshold, calculated taking into account observed data. A new variable is created using the 'progressive individual' when there are two or more components with $d<\delta$. For example, if we have a household with 5 components and individuals 1, 3 and 4 have surnames with $d<\delta$ according to the distance measure in [1] the new variable is string=1-3-4 (Figure 2). For each household the choice of the string depends on the greater number of components with the same surname; in case of two or more strings with the same number of components with the same surname the chosen string was that including the RP.

Figure 2: Example of the string comparison process of surname



ID family	Prog. Individual	Surname	Relationship
0000001	1	**Xxxxxxxx**	Ref. Person
0000001	2	Yyyyy	Partner
0000001	3	**Xxxbxxxx**	Daugther
0000001	4	**Xxxxxxxx**	Son
0000001	5	Kkkk	Other relative

A new variable is created using the 'progressive individual' when there are two or more components with ***d<δ***

String=1-3-4

Individuals N° 1,3 and 4 have surnames with *d<δ*

22. Another auxiliary variable, widely used in previous censuses, is the one that allows potential couples to be identified (Bianchi et al., 2020). The identification of potential couple (Figure 3), based on gender, age, relationship with RP, marital status and year of marriage or civil union of the two partners, allows:

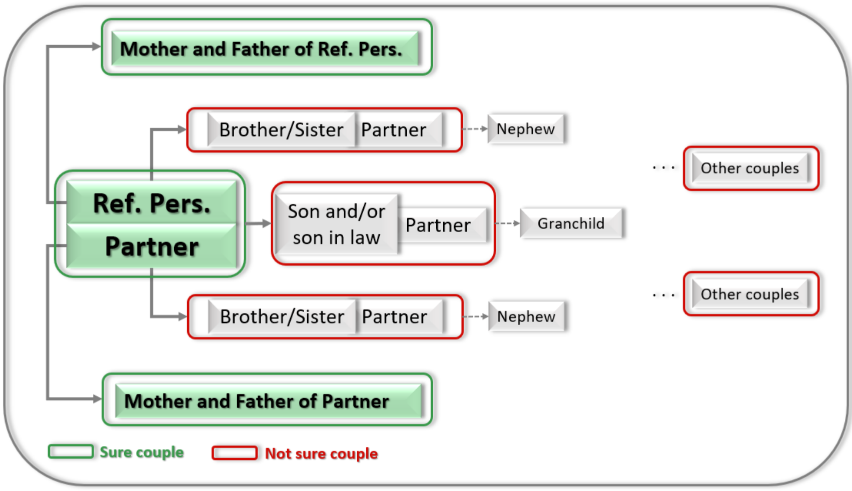
- identifying component of couple having unique or non-unique relationship with the RP;
- computing score using the responses provided to the demographic variables and the relationships of people that form couple within the household, by defining sure and not sure couples.

23. On the basis of the available information it is possible to identify only these types of sure couples (Figure 3), in the green rectangles, whose components are identified by means of appropriate control rules:

- Reference person with his/her partner
- Mother/father of reference person with his/her partner
- Mother/father in law of reference person with his/her partner

24. The others, in the red rectangles, are not sure couples (Figure 3) because there is a non-unique and well-defined relationship with the reference person. For this type of couple, it is difficult to solve the problem of checking and eventually correcting the components.

Figure 3: Identification of potential couples for household reconstruction



25. Further individual auxiliary variables were calculated because they were functional to the E&I process, such as duration of marriage or civil union, age at marriage or civil union, etc.

B.3. Phase II: Recursive Editing and Imputation process after the "Families Procedure"

26. The identification of households requires checking and correcting both the incompatibilities between the values of the variables relating to a single component (intra-component controls) and the incompatibilities between the values of the variables relating to the different components of the same households (controls inter-components). The process of rebuilding households is quite complex, especially if we consider the constant changes that occur within households following the demographic and social transformations observed over the years. Just think of the increase in the number of "reconstituted households" and the so-called "extended households" due to an increase in separations and divorces in couples, whose members often enter into a new union with people who are themselves separated or divorced and with children, rather than celibate or single. Such households are very difficult to deal with both from a statistical and computational point of view, given the complexity of the structure and the relationships existing between individuals.

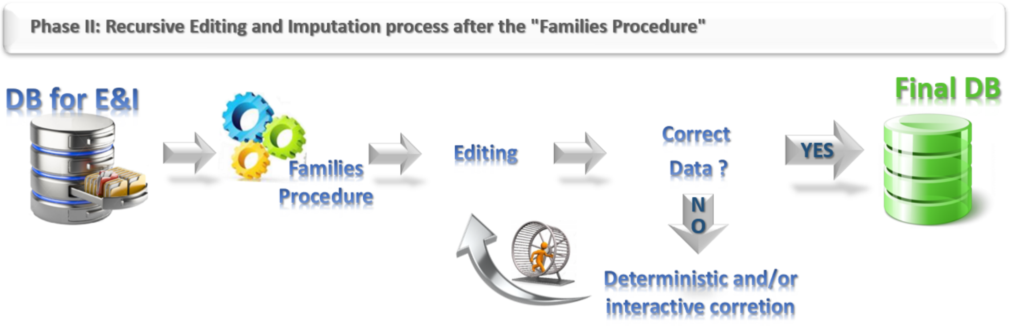
27. The reconstruction of a household takes place starting from the formation of the couples, on the basis of the individual information collected, then household and nuclei types and their respective positions are identified, both in the nucleus and in the household.

28. After having reclassified the variables *relationship with RP* and *marital status*, and having determined the value of the auxiliary variables, it was possible to proceed with the editing using the compatibility rules between individual data and family variables; subsequently incorrect or missing data were imputed to restore consistency between the variables. Only at the end of these activities the "Families Procedure" (FP) was carried out, which checked and corrected some variables, subsequently calculating the household and nuclei types.

29. At the end of the FP, familial editing was carried out to verify the coherence between the household members with respect to age, sex, marital status, kinship relationship and duration of marriage or civil union in order to identify any anomalous household. This process allowed the first level validation to be carried out before the release of the data for the 2nd level validation by the thematic experts.

30. The process described is cyclical and reiterated until the optimal result was achieved (Figure 4). The application is a complex procedure because the editing and imputation process do not end in a single "step" but require a reiteration on the data, to restrict the errors in increasingly smaller subsets until they have zero numbers.

Figure 4: Phase II of the E&I and household reconstruction process


Phase II: Recursive Editing and Imputation process after the "Families Procedure"
DB for E&I
Families Procedure
Editing
Correct Data ?
YES
Final DB
NO
Deterministic and/or interactive corretion

31. The FP is a software package (Budano et al., 2010) used by the social surveys for the reconstruction of the household type and can be adapted for the specific needs of the survey conducted. This procedure defines a set of IT and editing steps for the correction of individual variables such as, for example, the kinship relationship, the marital status, etc. in relation to those of the other members of the household.

32. It is worth underlining that the FP was adapted, for the first time, to a huge amount of data as RBI_CENS2021 which contains about 58,7 million of resident population in 26,2 million of households. Usually the FP is used for social surveys (maximum 129,000 individuals in 58,000 households as in Labour Force Survey) and recently for the permanent population census of 2018, 2019 and 2021 (maximum 4,8 million individuals in 2,4 households as in 2021) too. The huge amount of data in RBI_CENS2021 required a well-defined strategy to adapt the FP. In order to ameliorate the performance and reduce the time to execute the FP it was necessary to subset households by the number of components and by grouping some provinces.

33. At the end of the correction phase of the kinship relationship variable and, consequently of all the variables connected to it (marital status, year of marriage, etc.), the FP, with a special function, allows reconstructing the household in its internal composition by creating household and nuclei types.

C. The main results of the E&I process

34. The missing data affected mainly the year of marriage or civil union (58%) with decreasing values for marital status (37.4%) and relationship with RP (4.6%). The trends are similar for the mentioned variables in each region. Referring to marital status, regions with the higher number of missing values (greater than 10%) are Lombardia, Emilia Romagna, Veneto and Lazio. For the year of marriage or civil union, there are two regions (Lombardia and Lazio) with missing values greater than 10%. Lastly, for the relationship with RP in addition to Lombardia and Lazio, there is Campania with missing values greater than 10% (Table 2).

35. After the imputation process, editing was launched, using 94 edit rules (13 individual and 81 familial edits) to identify any inconsistencies. The number of failed edits, with at least one incorrect individual error, involved 5,497,417 people (9.4%) of the total population. Lazio, Campania and Liguria, considering their respectively resident population, are the regions with the higher percentages (more than 11%) of individuals with at least a correction (Figure 5).

36. Individual editing identified 937,328 failed edits (1.6%) out of the total units. Most of the errors concerned marital status (86.5%). The inconsistencies found between two or more variables were mainly between relationship with RP and marital status (51.2%) (Lipsi and Pezone, 2024).

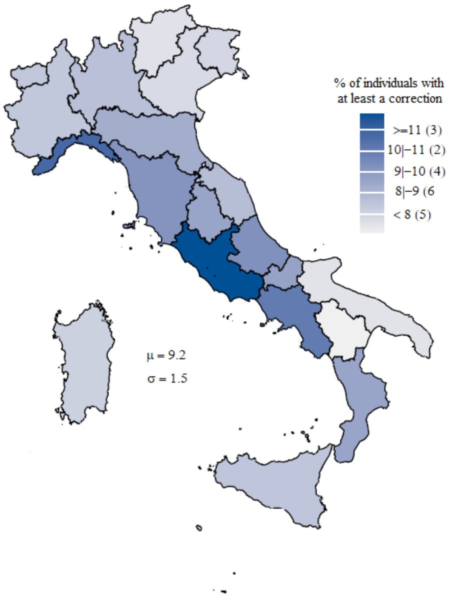
37. The number of errors of marital status (Table 3) were similar for women (50.88%) and men (49.12%); the 30-59 age group is the most affected by errors (62.44%). Finally, with respect to citizenship, the highest number of errors were observed among Italians (80.22%), especially women (40.62%).

Table 2: Distribution of missing data for marital status, year of marriage or civil union and relationship with reference person (RP) by region. Percentage values.

Regions	Marital status	Year of marriage or civil union	Relationship with RP
Piemonte	6.38	7.03	4.75
Valle d'Aosta	0.22	0.23	0.06
Lombardia	25.45	23.19	16.98
Trentino-Alto Adige/Südtirol	2.43	1.84	1.80
Veneto	10.01	8.34	8.54
Friuli Venezia Giulia	2.81	1.94	2.09
Liguria	4.08	3.85	2.16
Emilia Romagna	11.09	9.41	7.47
Toscana	9.52	8.59	9.76
Umbria	1.49	1.39	0.94
Marche	3.26	2.62	2.20
Lazio	10.46	11.15	15.07
Abruzzo	1.40	1.86	1.53
Molise	0.15	0.28	0.97
Campania	3.56	7.30	10.78
Puglia	2.74	3.15	4.52
Basilicata	0.33	0.36	0.45
Calabria	0.91	1.61	3.07
Sicilia	2.44	4.29	5.97
Sardegna	1.25	1.57	0.88
Total	**100**	**100**	**100**

Source: Our elaboration on Istat data

Figure 5: Percentage of individuals with at least a correction by region (in brackets the number of regions).



Source: Our elaboration on Istat data

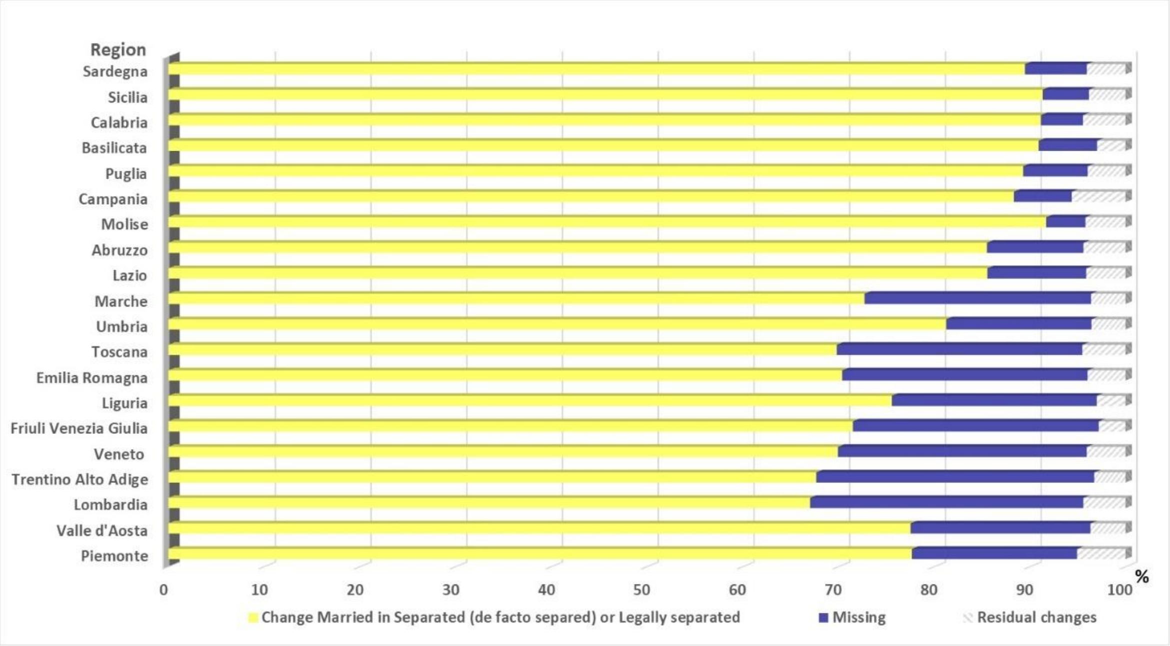
Table 3: Distribution of the errors of marital status, by age groups, gender and citizenship (Italian (It) and Foreign (For)). Percentage values.

	Women			Men			
Age groups	**It**	**For**	**TotW**	**It**	**For**	**TotM**	**Total**
0-16	0.11	0.13	0.24	0.13	0.14	0.27	0.51
17-29	0.61	1.61	2.22	0.35	2.04	2.39	4.61
30-59	26.18	6.56	32.74	23.41	6.29	29.71	62.44
60-84	12.93	1.89	14.83	14.93	1.02	15.95	30.78
85 and over	0.79	0.07	0.85	0.78	0.03	0.81	1.66
Total	**40.62**	**10.26**	**50.88**	**39.59**	**9.53**	**49.12**	**100**

Source: Our elaboration on Istat data

38. The analysis of the *marital status* before and after the E&I process highlights that the changes observed among categories are those mainly due to married people and de facto or legally separated people, categories present in the census, but not in ANPR. These changes especially involved smaller regions (Figure 6). Excluding missing data, the residual changes for the other categories of the marital status are less than 6%.

Figure 6: Distribution of the marital status (married) before/after the E&I process. Bars are % of each category.



Source: Our elaboration on Istat data

39. Referring to the *year of marriage or civil union*, there were more imputations of missing data (66.84%) and few inconsistencies with the year of marriage (or union of the partner) or with the year of birth.

40. The corrections of the *relationship with RP* were more complex. The number of errors of this variable (Table 4) were higher for women (54.83%) than men (45.17%); the 30-49 age group is the most affected by errors (39.79%). With respect to citizenship, as marital status, the highest number of errors were observed among Italians (71.23%), especially women (39.26%).

Table 4: Distribution of the errors of relationship with RP, by age groups, gender and citizenship (Italian (It) and Foreign (For)). Percentage values.

	Women			**Men**			
Age groups	**It**	**For**	**TotW**	**It**	**For**	**TotM**	**Total**
0-14	4.30	1.44	5.73	4.52	1.50	6.02	11.76
15-29	4.94	2.78	7.73	4.58	2.39	6.97	14.70
30-49	14.19	8.16	22.35	10.49	6.96	17.44	39.79
50-64	9.68	2.50	12.18	8.16	1.94	10.11	22.29
65-84	4.91	0.65	5.56	3.79	0.39	4.18	9.74
85 and over	1.24	0.03	1.27	0.42	0.02	0.44	1.71
Total	**39.26**	**15.57**	**54.83**	**31.97**	**13.20**	**45.17**	**100**

Source: Our elaboration on Istat data

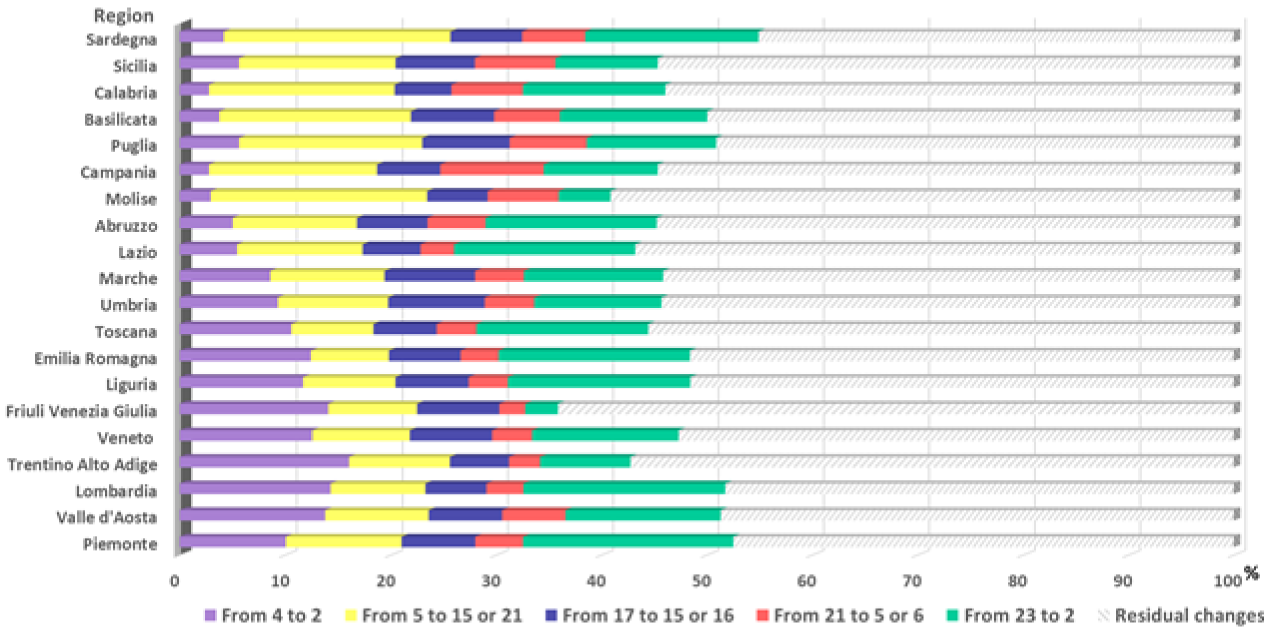
41. For foreigners, the low number of changes could be due to a high number of individuals classified, especially in large families, with kinship relationship 23 (other cohabiting person without being a member of a couple, a relative, or extended family); this occurs when:

- the registry officer is not be able to correctly identify kinship relationships for language problems or in absence of certain certification or cannot do otherwise,
- foreigners live together in the same household without having any kinship relationship for economic opportunities or otherwise.

This ensures that both editing and FP do not detect anomalies and therefore there are no corrections to be made.

42. The distribution of the *relationship with RP* before and after the E&I process with the higher percentage of changes, mainly involved 9 categories of the variable on 23. The changes for the remaining categories were negligible, so they were grouped together. The main changes were different among regions (Figure 7). For the most of the southern regions, the major changes referred to the “son/daughter of RP and of the spouse/civil partner/partner” that changed into “brother/sister of RP” or “grandson/granddaughter of RP and/or of the spouse/civil partner/partner of RP” (yellow bar of Figure 7). For the most of the northern regions, the major changes referred to the “partner of RP (consensual union)” that changed into “spouse of RP” (violet bar of Figure 7). There is not a regional profile for the changes observed from “other cohabiting person without being a member of a couple, a relative, or extended family” to “spouse of RP” (green bar of Figure 7).

Figure 7: Distribution of some categories of the relationship with RP before/after the E&I. Bars are % of each category.



2 = Spouse of RP
4 = Partner of RP (consensual union)
5 = Son/daughter of RP and of the spouse/civil partner/partner
6 = Son/daughter of RP only
15 = Brother/sister of RP
16 = Brother/sister of the spouse/civil partner/partner of RP
17 = Spouse of the brother/sister of RP or of the spouse/civil partner/partner of RP
21 = Grandson/granddaughter of RP and/or of the spouse/civil partner/partner of RP
23 = Other cohabiting person without being a member of a couple, a relative, or extended family

Source: Our elaboration on Istat data

D. Final remarks

43. In this paper the whole process to produce statistics on households and their characteristics has been briefly described, focusing on the revision of the overall Editing and Imputation system, involving innovative generalized solution and specific adaptations of the "Families Procedure" (for the reconstruction of the household and nuclei types) to a huge amount of data, usually used for social surveys, highlighting the complexity linked both to the integrated use of data gathered from registers, administrative sources and surveys.

44. It is important to underline that FP was used for the first time on integrated data, without never having tested it on big dataset, relating to individuals and households belonging to the all resident Italian population. In addition, this process improved the quality of data released to Eurostat with reference to census hypercubes involving household and nuclei types.

45. However, further studies, both on sources and methods of E&I, will be useful to reduce missing data and errors as much as possible. It will be interesting to apply Machine Learning methods or Artificial Intelligence to improve the E&I process in order to minimize errors in household reconstruction, especially for households with numerous members which internal composition is difficult to detect.

46. Another hope would be to reengineer the FP aiming to optimize the performance by reducing some anomalous household and the speed of its execution.

Acknowledgment

The authors wish to thank Silvia Dardanelli and Salvatore Pagliara for their suggestions and contribution.
The work is the result pursued jointly by the two Authors. Opinions expressed in this article are those of the Authors but do not involve the institutions they belong to.

References

ANPR (2024). *Anagrafe Nazionale Popolazione Residente*. https://www.anagrafenazionale.interno.it

Bianchi G, Filippini R, Lipsi RM, Pezone A, Scalfati F. (2020). *An overview of the editing and imputation process of the 2018 Italian Permanent census*. UNECE, online workshop on Statistical Data Editing.

Budano G. and P. Piergentili (2010), La Procedura Famiglie in G. Budano e S. Demofonti, La misurazione delle tipologie familiari nelle indagini di popolazione in *Metodi e Norme*, 2010, n. 46. Istat.

Eurostat, (2017). European Commission. Commission Regulation No 763/2008 of the European Parliament and of the Council, OJ L 105, 21.4.2017, p. 1–11.

GDPR (2016). *General Data Protection Regulation* (GDPR – Regulation 2016/679).

Istat (2022). Nota tecnica sulla produzione dei dati del Censimento Permanente: *la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2021*. pp.14.

Lipsi R.M. and A. Pezone (2024), An innovative approach to improve the quality of the household and nuclei types reconstruction in Italy, *Q2024*, Estoril, 04-07 June 2024, pp. 10.

Sogei (2024), Sogei - Società Generale d'Informatica S.p.A., società di Information Technology, https://www.sogei.it/it/sogei-homepage.html