VOLUME LXXVIII – N. 2 APRILE – GIUGNO 2024

RIVISTA ITALIANA DI ECONOMIA DEMOGRAFIA E STATISTICA

DIRETTORE
CHIARA GIGLIARANO

GUEST EDITOR
LIVIA CELARDO, DANIELE GRECHI, ENRICO MORETTO

ASSOCIATE EDITOR
MARIATERESA CIOMMI, ALESSIO GUANDALINI, LUCA SALVATI

COMITATO SCIENTIFICO
GIORGIO ALLEVA, EMANUELE BALDACCI, GIAN CARLO BLANGIARDO, CLAUDIO CECCARELLI, FRANCESCO M. CHELLI, CONCHITA D'AMBROSIO, CASILDA LASSO DE LA VEGA, MIKHAIL DENISENKO, LUIGI DI COMITE, PIERPAOLO D'URSO, ALESSIO FUSCO, MAURO GALLEGATI, ANTONIO GIMENEZ MORENO, RARES HALBAC COTOARA ZAMFIR, ALBERTO QUADRIO CURZIO, CLAUDIO QUINTANO, JESUS RODRIGO COMINO, KOSTAS RONTOS, SILVANA SCHIFINI D'ANDREA, SALVATORE STROZZA, PHILIPPE VAN KERM, PAOLO VENERI, PAOLO VERME, ROBERTO ZELLI

REDAZIONE
OHIANA ARISTONDO, ALESSIO BUONOMO, LIVIA CELARDO, LIDIA CERIANI, ANDREA CUTILLO, GIUSEPPE GABRIELLI, DANIELE GRECHI, FRANCESCA MARIANI, ENRICO MORETTO, SIMONA PACE, FAUSTO PACICCO, GLORIA POLINESI, CECILIA REYNAUD, STEFANIA RIMOLDI, GIUSEPPE RICCIARDO LAMONICA, ANDREA SPIZZICHINO, ANDREA VENEGONI

SIEDS
SOCIETÀ ITALIANA
DI ECONOMIA DEMOGRAFIA E STATISTICA

CONSIGLIO DIRETTIVO

Presidenti Onorari: LUIGI DI COMITE, FRANCESCO MARIA CHELLI

Presidente: SALVATORE STROZZA

Vice Presidenti: LEONARDO BECCHETTI, CHIARA GIGLIARANO, VENERA TOMASELLI

Segretario Generale: ALESSIO GUANDALINI

Consiglieri MARINA ALBANESE, MARCO ALFÒ, GIUSEPPE GABRIELLI, MARGHERITA GEROLIMETTO, MATTEO MAZZIOTTA, SIMONE POLI, MARIA CRISTINA RECCHIONI, LAURA TERZERA

Segretario Amministrativo: FABIO FIORINI

Revisori dei conti: MICHELE CAMISASCA, GIUSEPPE NOTARSTEFANO, DOMENICO SUMMO

Revisori dei conti supplenti: CLAUDIO CECCARELLI, CECILIA REYNAUD

SEDE LEGALE:
C/O Studio Associato Cadoni, Via Ravenna n. 34 – 00161 ROMA

info@sieds.it

rivista@sieds.it

VOLUME FUORI COMMERCIO – DISTRIBUITO GRATUITAMENTE AI SOCI

INDICE

BUSINESS INTELLIGENCE FOR ANALYSING ISTAT HUMAN RESOURCES DATA BASE

Alessandra Dentini, Iole Zeppieri[1]

Abstract. In the recent years, the use of administrative data plays a central role in Official Statistics for reducing statistical burden. The internal administrative data could represent a strong tool to increase efficiency of management system. Using Business Intelligence techniques, the Italian Institute of Statistics (Istat) Duty Travel database is analysed in order to realise strategic information for stipulating agreements with transport suppliers. This innovative approach in Public Administration allows exploiting data that have a strong benefit in order to allow substantial financial savings to be allocated to scientific research.

1. Introduction.

The Italian Institute of Statistics (Istat) provides the official information for understanding and deciding the complex reality of Italy. Istat carries out surveys (exhaustive and sample) and collects data from different official sources. In this paper, an administrative database for internal use is treated to show the power of the Business Intelligence for increasing the efficiency of the management system. It is not just a tool for understanding the "status quo" of Istat Human Resources (HR) but it is a set of statistical analysis to construct a strategy for improving the "idea" of modern Public Administration.

The Business Intelligence (BI) is a technology-driven process for analyzing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process, organizations collect data from internal IT systems and external sources, prepare it for analysis, run queries against the data and create data visualizations, BI dashboards and reports to make the analytics results available to business users for operational decision-making and strategic planning.

[1] The paper is the results of the common work of the authors. In particular, A. Dentini has written Sections 2, 4 and 5; I. Zeppieri has written Sections 1 and 3.

The duty travel information selected from administrative Istat database are made up of a several variables which include employee identification data (registration number, profile, staffing plan), duty travel administrative data (number of assignments, duration, destination, object, and institution visited, cost center) and finally, the detail of the cost items for single duty travel expenditure (type of transport, board and lodging). The data analysis covers the years 2009-2022 (thirteen years of time series!) even if the last years have not presented many duty travels due to Covid pandemic.

The first phase towards the construction of the database involved the transposition of the raw matrix to obtain the single duty travel for each record (row). The statistical analyzes used have extracted from a very large matrix of unstructured administrative data the information necessary to design strategic company policies aimed at saving and investing towards new strategies for enhancing professional skills. In fact, the ability to reduce the costs of missions through the signing of agreements with service providers is a goal that Istat will pursue in the coming years.

The paper is one of the first example of Business Intelligence (BI) in Istat, in which data purely used for administrative matters are used to obtain useful information for the management of activities, economic savings policies, relations with other institutional organizations and much more. We are talking about a wealth of information to be exploited to increase the effectiveness of some internal and external processes of the Istat. The paper is structured as follows. In the second section, the approach of BI in the Public Administration is presented; in the third section, the database construction is focused; in the fourth section the results of BI techniques are presented, and, in the last section, the concluding remarks are highlighted.

2. Business Intelligence in the Public Administration.

In addition, in the light of the regulatory reforms, the Public Administration is today engaged in a renewal process that focuses on the internal efficiency of each individual administration, on greater transparency and on more accessible, flexible and timely services, to facilitate the relationship between the public and citizen.

These levers drive the Public Administration to equip itself with tools for decision support, with regard to Business Intelligence as a solution capable of combining planning and strategic control, as well as the analysis of processes and the detection of recurring phenomena, with all the implications relating to their territorial dimension. The Business Intelligence (BI) is a technology-driven process for analysing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process,

institutions collect data from internal Information Technology systems (IT) and from external sources, prepare it for analysis, run queries against the data and create data visualizations, BI dashboards and reports to make the analytics results available to business users for operational decision-making and strategic planning (De Vivo, Polzonetti, Tapanelli, 2011).

In recent years, business intelligence has developed to include more processes and activities to enable performance improvement. Such processes include a statistical approach.

• Data Mining: Use databases, statistics, and machine learning to uncover trends in large datasets.

• Reporting: Sharing data analytics with stakeholders so they can draw conclusions and make decisions.

• Performance metrics and benchmarking: Compare current performance data with historical data to monitor performance against goals. Typically, this is done using custom dashboards.

• Descriptive analytics: Using preliminary data analytics to understand what happened.

• Query Execution: Querying data with specific questions, for which BI extracts answers from datasets.

• Statistical analysis: starting from the results of the descriptive analysis, further exploration of the data using statistics, for example in relation to how and why a certain trend has occurred.

• Data visualization: Transform data analysis into visual representations, such as graphs, charts and histograms, for easier data consumption.

• Visual analytics: Exploration of data through visual representations to communicate information on the fly and follow the flow of analysis.

• Data Preparation: Compiling various data sources, identifying their dimensions and measurements, and preparing them for data analysis.

The Corporate Information System (URBI) collects, organizes, processes, and manages the data necessary for running Istat. Such data can be produced directly within Istat in the various business processes or be acquired because of relationships with external parties.

Similarly, such data can be intended for internal consumption or be intended for external purposes. URBI consists of a computerized part called Business Information System (BIS) and a non-automated part. The components (both computerized and not) of a Corporate Information System (URBI) can be divided into two categories according to their purpose:

1) components for the support of operational activity. This part of the URBI is responsible for archiving, managing and processing all the information for carrying out daily activities.

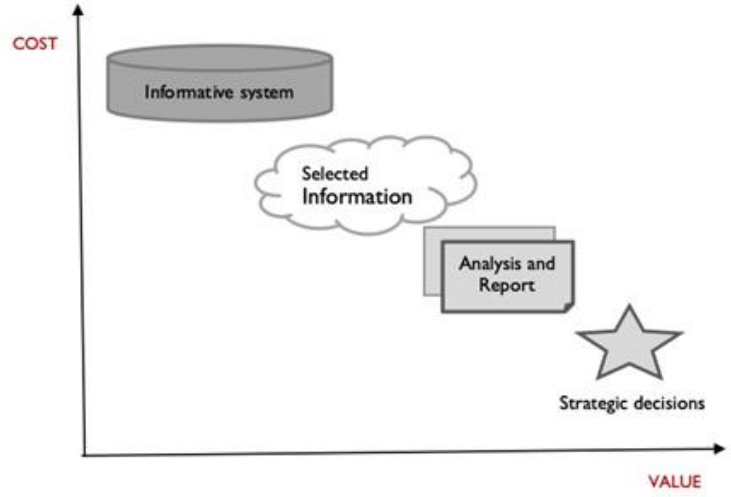
2) decision support components. This part of the URBI concerns management, production and archiving to support managers in strategic choices.

Internally, URBI presents a series of data that are redundant and non-coherent, consequently they are not supportive for decision-making choices.

In this case, the "data" is different from the "information". In fact, the terms “data” and “information” are sometimes used interchangeably, but they are not the same. In general, data is defined as individual facts (without meaning), while information is the organization and interpretation of those facts: information puts those facts into context. Data is unorganized, while information is structured or organized. Information is an uncountable noun, while data is a mass noun. Data is not typically useful on its own, but information is. By carrying out operations of aggregation and interpretation of the data present in the Human Resources Database, we transform them into information useful for analysis and for the choice of strategic decisions.

Figure 1 represents one of the main objectives of BI, i.e., the enhancement of information starting from unstructured data systems. This enhancement translates, through a work process, into a sharp reduction in costs in management terms. We want to extract information from the vast amount of operational data present in corporate IT systems using two techniques: data warehouse and data mining. This does not involve entering new data into the corporate IT system but is responsible for reorganizing, integrating, filtering, aggregating data already available to obtain strategic information (or summary data). In detail, in our case study, the information system (URBI) contains a lot of information in an unstructured and inconsistent way so that the information value is very low, and the cost reduction objective is far away. By selecting a series of variables so that indicators can be constructed, it is possible to develop many statistical analyzes and, therefore, to produce reports useful for identifying intervention policies on Istat's organizational management. The last step involves the exploitation of these policies in order to undertake strategic decisions with a high information content for the Institute with a view to reducing costs and investing the savings obtained in other research activities. Therefore, thanks to BI, as the value of information increases, the cost of processes decreases.

Figure 1 – *The same data but…with more value*



3. Data source and database construction

The Human Resources sector of Istat has been using URBI Smart as a management operating system since 2016. The application system manages various areas of human resources, including the legal and accounting management of employee duty travel and external staff who carry out activities for the Institute. In order to build a database that is statistically informative for the study of duty travel in Istat, it is necessary to review the work process divided into phases, starting from the rationalization of the information present in URBI in order to select the variables to be analysed for statistical purposes. The goal is to transform an administrative database, made up of multiple management variables, mainly used as a repository of administrative data, into a statistical database, capable of making the data classifiable and usable for statistical purposes. The duty travel information selected from URBI are made up of a several variables which include employee identification data (registration number, profile, staffing plan), duty travel administrative data (number of assignments, duration, destination, object and institution visited, cost centre) and finally, the detail of the cost items for single duty travel expenditure (type of transport, board and lodging). The data analysis covered the period 2009-2019 (eleven years) since this last year essentially represented the end of the duty travel due to the health emergency. The first phase towards the construction of the database involved the transposition of the raw matrix in order to obtain the single duty travel for each record (row) (Dentini, Mazziotta, Zeppieri, 2022).

Subsequently, the matrix underwent some changes summarized as follows:

• normalization of the cost items of duty travel for the years 2009 and 2015 when the management information system was different (SIGED was before URBI);

• normalization of the “Destination” variable in foreign countries and Italy, with a further subdivision by region, province and municipality;

• normalization of the “Destination” variable since the names of the destinations themselves were registered in a non-univocal way;

• normalization of the "Object of the duty travel" variable into four categories:
  - scientific meeting;
  - training;
  - internal meeting;
  - other institutional activities.

• classification by items of expenditure (travel, food and accommodation).

Table 1 – *Departures from Rome.*

Arrivals	Departures from Rome	…in percentage
Ancona	116	3,3
Bari	117	3,3
Bologna	474	13,5
Cagliari	109	3,1
Campobasso	63	1,8
Catanzaro	41	1,2
Firenze	415	11,9
Genova	104	3,0
Mestre	6	0,2
Milano	812	23,2
Napoli	473	13,5
Palermo	249	7,1
Perugia	98	2,8
Pescara	10	0,3
Potenza	6	0,2
Torino	159	4,5
Trieste	71	2,0
Cagliari	109	3,1
Venezia	179	5,1
Total	3.502	100,0

Our elaborations on Duty travel data base.

4. Results

In order to study the characteristics of duty travel, it is necessary to select from the database created a subset that focuses attention on the most important cases. In this regard, only the regional capitals are considered. Therefore, a double-entry table (not presented in the paper for reasons of size, 19x19) provides all duty travel to and from the cities hosting Istat regional offices. Just over twenty thousand routes in eleven years are considered. Of these, about eleven thousand routes insist on Rome, which, obviously, is the main headquarters of the National Institute of Statistics.

In table 1, the main destinations from Rome are presented; over time, 23.2% of the routes are towards Milan, 13.5% towards Naples and Bologna, 11.9% towards Florence. Palermo (7.1%), Venice- Mestre (5.3%) and Tourin (4.5%) present lower values than the cities already mentioned.

Table 2 – *Arrivals to Rome*.

Departures from…	Arrivals to Rome	…in percentage
Ancona	346	6,2
Bari	599	10,8
Bologna	330	5,9
Cagliari	319	5,7
Campobasso	156	2,8
Catanzaro	238	4,3
Firenze	477	8,6
Genova	275	4,9
Mestre	248	4,5
Milano	309	5,6
Napoli	487	8,8
Palermo	352	6,3
Perugia	252	4,5
Pescara	357	6,4
Potenza	263	4,7
Torino	249	4,5
Trieste	246	4,4
Venezia	57	1,0
Total	5.560	100,0

Our elaborations on Duty travel data base.

In Table 2, the main departure cities for Rome are presented; in eleven years, if we exclude Bari (almost 11%), Naples and Florence (almost 9%) which show outliers, all the other cities (including Milan) have a constant distribution of trips to Rome (about 5%). In fact, even the values of Venice and Mestre, if added together,

reach approximately this value. It seems interesting to study the situation of the route from Milan to Rome; in fact, the percentage (equal to 5.6%) is lower than other smaller cities that have smaller territorial offices. Probably because the Istat office in Milan (representing Lombardy) has such a high importance that it is more attractive than others. Moreover, for this reason, it organizes many events and has a higher number of incoming rather than outgoing duty travel.

Table 3 – *The cities most reached by Rome over time.*

Departures from Rome	Bologna	Florence	Milan	Naples
2009	53	42	89	30
2010	48	41	56	20
2011	80	37	47	18
2012	55	24	52	62
2013	66	58	108	41
2014	40	30	71	59
2015	36	37	101	41
2016	23	32	57	75
2017	36	50	51	30
2018	19	25	66	48
2019	18	39	114	49
Total	474	415	812	473

Our elaborations on Duty travel data base.

Table 3 shows the major destinations reached by Istat workers whose place of work is Rome; these destinations are easily reachable by train, which is the most used compared to the plane, which is practically unused. The use of the train has also been encouraged by the increase in High-Speed Train (HST) over time. The frequency over the years towards these destinations appears to be constant.
There are peaks for Milan in 2013 and 2019; this phenomenon is due to two editions of Census of Companies that sees Milan as an operational hub.

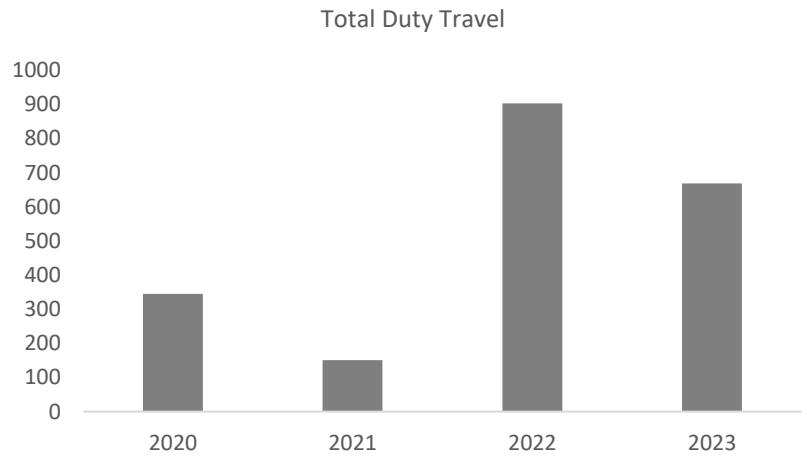
The Table 4 shows that the most used vector is the train. With approximately 23,000 duty travel in Italy from 2009 to 2019, there is a constant trend between 64.2% and 78.2%, except for 2011 where the use of the train drops to 43, 5% and increases the use of plane (19.6%). This phenomenon can be connected to the Population Census was carried out for the last time in the traditional way. The greater use of the plane is also due to more destinations in the same duty travel.

Table 4 – *Trasportations in percentage*

Trasportations in percentage	Plaine	Train	Boath
2009	13,9	64,2	0,0
2010	12,6	69,5	0,0
2011	19,6	43,5	0,0
2012	8,1	64,8	0,1
2013	9,7	78,2	0,0
2014	13,2	74,8	0,2
2015	12,6	68,1	0,1
2016	10,6	76,3	0,1
2017	16,1	68,2	0,2
2018	6,6	62,1	0,1
2019	5,9	64,5	0,0

Our elaborations on Duty travel data base.

Figure 2 – *Of duty travel during and after Covid.*



Our elaborations on Duty travel data base.

The SARS-CoV-2 pandemic emergency of 2020 prompted the Public Administrations to adopt gradual and timely managerial, organizational and technical choices for the protection of the health and safety of workers from this new risk, also limiting duty travel (Camisasca, Pietrantonio E., Magro, Arborea, Fabiani, Giordano, Pietrantonio M., Sparano, 2022). Before the pandemic, Istat researchers carried out around 3,000 duty travel in a year. The figure 2 represents duty travel

made in the pandemic and post-pandemic period. Duty travel made in 2020 mainly concern the pre-pandemic period (January and February 2020); a substantial decrease can be seen in 2021. Only at the end of 2021 does a weak recovery of duty travel begin, which increases more in 2022. In 2023, the data are encouraging, there is a clear recovery in travel with a projection for the end of the year of exceeding 2,000 duty travel.

5. Conclusions

The Italian and global Big Data and Analytics market is constantly evolving, with a high growth rate for both large companies and SMEs. In an economic phase dominated by an abundance of data and limited time to make decisions, it is important to exploit information to be competitive and generate business actions. Discover the advantages of Business Intelligence and Analytics to increase the productivity of your institution and simplify strategic choices.

Istat is starting a process of managerial growth in order to undertake activities for the coordination and use of strategies in order to provide technical-organizational support to the government structures of the Institute of Statistics.

This research experience must be understood as inserted within the context of profound innovation that statistics (not only official) are going through from the point of view of the use of administrative sources in order to represent complex realities with increasingly clear images that can assist the stakeholders in strategic choices.

The exploitation of the Istat duty travel database is a classic example of the use of administrative data for the development of BI techniques. In fact, BI refers to the ability to make better decisions, take informed action, and implement more efficient business processes. BI capabilities allow to: accept updated data from your organization; present data in easy-to-understand formats; provide data in a timely manner to make strategic decisions.

The objective of this contribution is to use internal Istat administrative sources for exploiting statistical information to rationalize internal processes and carry out economies of scale for improving the organizational structure of the Institute.

Besides, the final aim is the possibility of stipulating agreements with transport suppliers and with hotel structures for reducing the costs of duty travel and to invest the savings obtained in other research activities. The results of the statistical analyzes demonstrate that the number of duty travel has been constant over time, therefore allowing forecasts to be made on the future amount. Furthermore, the trend of the last year is strongly increasing therefore the prospect of returning to pre-pandemic levels has a very high probability. The analyzes agree in designing a scenario in

which the agreements to be stipulated can create considerable advantages in the organizational and financial management of Istat.

The first steps have been taken and have shown that statistical analyzes on databases for internal use can have a yet unknown value. The next steps guide the authors towards:

- interacting with Istat Mobility Manager;
- agreements with hotel facilities;
- analyzing the duty travel of external people and foreign destinations;
- implementing a data base available for consultation in real time;
- updating statistical reports to address decision strategies.

Official statistics, not only in Italy, are in a moment of great change. Administrative sources integrated with sample surveys are a new resource for providing quality statistics with less statistical burden for citizens and businesses.

Istat is acting in a new direction, exploiting internal administrative sources to increase process efficiency and financial savings.

Knowledge of BI statistical techniques will allow Istat to combine methodologies and unstructured data towards a new and modern working paradigm.

References

DE VIVO M.C., POLZONETTI P., TAPANELLI P. 2011. Open Data, Business Intelligence e Governance nella Pubblica Amministrazione, "Informatica e diritto", XXXVII Annata, Vol. XX, 2011, n. 1-2, pp. 239-262.

DENTINI A., MAZZIOTTA M., ZEPPIERI I. 2022. The destinations of research activity in Istat: statistical methods of business intelligence for enhancing territorial information, *Rivista Italiana di Economia Demografia e Statistica* - Vol. LXXVII-1 January-March 2023, pp 125-136

CAMISASCA M., PIETRANTONIO E., MAGRO R., ARBOREA A., FABIANI A., GIORDANO C., PIETRANTONIO M., SPARANO A. 2022. A pandemic health risk management model for the protection of workers: The Istat experience *Rivista Italiana di Economia Demografia e Statistica* - Vol. LXXVII-1 January-March 2023, pp 53-64

Alessandra DENTINI, ISTAT, aldentin@istat.it
Iole ZEPPIERI, ISTAT, iozeppieri@istat.it

EXPERIENCING AND REPORTING DISCRIMINATION AT WORK AGAINST LGB PEOPLE[1]

Eugenia De Rosa, Vincenzo Napoleone, Francesca Scambia

Abstract. This article shows the results that emerged from the Istat-Unar "Survey on Labour Discrimination against LGBT+ people" addressed to people in Civil Union or formerly in union) 2020/2021" with reference to several adopted subjective and more "objective" indicators of workplace discrimination. Logit regression models study the probability of experiencing discrimination events and reporting them as a function of some variables involving, for example, socio-demographic aspects, context, coming out, support by family and friends, perception of discrimination, and socio-economic status. This article provides a contribution on both, the methodological side of measuring discrimination, and the substantive side of knowledge about a complex and multidimensional phenomenon.

1. Introduction

Discrimination is a complex phenomenon not easy to detect. This refers to a situation, behaviours, and practices whereby a person is treated less favourably than others because of some of their characteristics (e.g., age, foreign origin, health problems, political ideas, gender, sexual orientation, gender identity, etc.) that in themselves are not relevant to the task or the context in which they are performing (direct discrimination). At the same time, discrimination refers to policies, practices, and behaviours that perpetuate inequalities among certain social groups (Krieger, 2014) and create obstacles in achieving equal opportunities and real equality for all (structural discrimination). Specifically indirect discrimination occurs when laws, policies or practices that appear neutral at face value, yet are discriminatory for population groups with certain characteristics (e.g. partner care leave, etc.). The focus of direct discrimination is on treatment consistency, both structural and indirect discrimination focus on outcomes (Praia group, 2020). There are several

[1] This article is the joint work of the authors, however paragraphs 1, 2, 3 are written by Eugenia De Rosa, paragraphs 4.2, 4.4 and 4.5 by Vincenzo Napoleone, paragraphs 4.1., 4.3 and 5 are written by Francesca Scambia.

ways to detect the discriminatory phenomena. A first practice is to use indicators that we might call outcome or indirect indicators to compare the situation of different population groups. A second way is to operationalize universal subjective experiences of discrimination. Finally, a third mode of conceptualizing and operationalizing discrimination is considering intersectional or certain groups' specific subjective experiences such as LGBT+ discrimination (De Rosa, 2022).

Multiple discrimination may be sequential (when a person suffers discrimination on different grounds on separate occasions) or additive (when a person suffers discrimination on the same occasion but on two grounds) but also intersectional discrimination when two or multiple grounds operate simultaneously and interact in an inseparable manner, producing distinct and specific forms of discrimination (Makkonen, 2002). The lived experiences of discrimination are also affected by the subjective perception. Perception is also affected by the level of awareness of one's own rights, the understanding of personal legal rights when it comes to discrimination and who to contact for support. Recognizing the signs of discrimination and reporting are key points in addressing discrimination and effecting change. Literature stresses the under reporting due to various reasons (e.g. reticence, underestimate of the incident, ineffectiveness, fear of coming out, fear of retaliation) and misrecording by those who receive reporting, namely police and other officials, that may record reasons other than discrimination (Praia group, 2020). These different phenomena and ways to conceptualized discrimination were operationalized in the Survey on Labour Discrimination targeting people in civil union, which Istat carried out in collaboration with Unar in 2020-2021. This census survey focuses on labour discrimination and aspects related to sexual orientation studying a segment of the LGBT+ population reached out by means of the municipal lists of civilly united persons. Since July 2016 the union of the over-18 same-sex persons has been regulated in Italy by a special institution named Civil Union.

The questionnaire collected data about personal information, family and socio-economic status, sexual orientation and coming out; aspects related to civil union, employment status, experiences of discrimination at work and in other contexts, reporting, experiences of discrimination, managing sexual orientation at work, aggression, hate speech, perception of discrimination against LGBT+ people in Italy, relationship with the LGBT+ community/associations and measures or initiatives that could be adopted in favour of the LGBT+ people in Italy. The main aim of this paper is to investigate labour discrimination against LGB people (in Civil Union) and the main associated factors as well as reporting discrimination, considering different aspects such as socio-demographic characteristics, coming out, support by family, friends and LGBT+ community, perception of discrimination and socio-economic status.

2. Conceptual and Measurement Framework: Discrimination and Reporting

The survey adopted a first working definition of employment discrimination as unfair and negative treatment of workers based on their personal characteristics that are not relevant to job performance (Chung, 2001). In line with the national and European laws[2] the main phases considered are: school/university, job search, work experiences and exit. We considered on the one hand indicators of general (not related to sexual orientation only) and specific labour discrimination: in the first case we listed subjective experiences conceptualized as universal and then we asked the respondent to indicate the characteristic(s) or reason(s) on the basis of which they believe they have been treated in this way; in the second case subjective experiences are conceptualized as specific to a particular population group ad for LGB people. On the other hand, we used outcome indicators of the working condition (e.g. employment status, hourly regime) that potentially allow comparisons among population groups. It should be noted that differences in outcomes do not in themselves constitute evidence nor do they indicate the extent of the prevalence of discrimination. From an analytical point of view, we also distinguished between formal and informal discrimination: the first concerns decisions and acts related to a worker's career (hiring, firing, promotion and retribution); the second concerns the working atmosphere, attitudes and interpersonal dynamics. The latter has been conceptualized considering the following dimensions: coming out at work, managing and coping with sexual identity at work, workplace relations, workplace microaggression. They are "brief daily interactions that communicate messages to some individuals as part of a group, subtle insults (verbal, non-verbal, and/or visual) directed at people often automatically or unconsciously" (Sue, 2010). Microaggressions allow to capture a more intangible and often unaware form of discrimination; hostile environment straddles the two dimensions (formal/informal).

We also distinguished between formal reporting of discrimination (trade unions, equal opportunity committee or trusted advisor, law enforcement agencies, employer) and informal reporting when people talk with family, friends, co-workers or employer. A range of possible actions has also been detected, namely: a legal action, labour conciliation, a direct approach to the offender and other actions.

Table 1 provides an overview of the conceptual and measurement framework.

[2] Legislative Decree No 216 of 9 July 2003 ('Implementation of Directive 2000/78/EC on equal treatment in employment and occupation').

Table 1 – *Conceptual and Measurement Framework of the Istat-Unar Survey on Labour Discrimination against LGBT+ people (in Civil Union or formerly in Union).*

Conceptual & Measurement Framework	Main indicator
1. Experiences of Discrimination at school/university, while looking for a job, as employee (current/last job) • by grounds of discrimination with reference to the last event	At least 1 event
2. Hostile environment and aggression in the workplace (current/last job) • by grounds of discrimination with reference to the last event	At least 1 event
3. Workplace microaggressions related to sexual orientation (current/last job)	At least 1 event
4. Disadvantage for sexual orientation during the working life • Career and professional development • Recognition and appreciation • Income and pay	At least 1disadvantage
5. Managing sexual orientation at work (current/last job) • Coming out • Avoiding to talk about private life, avoiding to hang out with people from the working environment in your free time, • Avoiding to attend corporate or other social events • Outing	Yes/No
6. Reporting (last event) • Hostile environment and aggressions in the workplace, by type of actor & non-reporting reasons • Microaggressions related to sexual orientation at work, by non-reporting reasons	Yes/No
7. Discrimination in other areas of life related to sexual orientation	Yes/No
8. Threats or aggression related to sexual orientation (last 3 years) and self-discrimination (during life)	Yes/No
9. Perception of Discrimination and desirable actions for LGBT+ inclusion	Level
10. Indirect/Structural Discrimination • Socio-economic statistics, comparison within and to other population groups	

3. Data and Methods

This study uses data from the "Survey on Labour Discrimination against LGBT+ people (in Civil Union or formerly in Union)", a CAWI survey based on a self-administered web questionnaire. The target population was made up of all resident individuals (over 21,000) who, as of 1 January 2020, were or had been in Civil Union (Law 76/20 May 2016). About 9,000 questionnaires were sent and validated; a post-stratification non-response was carried out.

Our analysis is restricted to those that by self-identification declared a homosexual or bisexual orientation (95.2% of the total).

This study investigates discrimination with a focus on hostile atmosphere or aggression at workplace and labour microaggression due to sexual orientation; reporting and actions following the experience of workplace discrimination. The population of the models varies depending on the response variable considered.

The two main research questions were: which are the main discrimination-associated factors? Which are the reporting-associated factors?

Descriptive analyses are presented. Then a multivariate analysis was carried out to study the experience of labour discrimination against LGB people in Civil Union (or formerly in Union) incorporating every single covariate at the $P < 0.05$ level, with stepwise multinomial logistic regression models, which allowed to calculate odds ratios (OR) with confidence intervals at 95%.

Our dependent variable is the probability of experiencing discrimination (hostile atmosphere or aggression at workplace, labour microaggression due to sexual orientation) and reporting such an event. Variables as regressors in the models concern the following dimensions: socio-economic characteristics, employment status and type of job, coming out and support, awareness of LGBT+ rights and involvement in the LGBT+ community. All regressors are dichotomous, exception made for quantitative variables (e.g. age, years in the same job).

4. Results

Descriptive statistics and models help to deepen the phenomena of discrimination against LGB people considering different dimensions and factors.

A combination of "objective or outcome indicators", "indicators about experiences of discrimination" and "group-specific" indicators (De Rosa, 2022) give a more accurate picture.

4.1. Labour market condition of LGB people and general discrimination

Same-sex couples in Civil Union represent a specific group of LGBT+ population living in Italy. This population is characterized by a high level of education and in their majority are currently out at work (for 92.5% of them, their sexual orientation is or was known to at least some of the people in their working environment). Standard indicators of labour market condition of LGB people in Civil Union or formerly in Union show a high level of participation in the labour market: 77% are employed and 22.5% have been employed in the past.

As it is for the entire population, women and young people are in a disadvantage position. A gender vertical and horizontal segregation in employment is observed for women who, more than men are employed in services as well as in executive or unskilled positions, and have more care responsibilities. Younger people report more discriminatory behaviours (De Rosa *et al.* 2022b). The same dynamics emerged when considering indicators of discrimination, not necessarily due to sexual orientation. A percentage of 46.9% homosexual or bisexual persons claim to have experienced at least one discrimination event at school/university (61.6% of the 18-34-year-olds), and 32% declare having suffered at least one event of discrimination in job search, not necessarily related to sexual orientation (28.3% of men), and gender is the most common reason they indicate (44.7%) in relation to the last event. Younger people as well, report more discriminatory behaviours than the total population in job search. Looking for a job as employee, 34.5% claim to have experienced at least one discrimination event (36.8% among women and 49.5% among the youngest).

4.2. Hostile atmosphere or aggression

The survey shows that about one in five homosexual or bisexual persons in Civil Union of formerly in Union (20.8%), employed or ex-employed in Italy, have experienced a hostile atmosphere or aggression not necessarily due to sexual orientation in their last job[3], with a slightly higher incidence among women (21.5% vs. 20.4%), young people (26.7%), foreigners or stateless people (24.7%) and people living in the South and Islands (22.6%).

[3] This means at least one incident of hostile atmosphere or aggression in the current/last job among: slandered, laughed at or played tricks on them, humiliated or insulted, deliberately excluded from meetings, conversations, etc.., offence, including making offers of a sexual nature, threatened verbally or in writing, totally deprived of tasks, subjected to unjustified disciplinary controls, physically assaulted.

The model puts into relation likelihood to have experienced at least one event of hostile atmosphere or aggression as the dependent variable. Sex, sexual orientation, age, education level, geographical area, employment status (employed/ex-employed), the type of job (dependent/independent), coming out at work (if the respondent's sexual orientation was known or not in the workplace), cultural capital of the family of origin (at least one graduated parent), number of years in the same job, working context size (less or more than 5 people) and type of contract (fixed-term or not) are regressor variables in the model (Table 2).

Table 2 – *At least one event of hostile atmosphere or aggression. Odds ratio.*

Parameter	Analysis of Maximum Likelihood Estimates			Odds Ratio Estimates		
	Estimate	Standard Error	P-value	Point Estimate	95% Wald Confidence Limits	
Intercept	0,163	0,171	0,3400	-	-	-
Age	-0,033	0,002	<.0001	0,968	0,964	0,972
Univ. degree and beyond	-0,145	0,045	0,0014	0,865	0,791	0,945
South and islands	0,270	0,063	<.0001	1,309	1,157	1,481
Employed	-0,655	0,055	<.0001	0,519	0,466	0,579
Graduated parents (at least 1)	0,120	0,061	0,0489	1,127	1,001	1,269
Coming out at work	0,754	0,104	<.0001	2,126	1,733	2,608
Years in the same job	-0,453	0,051	<.0001	0,636	0,576	0,702
Association of Predicted Probabilities and Observed, Percent Concordant=59.2						

Source: Survey on Labour Discrimination against LGBT+ people in Civil Union (Istat, 2022)

Sex is not a significant variable neither are the sexual orientation, employment status, coming out at work, cultural capital of the family of origin, years in the same job. Likelihood increases for those living in the South and Islands, and for young people who are more aware and sensitive to discrimination events. In particular being employed halves the probability to have experienced at least one event of hostile atmosphere, which probably means people leave a job when the environment is not fair. On the other hand coming out seems to give room to a hostile atmosphere, and this might explain why in Italy LGBT+ people are not so ready to it. An evidence of this is that sexual orientation is the characteristic most frequently mentioned (66.7%) among those for which they have been treated less favourably than others with reference to the last incident occurred.

4.3. Reporting hostile atmosphere or aggression

Descriptive statistics show people who experienced a hostile atmosphere at work prefer to talk informally about it with either co-workers/superiors within the work environment, or out of it with family members and friends. Formal reporting is hardly performed; it means people do not report to trade unions, equal opportunity committee, trusted advisor or law enforcement agencies (Istat, 2022). In order to deepen this phenomenon another regression model has been created. In it the dependent variable is likelihood of reporting events of hostile atmosphere or aggressions due to sexual orientation in the working environment[4]. Variables as regressors in the model are sex, sexual orientation, age, level of education, geographical area, employment status, the type of job, coming out at work, cultural capital of the family of origin, submitted request of marriage leave[5], number of years in the same job, involvement in LGBT+ community[6].

Table 3 – *Reporting events of hostile atmosphere or aggressions due to sexual orientation in the working environment - Odds ratio.*

Parameter	Analysis of Maximum Likelihood Estimates			Odds Ratio Estimates		
	Estimate	Standard Error	P-value	Point Estimate	95% Wald Confidence Limits	
Intercept	-2,176	0,439	<.0001	-	-	-
Homosexual	0,613	0,243	0,0117	1,846	1,146	2,974
University degree and beyond	0,266	0,103	0,0102	1,304	1,065	1,597
Centre	-0,291	0,112	0,0095	0,748	0,600	0,931
Coming out at work	1,736	0,371	<.0001	5,675	2,745	11,732
Request marraige leave	0,479	0,098	<.0001	1,614	1,331	1,956
Involvment in LGBT+community	0,459	0,097	<.0001	1,582	1,309	1,912
Association of Predicted Probabilities and Observed, Percent Concordant=58.7						

Source: Survey on Labour Discrimination against LGBT+ people in Civil Union (Istat, 2022)

[4] The items included: talking with superiors, co-workers, lower rank employees, equality bodies and trade unions. All these items are aggregated vs. talking to none, in order to create the dichotomy

[5] The law provides the right of marriage leave when a same-sex couple celebrates Civil Union. The dichotomy is between the respondents who declared they submitted this request vs. those who did not though entitle, whatever the reason why.

[6] A question investigated current or former participation in LGBT+ associations/groups vs. no participation/ prefer not to say.

Table 3 shows the results of the estimation of the model. Likelihood of talking about this event with people in the working environment increases for the homosexuals, people living in the South and Islands and in the North, among those who are more educated and aware of their rights and are close to LGBT+ associations. So people who know discrimination and who are committed to tackle it are those who more likely speak about it within the working environment probably with the idea of changing the environment from within.

4.4. Microaggressions for sexual orientation

According to the survey, approximately six out of ten people (61.8%) have experienced at least one form of microaggression at work related to sexual orientation[7]. The incidence is similar for men and women but is more frequent for homosexuals than for bisexuals (62% to 58.9%), by people with a medium-high educational qualification (62.7% with at least a degree compared to 58.9% with at most a law-secondary school diploma) and among employees/former employees (62.3% to 60.3% of the independents/former independents). A third model was developed.

Table 4– *At least one workplace microaggression for sexual orientation. Odds ratio.*

Parameter	Analysis of Maximum Likelihood Estimates			Odds Ratio Estimates		
	Estimate	Standard Error	P-value	Point Estimate	95% Wald Confidence Limits	
Intercept	1,770	0,144	<.0001	-	-	-
Age	-0,021	0,002	<.0001	0,979	0,975	0,982
Employed	-0,320	0,050	<.0001	0,726	0,659	0,801
Graduated parents (at least 1)	0,118	0,051	0,0201	1,125	1,019	1,242
Coming out at work	0,249	0,069	0,0003	1,282	1,120	1,468
Fixed-term work	-0,168	0,058	0,0037	0,846	0,755	0,947
Years in the same job	-0,650	0,044	<.0001	0,522	0,479	0,569
Association of Predicted Probabilities and Observed, Percent Concordant=57.5						

Source: Survey on Labour Discrimination against LGBT+ people in Civil Union (Istat, 2022)

[7] This means at least one microaggression related to sexual orientation in the current/last job among: hearing someone refer to someone as a faggot or use the terms 'lesbian', 'gay' or similar in a derogatory way, being asked about your sex life, that your gestures, speech and dress were imitated in order to make fun of you, that your sexual availability was taken for granted, that your partner was not invited to social events, that it was implied that you only got your job because you were homosexual or bisexual.

The dependent variable is the probability to have experienced at least one microaggression for sexual orientation in the current/last job. Variables as regressors are the same of model n.1 plus involvement in the LGBT+ community. According to the model, sex is not a significant variable; neither are sexual orientation, level of education, geographical area, the type of job, working context size, involvement in LGBT+ community (Table 4).

Likelihood increases among persons with high cultural heritage in their family of origin, among those who have come out at work. It decreases among those in fixed-term employment and as age increases.

4.5. Reaction to workplace microaggressions for sexual orientation

Looking at the reaction, 58.5% of the people who experienced a microaggression did not carry out any action as a reaction to the last incident occurred. Microaggressions are not direct acts of discrimination and therefore they are not always perceived as negative actions, this may explain why people do not openly react to them.

Table 5 – *Reaction to workplace microaggressions for sexual orientation (last event). Odds ratio.*

Parameter	Analysis of Maximum Likelihood Estimates			Odds Ratio Estimates		
	Estimate	Standard Error	P-value	Point Estimate	95% Wald Confidence Limits	
Intercept	-0,636	0,097	<.0001	-	-	-
Female	0,246	0,046	<.0001	1,279	1,168	1,400
Univ degree and beyond	-0,195	0,046	<.0001	0,823	0,753	0,900
South and islands	0,176	0,070	0,0121	1,192	1,039	1,368
Employed	-0,141	0,058	0,0150	0,868	0,775	0,973
Coming out at work	0,765	0,097	<.0001	2,148	1,777	2,596
Request marriage leave	0,243	0,049	<.0001	1,275	1,158	1,404
Involvment in LGBT+community	0,359	0,044	<.0001	1,432	1,314	1,562
Association of Predicted Probabilities and Observed, Percent Concordant=56.1						

Source: Survey on Labour Discrimination against LGBT+ people in Civil Union (Istat, 2022)

In the fourth model the dependent variable is the probability of performing an action as a reaction to microaggressions[8]. Regressors are sex, sexual orientation, age, level of education, geographical area, employment status, the type of job, coming out at work, cultural capital of the family of origin, fixed-term work, request marriage leaves, year in the same job and involvement in LGBT+ community. Results show that sexual orientation is not significant a variable, neither are age, the type of job, cultural capital of the family of origin, fixed-term work, year in the same job. Table 5 shows the results of the estimation of the model.

The likelihood of performing an action as a reaction to microaggressions is higher among females, people living in the South and Islands, those who have come out at work, those who are aware of their rights and are close to LGBT+ associations.

5. **Conclusion**

As above mentioned people in Civil Union are a collective with particular characteristics, made up of individuals who wanted to make use of the tools provided by the legal framework to have their status as a legally recognized couple. However, this study has great potential since it is a total investigation shedding light on the functioning of different indicators of discrimination. The gender gap is confirmed in some aspects of labour discrimination (e.g. firing, career, segregation). Age is another relevant aspect in all the phenomena/indicators considered: younger people are more vulnerable and show a higher rate of declaring "discrimination" incidents. The regression models evidenced how age and parent's high level of education make people more aware of their rights and able to recognize discrimination as such. Obviously, people who came out at work see an increased likelihood of being discriminated - not always related to their sexual orientation, though this is a factor that increases for this people the possibility of undergoing such events - but also of suffering and react to workplace microaggressions due to sexual orientation. Something similar can be said with reference to the act of reporting an hostile atmosphere in the workplace where one of the reasons reported is sexual orientation: likelihood increases for people who want to openly affirm their sexual orientation and claim for their rights; it is the case of requesting marriage leave, but also the active participation in the LGBT+ community. This study highlights how reporting in the working environment is a key issue in order to find solutions and responses to the discrimination incidents. Still it is not always performed. It means that education embracing difference is crucial for building awareness as also reported by the

[8] The dichotomy was between: I did nothing, or I did not know who to turn to or what to do vs. all the other items on possible actions.

interviewees according to whom training, awareness-raising activities or campaigns on LGBT+ diversity by public institutions are urgently required to foster the inclusion of LGBT+ people in the labour market (71.7%); in 89.1% of cases, they were also very much in favour of a national law against homolesbobitransphobia.

References

CHUNG, Y. B. 2001. Work discrimination and coping strategies: Conceptual frameworks for counseling lesbian, gay, and bisexual clients. *The Career Development Quarterly*, Vol.*50*(1), pp. 33–44. https://doi.org/10.1002/j.2161-0045.2001.tb00887.x.

DE ROSA, E. 2022. Intersezionalità e discriminazioni LGBT+: paradigmi, concetti e indicatori. *AG AboutGender*. Vol. 11, No. 22, pp. 306-336.

DE ROSA, E., de MARTINO, V., SCAMBIA, F., NUR, N. 2022. Perspectives on LGBT+ working lives: stakeholders, employers and LGBT+ people, RIEDS - *Rivista Italiana di Economia, Demografia e Statistica - The Italian Journal of Economic, Demographic and Statistical Studies*, SIEDS Societa' Italiana di Economia Demografia e Statistica, Vol. 76(2), pp. 4-12, April-Jun.

ISTAT. 2022. *Survey on Labour Discrimination toward LGBT+ People (in Civil Union or formerly in union). Year 2020-2021*. Roma, https://www.istat.it/it/files//2022/05/REPORTDISCRIMINAZIONILGBT_2022_en.pdf.

KRIEGER, N. 2014. Discrimination and health inequities. In: Berkman L.F., Kawachi I., Glymour M. (Eds). *Social Epidemiology*. 2nd ed. New York: Oxford University Press, 2014, pp. 63-125.

LEGISLATIVE DECREE NO. 216 OF 9 JULY 2003 - Implementation of Directive 2000/78/EC on equal treatment in employment and occupation.

PRAIA Group. 2020. *Praia Handbook on Governance Statistics* https://www.ohchr.org/sites/default/files/Documents/Issues/HRIndicators/handbook_governance_statistics.pdf.

MAKKONEN, T. 2002. *Multiple, Compound and Intersectional Discrimination: Bringing the Experiences of the Most Marginalized to the Fore*. Turku, Finland: Abo Akademi University.

SUE, D.W. 2010. *Microaggressions in Everyday Life: Race, Gender, and Sexual Orientation*. NJ, Wiley & Sons.

Eugenia DE ROSA, Istat, eugenia.derosa@istat.it
Vincenzo NAPOLEONE, Istat, vincenzo.napoleone@istat.it
Francesca SCAMBIA, Istat, scambia@istat.it

RESPONDENTS AND NON RESPONDENTS TO POPULATION AND HOUSING CENSUS: SOME STRATEGIES FOR DATA COLLECTION DESIGN IN THE ERA OF LOW RESPONSE RATE AND HIGH RESPONSE BURDEN. AN APPLICATION OF A DECISION TREE MODEL.[1]

Manuela Bussola, Novella Cecconi, Elena Donati, Linda Porciani

Abstract. The Italian National Institute of Statistics (Istat) has adopted the methodology of permanent Population and Housing Census in 2018, which is planned as an annual cycle. This study aims to identify the profile of respondents and non-respondents to the Census, in order to put respondents at the centre of the design and management of data collection and to develop specific strategies focused on subsets of the sample population. The methodology used was the 'Tree decision model', which was able to present the relative weight of the included variables on the independent variable through a good and direct visual layout. The first results confirm the relevance of socio-demographic variables, such as the age structure, education level, and occupational status of the sampled household, and present some new analytical insights capable of describing first the profile of non-respondents and then the profile of respondents by channel.

1. Introduction

The information collected by the Census does not simply provide a count of the people living in a given area but allows a picture to be drawn of how different groups, with specific characteristics, are located within the national territory. Moreover, census data form the statistical pillar of other official surveys and the backbone of the knowledge about different aspects of a population: households, transport systems, schools, hospitals, neighbourhoods and cities.

The essential need for this information must be combined with the challenges that are driving changes of the Census strategies (from a traditional to a permanent strategy): the management of a reduced financial budget, the respect of the timeliness of the data information; the decrease in the response rate. The answers are: a

[1] The paper was carried out by the joint work of the authors. More in detail, the single paragraphs are attributed as follows: paragraphs 1 and 3.1 to Manuela Bussola; paragraphs 2, 3, 3.2 to Novella Cecconi; paragraphs 2.1 and 4 to Elena Donati; paragraph 3.3 to Linda Porciani.

sampled, annual and integrated (both as data sources, survey and register, and as techniques, CAWI, CAPI and CATI) Population Census.

One of the most focused aspects of the Census process is the web response rate, because of its potentialities for responding to the real challenges: the 2022 edition of the Population Census shows a CAWI response rate of 44.3 percent (48.7 percent of responding households), 3 to 5 points lower than previous editions, with a huge difference between northern and southern regions. According to the literature, this gap could be due to several factors: territorial digital divide, different proximity of institutions to citizens in the territory, different organisational environment of municipalities regarding census operations, delays in the delivery of information letter, different living contexts and characteristics of households.

A major challenge is therefore to understand how to increase the CAWI response, by overcoming on the one hand the declining level of trust and confidence of people to respond to surveys, and on the other hand the lower participation of people living in hard-to-reach areas, those with low literacy levels, or those who by definition escape the official count, making the enumeration of the whole population a long, slow, and costly process (UNECE, 2021).

The aim of the present work is to outline the profile of responding and non-responding households in the 2022 Census of Population and Housing, in order to develop specific strategies aimed at increasing the household participation rate and, in particular the CAWI response rate.

To this end, the analysis of the characteristics of the survey sample is based on three aggregates by the most relevant variable: sampled households by response rate; respondents by technique; non-respondents by those contacted by the survey network.

This study presents the methodology used, and the main results of the three aggregates. At the end, it is possible to define some specific profiles of the census respondents, which could be the core for improving future editions of the census.

2. The model: the decision tree model

The population of sampled households in the Census is very heterogeneous in terms of socio-demographic characteristics, household conditions and place of residence.

In order to classify this population, a decision tree model is used because it allows to have homogeneous subgroups of predictors for the three aggregates: sampled households, respondents and non-respondents.

The classification algorithm is CHAID (CHI-squared Automatic Interaction Detection), which detects the interaction between variables in a dataset and it is best suited to the objectives of the analysis and the nature of the variables selected. CHAID identifies discrete groups and then, by examining the responses to the explanatory variables, attempts to predict the effect on the initial variable (KASS, 1980).

This multiple segmentation technique is based on the $\chi 2$ test to test the hypothesis of statistical independence between the dependent variable and the explanatory variables. For each modality of selected variables and for each combination of modalities, the model generates a contingency table (starting from the dependent variable) by calculating the $\chi 2$ and the corresponding p-value.

The Xi attribute with the smallest p-value (pmin) is compared to the threshold value α (this could be the maximum tree size, the maximum number of levels, or the minimum number of elements in a node):

If pmin$< \alpha$ the modality Xi is considered as an attribute of the partition
If pmin $> \alpha$ a leaf is identified.

If an attribute has more than 2 values, the model allows to group them in order to have homogeneous values with respect to the dependent variable.

2.1. The variables selected

The total number of sample households in 2022 is 998,540, of which 58,952 are non-target (moved, deceased, homeless), of the remaining 939,588 households, 874,976 were contacted and 64,612 were not reached by the survey network.

The analysis was carried out on "valid" sampled households (939,588 units) by response rate, on respondents (855,595 units) by channel; and on non-respondents (84,293 units) by whether or not they were contacted by the survey network.

Specifically, a dependent variable has been defined for each of the three aggregates: the response rate for sampled households; the percentage of response by channel for respondent households; the contact rate for non-respondent households[2].

The first step in the analysis is to define the theoretical dimensions behind the Census response rate. The propensity to respond to the Census is at the core of NSIs' the data collection strategies of the (ONS, 2011). In Italy in particular, it has been studied since the traditional population census in 2011 (Bernardini *et al*., 2014).

According to the literature (ISTAT, 2022a) and the working experience of the research group, they could be classified into: 1.the living context of the household, 2. the organizational characteristics of the Municipal Census Office (MCO), 3. the socio-demographic characteristics of the sample households.

The first dimension includes geographical variables in the form of administrative boundaries and the Inner Classification[3], which takes into account both the demographic dimension and the proximity of citizens to essential services (health, education and mobility) (ISTAT, 2022b).

The organisational dimension of the MCO is represented by the presence or absence of a statistical office, the number of sampled households per operator, the size of the MCO in relation to the sample size, the typology of participation in the census (every year/only one year in the census cycle), the completeness of the training activities of the MCO operators.

The third dimension, the socio-demographic characteristics of the households, is described as age structure, citizenship of household members, educational level and occupational status.

Other contextual information is also taken into account: trust in institutions, use of the Internet, sending of information letters, postal monitoring, and declaration of receipt of Istat letters by households.

[2] In 2022 there were 855,595 responding units and 84,293 non-responding units: 416,476 households responded via web, 81,932 households responded via telephone interview, and 356,887 were interviewed face-to-face by a municipal surveyor/operator. Among the nonresponding households, 19,681 were contacted at least once by a municipal surveyor/operator and 64,612 were never contacted.

[3] Inner Areas classifies municipalities according to their distance from three essential services: health, education and mobility. The map of Inner Areas is a tool that looks at the entire Italian territory in its articulation at the municipal level and identifies municipalities with a combined supply of three types of services - health, education and mobility - named Multi Municipality Service Center/Single Municipality Service Center. It also presents all other municipalities according to their distance from these Centeres (in terms of actual average road travel time), and classifies them into four bands of increasing relative distance – Belt Area, Intermediate Area, Remote Area, Ultraremote Area - and, therefore, with potentially greater inconvenience in accessing services. The municipalities classified as Intermediate, Remote and Ultraremote represent all of the Inner Areas of our country (www.istat.it/it/archivio/273176).

Based on the selected variables (Table 1), composite variables were developed to optimize the analysis without losing core information. Some other variables were excluded because of their excessive complexity or low variability.

The strategy of reclassification of variables was relevant for the dimension related to the household. In particular, it is necessary to construct variables for multi-component households that combine information on the age, citizenship, occupational status, and educational level of the members. The variable "Households by generation" takes into account both the number of members and their age (before/after the baby boom, i.e. is 1964) (ISTAT, 2017). The educational level has been identified as the highest level of education of the household members; and the same process has been adopted for the occupational status, defined as the most qualified occupation in the family.

Table 1 – *The selected variables.*

Household	**Territory**	**Fieldwork organisation**
Types of households by citizenship - Italian - Non Italian - Italian + non Italian	**Territorial aggregation** - North - Centre - South	**Statistical Office in Municipalilty** - Yes - No
Households by generation - 1 member until baby boom - 1 member after baby boom - More than 1 member until baby boom - More than 1 member after baby boom - More than 1 member mix	**«Inner Areas»** - Single-Municipality Service Centre - Multi-Municipality Service Centre - Belt Area - Intermediate Area - Remote Area - Ultraremote Area	**Field Workload** *[households for interviewer]* - Less than 50 - 50 - 100 - 100 - 150 - 150 - 200 - 200 - 250 - More than 250
Occupational status - Low - Medium - High		
Educational level - Low - Medium - High		

3. The main results

The main results can be represented through three tree decision models focused on:

1. the profile of households by response rate;
2. the profile of households by channel of questionnaire return
3. the profile of households not responding by contact with survey network.

3.1. The profile of sampled households by response rate

Figure 1 shows the results of the *decision tree* where response or non-response to the Census is considered as an independent variable[4].

The most significant variable for response rate is citizenship. Italian households are more likely to respond to the census. On the other hand, a quarter of all foreigners or mixed households did not respond.

The tree decision model makesit possible to identify specific significant variables that affect the three groups: Italians, foreigners and mixed households. Italian households are more influenced by the organisational context of the MCOs: they have a higher propensity to respond if they live in a more structured MCOs; while foreign households are influenced by the age composition of the household, and finally the response rate of mixed households is more related to the territory, according to the classification in "Inner Areas."

For Italian respondents, it is common that a higher workload for the municipal operators corresponds to a higher level of non-response: a high percentage of non-responding households have not been contacted by the local survey network, and have only received postal reminders and via the “IO App”[5] carried out by Istat. This is the "hard core" of people who are not used to responding online and who need to be "reminded" to respond.

On the other hand, if the workload of the MCOs is not particularly heavy, the area is likely to have an impact on the Census participation; while if the workload is high, the presence of a statistical office, with staff dedciated to statistical activities and probably better organised, facilitates contact and participation by Italian households, more than in a municipality without a statistical office.

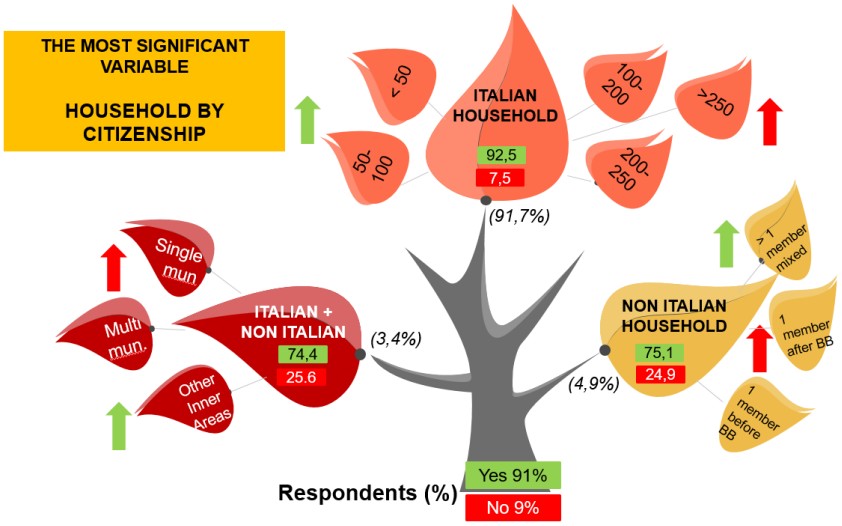
Foreign households, on the other hand, are the least likely to respond if the household consist of one person, regardless of the age: almost half of the one-person member households did not respond to the questionnaire. In particular, foreigners living in Single Municipality Service Centre (SMSC), Multi Municipality Service

[4] Respondents are 91 percent and non-respondents are 9 percent, the latter adjusted for non-targets.
[5] Io App is the APP for the communication among citizens and public administration.

Centre (MMSC) (both characterised by little or no distance from the three basic services) or Ultraremote Area (UA) (municipalities more distant from basic services) are more reluctant to partecipate in the Census.

On the other hand, living in municipalities with essential services related to health, education, and mobility seems to be a barrier to responding for mixed-citizenship households: the response rates are 65.5 percent of households in SMSC, 72.3 percent in MMSC and almost 80 percent in more remote areas.

Figure 1 – *Decision tree sampled households by response rate.*



3.2. The profile of respondents by channel of questionnaire return

The second *decision tree* was applied to the respondents according to the technique used to answer the questionnaire: CAWI, CATI, CAPI.

Educational level is the most important variable in the choice of response mode to the Census. As the level of education increases, so does the likelihood of responding online: almost 65 percent of households with at least one member with high level of education opted to complete the questionnaire via web, while 34 percent of households with a low level of education opted for a face-to-face interview at home or in Municipal Survey Centres (MSCs). The telephone is the less used channel and, moreover, it is almost exclusively an “outbound” response, i.e., due to

interviewer's reminder rather than the household's own request. The use of CATI increases as the level of education decreases.

For the most and least educated household the influence of the living places on the choice of the channel used to complete the census form is common. Living in municipalities provided with basic services (or not far from them), increases the propensity to respond via web. In more peripheral areas, the propensity decreases by at least ten percentage points.

Even if the effect of living places is the same, the response rate differs according to the level of education: the most highly educated households living in SMSC and MMSC show an increase in the CAWI response, especially if the highest occupation in the household is medium-high (about 70 percent of the households living the "central" municipalities).

On the other hand, more educated households living in a peripheral territorial context are more influenced by the number of members and generation: multi-component households composed of people born after the baby boom prefer web-based responses.

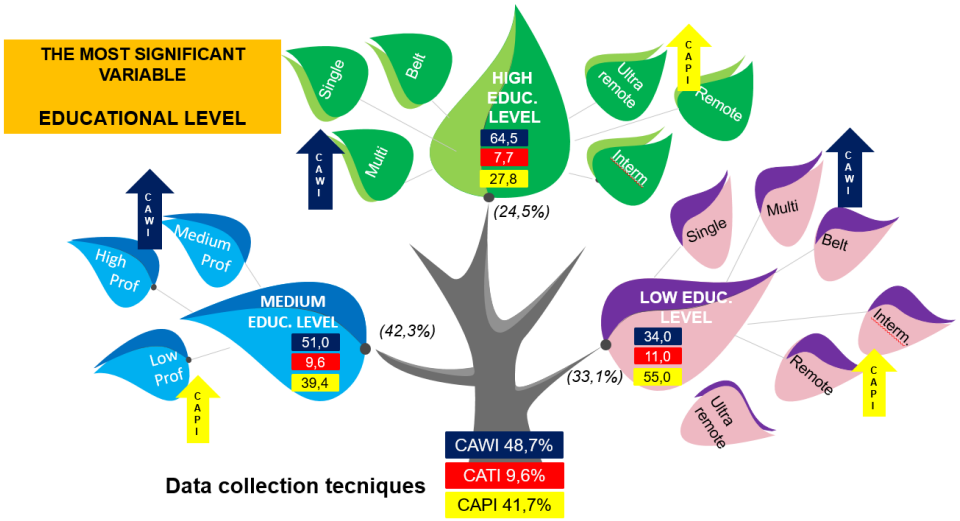
Finally, education level and living context are the most important variables in the choice of the web channel, regardless of the age of the respondent.

For households living in SMSCs, regardless of the number of household components, the propensity to use the face-to-face interview is higher for younger households, while the propensity to use the web is higher for older households. It is possible to observe the same factor influencing the web response of low educated households in the other territorial contexts, except for those living in MMSC, which are more affected by citizenship.

Occupational status is a significant factor in web use for households with a high school degree as the highest educational qualification.

A higher percentage of CAWI respondents is observed when the highest occupation is medium (64.5 percent) or a high (56.1 percent). The second influencing factor in this "leaf" is the living place: households living SMSC show a high propensity to use the web channel.

Figure 2 – *Decision tree responding households by channel of questionnaire return.*



3.3. The profile of non-respondents by contact with the survey network

The third *decision tree* was applied to non-responding households with the aim of describing their characteristics and to optimising the design process, especially at the stage of planning targeted field recovery activities.

The majority of non-responding households were not contacted by the local survey staff, especially if the households lived in SMSC and MMSC, where 80.0 percent of non-responding households had no contact with the survey staff: 83.6 percent and 79.8 percent respectively.

The probability of being contacted increases for non-respondent households living in the Belt or Intermediate Areas: 15 percent and 11 percent higher respectively compared to Single/Multi Centre.

Finally, households living in more remote municipalities (representing 6.8 percent of non-respondents) are also more likely to be contacted, at least once, by the municipal survey network: 71.8 percent of peripheral households escaped any kind of contact with census staff.

Following the tree model, the second most significant variable is the workload of the municipal census network, which is positively correlated whit the absence of contact with the sampled household.

In the case of the SSMCS, this relationship is evident: the higher the number of households assigned per operator, the higher the percentage of non-responding never

contacted households. When the local census network is overloaded, more than 250 households per operator (0.7 percent of the initial population) were no reached by the network.

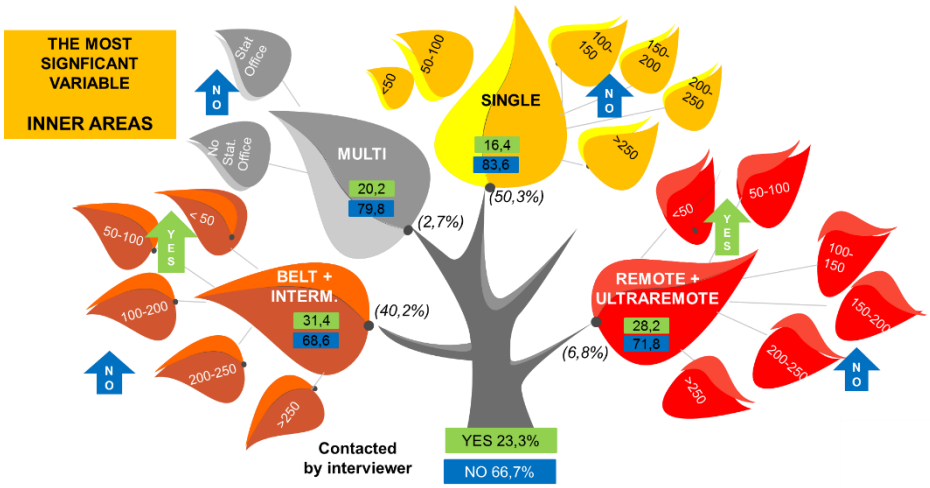
Households living in the Belt or Intermediate Areas (40.2 percent) show a lower non-contact rate when the workload is lower; in particular, the probability of being contacted is higher when an operator manages than 100 households, and when the workload is minimal (less than 50 households per operator). In the latter case, 41.7 of non-respondents are contacted at least once.

Looking at the more distant leaves of the tree, another important factor is the generational composition of the households.

In general, households with a one-member born after the baby boom are the most difficult subgroup to reach, regardless of the workload per operator. In fact, in the municipalities with the highest proportion of non-respondents, the generational composition becomes the second significant factor explaining the probability of not being contacted: 85.3 percent and 88.0 percent respectively in the municipalities with low and medium workload municipalities.

The same influence is present in the suburban and ultra-peripheral municipalities, where, although the presence of non-respondents is lower than in the Single and Multi-Service Centre, households with one or more members born after the baby boom are more likely to escape contact than other households (75.1 percent against 64.6 percent in municipalities with 50 to 100 households per operator).

Figure 3 – *Decision tree households not responding by contact with survey network.*



4. Conclusions

The Population Census is the largest statistical operation carried out by the NSIs and it presents complexity in each phase of the statistical process design and management; it should face the societal changes in terms of work organisation and household behaviour. These are some of the reasons of the need to have study and reaserch about Popolutation Census data and process to improve the quality and to update the process steps. The present study provides some insights based on the census results.

From the analysis of the *decision trees*, it is clear that the behaviour of non-respondents and respondents and the survey techniques show a homogeneity that crosses geographical boundaries. This means that strategies for implementing Computer-Assisted Web Interviewing (CAWI) and optimising non-response recovery can be applied across the country, following paths that do not necessarily correspond to administrative boundaries.

The results show us that the core of a renewal of the design process should be the characteristics closest to people's lives and to the organization of fieldwork. In particular:

- a high workload has a negative impact on both response rates and the probability of households being contacted;
- households with at least one foreign member have a low propensity to complete the census questionnaire, especially if they are single-member households or households living in Single Municipality Service Centre or Ultraremote Area;
- living in a Single Municipality Centre has a negative effect on both response rates and the probability of households being contacted, while at the same time, it has a positive effect on the use of the CAWI technique.

These results suggest that the planning of the future censuses edition could be enriched by the inclusion of some new elements in order to increase the response rate, especially the web response, and to reduce the number of sampled units not contacted:

- reducing the workload of the local survey network;
- an improvement in communication campaigns targeted at households with low levels of education and suburban contexts;
- the introduction of a smartphone questionnaire: "responsive" and accessible via QR Code;
- increasing of the contact information of the sampled units: i.e. mobile phone numbers available to the local census staff;
- setting up of an on line booking system for the sampled units to make an appointment to complete the questionnaire.

Further and detailed analysis of the different profiles of sampled households should be developed to be able to better manage this complex operation.

References

BERNARDINI A., FASULO A., TERRIBILI M.D. 2014. A model based categorisation of the Italian municipalities based on non-response propensity in the 2011 census, *Rivista Italiana di Economia Demografia e Statistica*, Vol. LXVIII, No. 3/4 (Luglio-Dicembre 2014), pp. 77-84.

ISTAT, 2017. *Rapporto annuale 2017*. Roma: Istat.

ISTAT, 2022a. *BES | 2021. Il benessere equo e sostenibile in Italia*. Roma: Istat.

ISTAT, 2022b. *La geografia delle aree interne nel 2020: vasti territori tra potenzialità e debolezze*. Statistiche Focus. Roma: Istat.

KASS G.V. 1980. An explanatory tecnique for investigating large quantities of categorical data, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 29, No. 2, pp. 119-127.

UNECE, 2021. *Keeping count. Conducting censuses during the covid-19 pandemic*. Geneve: Unece.

ONS, 2011. *Predicting patterns of household non response in the 2011 Census.* London: ONS.

Manuela BUSSOLA, Italian National Institute of Statistics, bussola@istat.it
Novella CECCONI, Italian National Institute of Statistics, ncecconi@istat.it
Elena DONATI, Italian National Institute of Statistics, eldonati@istat.it
Linda PORCIANI, Italian National Institute of Statistics, porciani@istat.it

REACHING THE UNREACHABLE. INNOVATIVE STRATEGIES TO APPROACH HARD-TO-REACH POPULATIONS

Nadia Nur, Federico Di Leo

Abstract. This paper presents an overview of the methodological approaches used by Istat to study hard-to-reach populations, focusing on Roma and Sinti people living in settlements and the homeless in Rome. These populations, characterized by their marginal status and preference for anonymity, challenge traditional sampling methods, necessitating innovative strategies. Our objective is to outline and compare innovative methodological strategies that Istat is implementing to carry on: a) a survey on housing conditions and social inclusion of Roma and Sinti people living in settlements; b) a count/survey of homeless people. For the Roma and Sinti survey, we employ a random walk technique for household selection in settlements, relying on self-identification and specific guidelines for contacting households. In contrast, the homeless count utilizes a point-in-time approach to estimate demographic and social characteristics. This involves a "snapshot" count conducted in a high-density homeless neighborhood. By comparing these methodologies, we aim to contribute to the advancement of official statistics in the realm of investigating hard-to-reach communities.

1. Data collection on Roma households living in settlements

1.1. Background

The Roma community in Italy has long faced significant housing challenges, residing in settlements with substandard living conditions and lacking access to proper infrastructure and basic services. These issues are compounded by historical discrimination, poverty, and social exclusion. To tackle these pressing concerns, policies aimed at relocating Roma households to adequate housing and enhancing social inclusion have been implemented. In pursuit of the objectives identified by the National Strategy for Roma, Sinti, and Caminanti inclusion 2012-2020 (Unar, 2022), Istat built up a set of surveys, with the cooperation of the National Platform of Roma

associations and Unar[1]. The overall project carried out by Istat adopted a two-stage selection approach to assess the effectiveness of policies and for conducting a comprehensive survey on housing conditions and social inclusion of Roma households. The first stage involved a broad survey covering municipalities with over 15,000 residents. The primary objectives were to identify Roma households that had been successfully relocated to adequate housing from informal settlements, evaluate the efficacy of policies related to such relocations, and quantify the number of Roma people still residing in both formal and informal settlements (Nur, 2021). Through stage one, municipalities with significant Roma populations have been selected for further investigation in stage two. Starting from the information collected from the municipalities, we built a list of households that were living in adequate housing, and we located Roma settlements (information has been updated in 2021). The second stage of the research focuses on in-depth investigation and involves two target groups: Roma households currently residing in adequate housing after relocation and households living in formal settlements[2]. The overarching objectives of the second stage are twofold: a. To assess the living conditions and level of social inclusion of Roma households that had been successfully relocated to adequate housing. This involves evaluating the improvements in housing conditions, access to basic services, and overall well-being of the residents. b. To examine the housing conditions and social inclusion status of households living in formal settlements, with an emphasis on identifying the challenges they face and informing future policy initiatives. To ensure accurate and culturally sensitive data collection, the study employs Computer-Assisted Personal Interviews (CAPI) with a structured questionnaire. This section outlines the use of the random walk technique for data collection and its adaptation and the potential benefits and challenges of implementing this technique in Roma settlements, providing a comprehensive overview of the process and procedures followed by interviewers to ensure the survey's representation of the Roma population[2].

1.2. Data collection techniques in Roma settlements

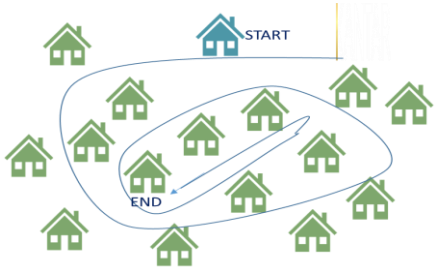
Studying Roma households living in settlements is crucial for understanding their living conditions and social dynamics and identifying areas for improvement although they often present unique methodological challenges due to their marginalized status, social complexities, and lack of official recognition. Traditional

[1] ISTAT-UNAR collaboration agreement for the establishment of a statistical information framework on social disadvantage and housing conditions of Roma, Sinti and Caminanti (Rsc) people. The execution of the agreement is supported by PON Inclusion ESF 2014-2020 to Axis4 - Specific objective 11.1, Action 11.1.3.

[2] Data on Roma settlements have been collected from March to June 2023. Results of the survey will be published by mid 2024.

sampling methodologies often face significant challenges, making them less effective or not suitable for these populations, Typically, they rely on a well-defined and accessible sampling frame, such as a list of households or individuals. Hard-to-reach populations, like Roma camps dwellers or homeless often do not have fixed addresses or are not included in standard registries, making it difficult to create a comprehensive and accurate sampling frame. Moreover, these populations can be highly mobile and difficult to locate and follow up with. Thus, traditional data collection methods may yield limited results. To achieve accurate and comprehensive data collection, an unconventional and tailored methodology is necessary, as outlined by other European surveys (Fra, 2023 and Fra, 2022)). The random walk technique offers a promising alternative for collecting data in such contexts, potentially mitigating some of the difficulties encountered by researchers. Among the key factors contributing to the survey's success, the correct implementation of household selection procedures plays a crucial role. These selection procedures are built upon an innovative random walk approach, which incorporated a direction matrix to enhance accuracy. To ensure systematic coverage during the random walk, researchers must delineate the boundaries of the informal settlement. By defining the survey area, the possibility of missing households or biased data collection is reduced. The approach adopted by the Italian National Institute of Statistics (Istat) involves a random walk technique with a predefined selection interval from an assigned starting point within the "Roma camp", usually the main entrance if visible (figure 1). Interviewers proceed to move randomly within the settlement, following specific rules for each step. To maintain consistency, interviewers move in concentric circles, following a counterclockwise route. This systematic movement allows for comprehensive coverage of the settlement, reducing the likelihood of missing households. This method eliminates potential interviewer bias, allowing for an equitable representation of households.

Figure 1 – *Random route.*



Interviewers apply a sampling step of five, meaning they make contact with every fifth dwelling unit in the camp. In the case of small camps, a sample interval of three is preferred to obtain a sufficient number of data points (Figure 2).

Figure 2 – *Selection interval.*

By employing a predefined selection interval and adhering to specific guidelines during data collection, the random walk technique proves effective in capturing the heterogeneity of the settlement while avoiding discretionary decisions and enhancing the screening process for accurate identification of the target population.

Owing to the sensitivity of asking if they have a Roma background, enumerating and screening for eligible households is usually a step conducted by the interviewers. A mandatory question to ascertain whether the respondent belongs to the Roma community is posed by the interviewers. This method helps avoid misclassification and enhances the accuracy of data collection, while ensuring that only willing participants are included in the survey. The interview can be done only if the head of the household, his/her spouse or partner, or an adult member of the family (+18) who lives in the selected house accepts to be interviewed. If no suitable respondent is found, the interviewer schedules an appointment. If nobody is found in the selected dwelling unit, the interviewer leaves his business card to schedule an appointment with the head of the family or another person suitable for the interview. By employing this approach, researchers can overcome some of the barriers associated with conventional methods and gather more comprehensive and representative data.

1.3. Challenges and potential benefits of random walk data collection

Data collection in Roma settlements is challenging due to various factors, including:

- mistrust and stigmatisation: the Roma community may be reluctant to participate in data collection activities due to historical discrimination and stigmatisation. Some may be reluctant to self-identify or be unwilling to participate in research for various reasons, including mistrust of researchers.
- language barriers that may hinder effective communication between researchers and residents, impacting the quality of data collected

- conducting field research in settlements can be logistically challenging since it requires gaining access to the camp and obtaining permission from its residents.

Engaging with the community and seeking their participation and input throughout the research process can help to mitigate some of these challenges and ethical concerns. Collaboration with associations (Roma and pro-roma) is essential to the success of this study. Local associations play a crucial role in facilitating access to Roma communities, building trust, and ensuring cooperation with the survey. The involvement of associations and Roma communities strengthens the survey's outreach and credibility. Moreover, this study includes Roma interviewers where possible and recognizes the significance of cultural mediation. Roma interviewers possess a unique understanding of the community's culture and norms, establishing a comfortable environment for respondents. Their presence fosters mutual trust, thus yielding more accurate and insightful data. Using random walk techniques in vulnerable communities like Roma settlements also raises ethical concerns. Interviewers must be mindful of the potential impact on the residents and ensure that their work respects the dignity and privacy of the individuals involved. Interviewers with Roma background and cultural mediators can assure more awareness of ethical aspects. To ensure data quality, proper training is provided to the interviewers regarding the methodology and approach. Additionally, regular supervision and monitoring of the data collection process are conducted to address any issues promptly. Although the random walk technique can potentially mitigate some of the difficulties that can be encountered by researchers, still we can outline certain challenges and limitations such as:

- irregular layout of the settlement that might not fit the assumptions of traditional random walk models, which are often designed for more regular environments. This irregularity can introduce challenges in accurately modeling the movements and interactions of individuals within the camp.
- the duration of data collection for random walk experiments can significantly impact the results. A longer observation period may provide a more comprehensive understanding of movement patterns, but it can also be resource-intensive and impractical in certain situations.
- there may be concerns about the representativeness of the sample, as the randomness of the approach might miss certain segments of the population.
- the technique can introduce biases, as certain areas or groups may be more likely to be included in the sample.

By addressing some of the challenges associated with traditional methods, the random walk technique offers a means to gather more comprehensive and representative data, leading to better insights into the realities of the Roma communities residing in settlements. However, its successful implementation

requires careful planning, ethical considerations, and adaptability to the unique circumstances of each Roma camp.

2. Methodological strategies to investigate Homelessness: the Street Count in Rome

2.1. Introduction and background

The aim of this section is to provide an overview of methodological approaches adopted by Istat and by other research teams to study a specific hard-to-reach population, homeless people (Marpsat, M., & Razafindratsima, N. 2010).

The city of Rome is used as case study, as the total homeless population has been repeatedly estimated over time using different techniques[3]:

- PIT - point-in-time sampling (in 2014);
- TLS - time-location sampling (in 2011, 2014 and 2021);
- AD - administrative data (Population Census 2011 and 2021).

This population constitutes a very small proportion of the general population but represents a very sensitive target for national and local authorities to fine-tune contrast policies due to their marginalised status. Homeless is a hard-to-reach population, for whom the traditional sampling methods are rarely feasible. The goals of the research are basically two: homeless count and the study of the individual condition and story. Census data are focused only on homeless count while the other approaches provide information on both aspects of the statistical investigation. In principle PIT is more directed to make a "snapshot" of people experiencing homelessness, on a single night. TLS is more suitable for and investigation of the condition of unsheltered people and the causes which led to a situation of extreme poverty. The Population Census, however, provided in 2011 [AD_PC 2011] and 2021 [AD_PC 2021] its own estimate of homeless people including an estimate of the presence of this population in Rome. In 2014 Istat considered also the possibility of a direct count of the people spending the night on the street, in public spaces, or in other places that are not considered appropriate shelter. The street count considered only the city of Turin because of the possibility to rely on a good network.

[3] A partial study has been carried out by Di Censi and others with the capture-recapture method, using a database (2022-2011) provided by the Municipality of Rome. The results are not used for this document. See Di Censi, L. (2013). Metodologie applicate per la misurazione della povertà urbana. FrancoAngeli.

In 2011 the ISTAT with the support of partners[4] realised the first National Research [TLS_NR 2011] on the condition of people living in extreme poverty. Three years later, in 2014, Istat and its partners carried out second research. [TLS_NR 2014]. "Fondazione Rodolfo Debenedetti" with the support of the Municipality (Roma Capitale) carried out in March 2014, with the Project "RacContami" a first attempt of direct count of homeless [PIT_RR 2014]. There are no further statistics on homeless people based on convenience sampling or "street count" at a national level or focusing on the situation in Rome after 2014. In 2021 a researcher with the support of the European University Institute[5] implemented a survey based on Istat approach, for the Municipality of Rome only [TLS_RM 2021].

2.2. Homeless count: a first comparison of the results

Homeless count in Rome according to various sources is summarised in Table 1. National Research (2011) estimated that the homeless population in Rome consisted of 7.827 while according to 2011 Population Census was 21.118 while ten years later, in 2021, was 22 182. These latter figures are mutually consistent, but they are different from the estimates based on the two National Research carried out by Istat (7.709 individuals in 2014). The "street count" conducted in 2014 with RacContami Project produced a result far different but consistent with the so-called "Rome Research". Before doing precise comparisons, it is necessary to understand why the homeless count according to the Population Census is so different from other estimates.

National Census defines homelessness according to official demographic records (*anagrafe*). The "anagrafe" registers considers people with two main characteristics:

- people who registered in conventional addresses and/or addresses of assistance bodies/institutions in which the homeless are present;
- people who are regularly present in Italy and therefore have the possibility to have an official residence address.

[4] The research was carried together with the Ministry of Welfare, the Italian Federation of Organizations for Homeless People (fio.PSD) and Caritas. See Istat (2014). L'indagine sulle persone senza dimora in Italia. Roma.

[5] See Gatta, A. (2023). A target on the head? Welfare targeting, conditionality and the persistence of disadvantage. PhD Thesis in Political and Social Sciences, European University Institute, Florence.

Table 1 - *Homeless count in Rome – Various sources and methodological approaches (2011-2021).*

Source	Point in Time	Time Location Sample	Administrative Data
National Research 2011		7.827	
Population Census 2011			21.118
RacContami 2014	3.276		
National Research 2014		7.709	
Population Census 2021			22.182

Istat (2011, 2014 e 2022), Braga (2014, Gatta (2022)

According to these two features many people who are registered in conventional addresses are not necessarily sleeping in open spaces or "unconventional shelters" but are persons who are eligible for an unconventional registration. In Rome, the Social Services Secretariat decides if there are conditions and there is not any official regulation to include people in this specific register. Therefore, data include many people who are in need of a special address but are not homeless like persons in collectives, civilian residents temporarily absent from the country, persons living in areas to which access is difficult, etc.

However, the estimates based on TLS and PIT approach are also partly inconsistent because of several reasons which is possible to list:

- difference in definitions;
- sampling vs snapshot approach;
- period of the year of the research (winter, spring, etc.);
- possible double counting;
- extension of period of the research (single night vs a period of research);
- special situations like the Covid-19 pandemic;
- non-sampling errors and quality of data collection.

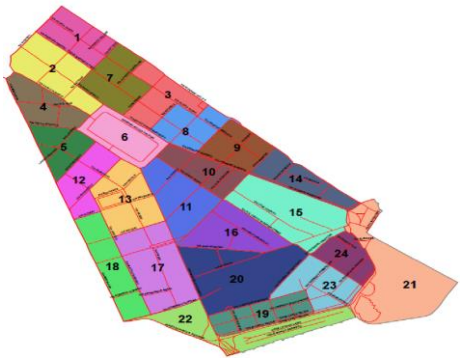
This list can be extended, and it is possible to find solutions to arrive at a (partial) reconciliation between the different methodologies. However, I want to stress the fact that each approach has its own validity and naturally PIT is much more expensive, complex and resource demanding than TLS. There are no advantages and/or limitations adopting the three different techniques examined: point-in-time sampling (PIT); time-location sampling (TLS); administrative data (AD) estimates are not suitable for direct comparisons. However, a few further clarifications are necessary. Estimates based on time-location sampling provide information on homelessness over a period that varies according to the approach taken; in the case of the Istat surveys the time window adopted was one month wide. Therefore, the

PIT estimate and the TLS estimate only coincide under the assumption that there were no "entries" and/or "exits" from homelessness condition during the period considered. As regards to the administrative data the estimate is affected by the source characteristics: part of the homeless are not counted, i.e. those who are not registered in the "registry office-anagrafe" (or those who are not identified as homeless because they are still registered at their previous address of usual residence) while part of the "registry" as homeless may not be homeless and/or may not be in severe housing deprivation (addresses dedicated to homeless are also used by municipalities for categories that do not belong to the target groups such as unaccompanied minors in care, women victims of violence, caravanners, people who for personal/fiscal reasons find it convenient to be homeless, etc.).

2.3. Street Count work in Paris and the Pilot Survey in Esquilino District.

Both TLS and PIT surveys face a set of problems concerning, from the respondent side, willingness or ability to answer the questions, language barriers, fear etc. Therefore, an option for avoiding bias that can lead to underestimating the target population might be to collaborate with community organisations with access to these groups. The Pilot Survey in Esquilino District was implemented to verify the possibility to make a new PIT estimation of homelessness in the Municipality of Rome according to the approach used in Paris and other major cities. Aware of possible biases of previous research, the first goal was to arrive at a new homeless count with few demographic features (age, sex, continent of origin). Following the methodology used in Paris, starting around 9pm, small teams of volunteers walked every street in one of 24 sectors of Esquilino District (figure 3).

Figure 3 – *Map of Esquilino district.*



Four more teams were dedicated to the area of the main railway station (Stazione Termini) where it is possible to meet a large number of homeless. Each team had a

leader, but all the members were responsible for the various aspects of the survey: filling out the forms properly, detecting the people to be counted, and identifying the path to be followed. For every individual, the team completed a questionnaire with basic observable information. For those willing to be interviewed, the volunteers could ask more targeted questions – with responses to remain strictly anonymous. Once the team covered every street in the assigned sector, they returned completed questionnaires to the neighbourhood headquarters for an initial quality control.

2.4. Challenges of PIT technique

Practical implementation challenges associated with street counts can affect the accuracy and richness of the data. Given the transitory nature of rough sleepers during the day, the counts generally take place at night: but if you start too early, people may not have settled in for the night and you'll miss them; if you start too late, many people are sleeping, and data collection will be limited. Street counts are also resource-intensive, relying on many enumerators (around 2000 for the Paris count alone), who require training and supplies. Street counts also depend on relatively favourable external conditions: extreme weather or a global pandemic, for instance, complicates data collection efforts and can limit comparability across time.

The Pilot Survey results as well as the specific characteristics of Rome require specific adjustments for the final research, scheduled for the end of 2023.

Among the main changes to be introduced the most relevant are connected with:

- the questionnaire;
- the introduction of a CAPI approach;
- the correct identification of the Municipality to be covered (Rome is much bigger than Paris in term of square meters but not in terms of population);
- the number of teams to be selected for the total snapshot;
- the support of people belonging to charities vs people selected from universities (enumerators);
- the possibility to have a partner responsible for the organisation.

3. Conclusions

Homeless and Roma communities living in settlements are considered hard-to-reach populations when it comes to data collection. However, obtaining accurate and comprehensive data on these vulnerable groups is crucial for designing inclusive policies that address their specific needs and challenges effectively. To achieve this, survey design must consider several key considerations. Both surveys pose methodological challenges, such as the lack of a reliable sampling frame and the high mobility of the population, necessitating the use of innovative and flexible approaches. These approaches should strive to balance the need for accurate,

representative data with the practicalities of reaching and engaging these groups effectively. Moreover, both surveys face a set of problems concerning from the respondent side, willingness, or ability to answer the questions, language barriers, fear etc. Therefore, an option for avoiding bias that can lead to underestimating the target population might be to collaborate with community organisations with access to these groups. These organizations have established connections and trust within the target populations, making them valuable partners in the data collection process. Their involvement can facilitate access to the communities and help overcome barriers that might hinder conventional data gathering efforts. Another approach that can encourage participation and ensure that respondents feel comfortable sharing their experiences and perspectives is to involve interviewers or enumerators with specialised skills in cultural mediation or deep knowledge of the target population to create contact with the target population and gain their trust. In addition, it is essential to provide proper training to interviewers on the nuances of handling 'official' data collection. Finally, official statistics agencies should embrace innovative and adaptable data collection methodologies and techniques to obtain high-quality data from these hard-to-reach populations.

References

APUR. 2021. Les personnes en situation de rue à Paris la nuit du 30-31 janvier 2020. Available at https://www.apur.org/sites/default/files/nuit_solidarite_2020.pdf?token=5LtIDAe7.

BRAGA M. 2014. *Raccontami 2014 - prima indagine point in time dei senza fissa dimora a Roma*. Roma: Fondazione Rodolfo De Benedetti.

EMPLOYMENT AND SOCIAL DEVELOPMENT CANADA ESDC. 2021. Everyone Counts – A Guide to Point-in-Time Counts in Canada. Available at https://homelessnesslearninghub.ca/wp-content/uploads/2021/06/2829-Pit-Count-Report-edition-3.1-EN54.pdf.

FRA - FUNDAMENTAL RIGHTS AGENCY. 2023. Roma Survey 2021 Technical report. Available at http://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-roma-survey-2021-technical-report_en.pdf.

FRA - FUNDAMENTAL RIGHTS AGENCY. 2022. Report Roma in 10 European countries – Main results: Roma Survey 2021. Available at https://fra.europa.eu/en/publication/2022/roma-survey-findings.

GATTA A. 2022. Non avere dimora a Roma. Accesso alla residenza, alle misure di sostegno al reddito e alle cure mediche per le persone senza dimora della Capitale.

Osservatorio Fio.PSD. Available at https://www.fiopsd.org/report-progetto-instrada/.

INGLESE F. AND MASI A. 2019. Una metodologia per la stima della homelessness attraverso le Unità di strada: il caso della città di Torino, *Istat Working Papers No.5*

Insee. 2021. Nuit nationale de la solidarité - Compter pour agir. Guide méthodologique, *Insee Méthodes*, No.139.

ISTAT. 2022, Popolazione residente e dinamica demografica. Available at https://www.istat.it/it/files/2022/12/CENSIMENTO-E-DINAMICA-DEMOGRAFICA-2021.pdf .

ISTAT. 2012. Le persone senza dimora. Available at https://www.istat.it/it/archivio/72163.

ISTAT. 2015. Le persone senza dimora. Available at https://www.istat.it/it/archivio/175984.

MARPSAT, M. RAZAFINDRATSIMA, N. 2010. Survey methods for hard-to-reach populations: introduction to the special issue. *Methodological Innovations Online*, Vol.5, No.2, pp.3-16.

NUR N. 2021. Abitare in transizione. Indagine sui progetti di transizione abitativa rivolti alle popolazioni Rom, Sinte e Caminanti. Available at https://www.istat.it/it/archivio/254436.

SCHNEIDER M., BRISSON, D. & BURNES, D. 2016. Do We Really Know how Many Are Homeless? An Analysis of the Point-In-Time Homelessness Count, *Families in Society*, Vol.974, pp. 321–329.

UNAR. 2022. Strategia nazionale di uguaglianza, inclusione e partecipazione di rom e sinti 2021-2030. Available at www.unar.it/portale/web/guest/strategia-rsc-2020-2030.

UNAR. 2012. Strategia nazionale di uguaglianza, inclusione e partecipazione di rom e sinti. 2012-2020. Available at https://www.minori.gov.it/it/minori/strategia-nazionale-2012-2020-dinclusione-dei-rom-dei-sinti-e-dei-caminanti.

Nadia NUR, Istat, nur@istat.it
Federico DI LEO, Istat, dileo@istat.it

THE DEMOGRAPHIC CRISIS CAN AFFECT QUALITY OF LIFE IN ITALY: A MULTI-DIMENSIONAL ANALYSIS[1]

Carolina Facioni, Isabella Corazziari

Abstract. Official data confirmed the inexorable aging of the Italian population. A process historically attributed to the effect of the demographic transitions. However, in Italy it also may be due to specific social/cultural aspects, considering the different performance respect to the other EU countries. Aim of the work is to analyse the progressive aging in Italy in terms both of the increasing weight of older people on youngers, and of other dimensions related to the aging of the population (as the labour force market, the households' structures, marriage behaviours, etc.), focusing the Italian demographic dynamics in a multidimensional approach. The hypothesis is that the aging of the population will inevitably affect the future quality of life in Italy. Anticipation practice need robust scenarios based on official data. For this reason, the work uses the Equitable and Sustainable Well-being data jointly with mortality statistics and other demographic indicators, analysed by the Dynamic Factor Analysis method.

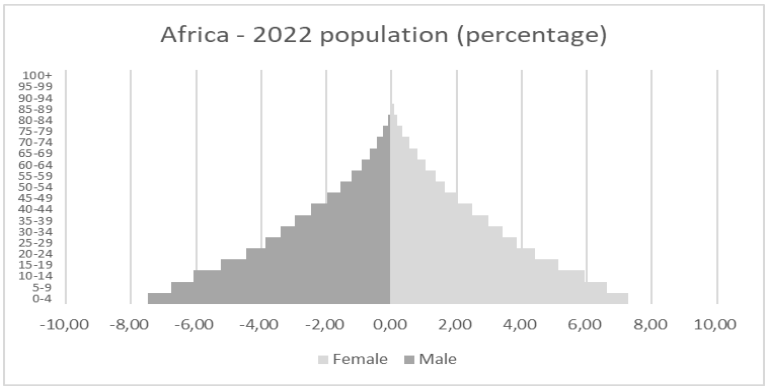
1. Age pyramids and social complexity

In 2022[2] the number of new-born children in the world, according to United Nations data, significantly decreased respect to past. People aged 0-4 are in a smaller percentage than the immediately higher age group (5-9 years). This is true even in countries where the fertility rate is still high. For centuries, everywhere in the world, the composition by age group has been similar to that of the African continent in 2022 (Figure 1). A significant disproportion between new-borns and the adult population can put the future dynamics of a country at risk. This is the case of Italy.

[1] The work is the result of the overall contribution of the authors. However, paragraphs 1, 2, 4 are attributable to Carolina Facioni, while paragraph 3 is attributable to Isabella Corazziari. The views expressed in the paper do not necessarily reflect those of the authors' affiliation institution.

[2] The age pyramids in the paper are the result of elaborations of the year 2022 revised data provided by United Nations, Department of Economic and Social Affairs, Population Division.

Figure 1 – *African population by age and sex: year 2022. (Per 100 individuals).*



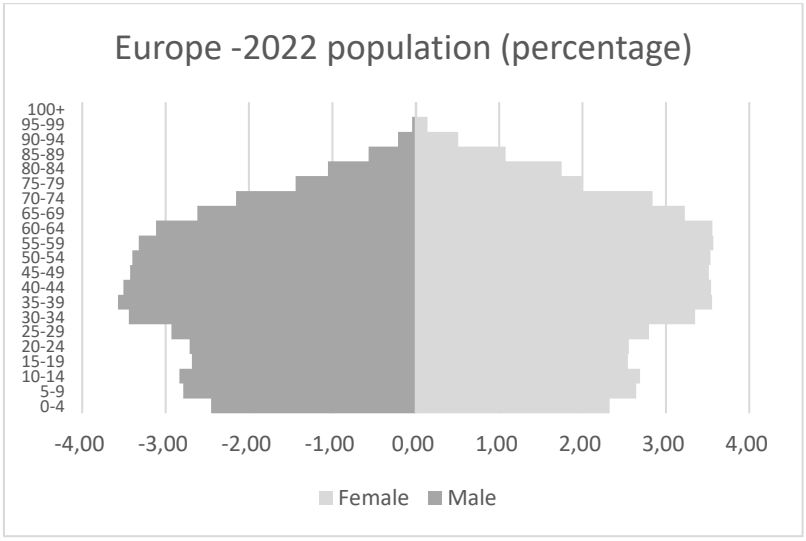
Source: United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2022 Revision. (Medium variant).

Population aging is the result of the combined effect of two distinct factors: a high life expectancy and a low fertility rate. These two elements are the result of two demographic transitions: the one that slowed down mortality and allowed people to live longer (the older of the two) and the one that reduced births (it started about in the 60s of the last century). Both of these factors were the result of civilizational achievements.

Furthermore, life expectancy increases not only due to scientific achievements, but also because of both a medical care available to everyone, and a widespread healthy lifestyle. In synthesis, a good life expectancy is due to a widespread good quality of life. On the other side, the possibility of birth control makes parenthood the result of a responsible decision. People, especially young people, can face it at the most appropriate moment. So, also birth control is concerning the quality-of-life topic.

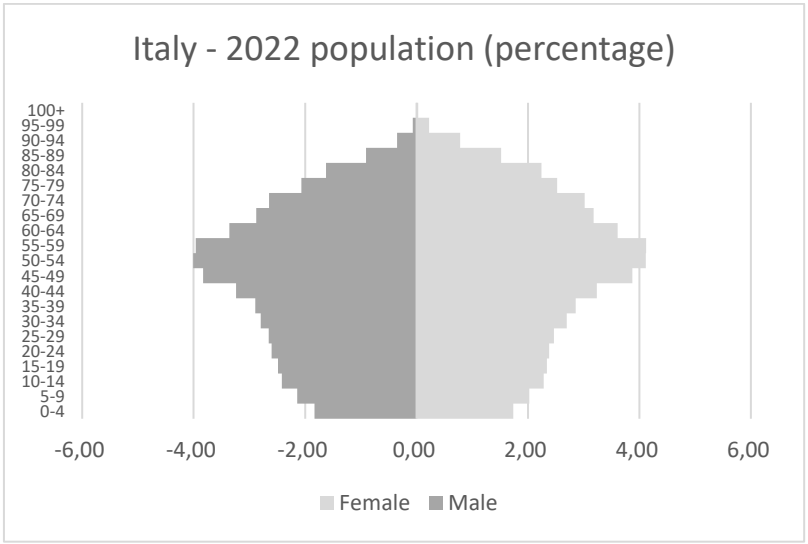
It is worth focusing on the social aspects related to the "right time" to have children in Italian context. In fact, from this specific point of view, Italy represents a sort of borderline case, due to a historic impasse that will be maybe extremely difficult to resolve (Golini, 1997, 2001). In 2022, the demographic crisis in Italy, respect to European context, is evident by comparing Figures 2 and 3.

Figure 2 – *European population by age and sex: year 2022. (Per 100 individuals).*



Source: United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2022 Revision. (Medium variant).

Figure 3 – *Italian population by age and sex: year 2022. (Per 100 individuals).*



Source: United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2022 Revision. (Medium variant).

The age pyramids make it clear that Europe's population is aging as a whole, but Italy is aging more drastically and faster respect to the European trend. In Italy, it is faster respect to other secular, prosperous and industrialized European countries[3] like Germany[4]. What future can we think of for the Italian population? Considering the UN data for 2022, the total Italian population is 59,037,474 people; the largest group is made up of people aged 50-59 (9,566,566 individuals); children in the 0-4 age range are instead only 2,106,488. Italian official data say that in 2022 the average number of children per woman is 1.24, while life expectancy at birth is 82.6 years (Istat, 2023a). The median scenario referring to 2040 (Istat, 2021b, 2022c) shows an increase in people living alone, as well as in couples without children. Can we imagine the quality of life in a country that in twenty years will be made up mainly of elderly people? It is evident that such a particular situation requires immediate attention and policy measures. Keeping a good level of quality of life (Maggino, 2022) in our country requires very serious anticipation (Poli, 2017) work. A work that can rely on reliable data, such as the official data available in Italy.

2. Being young in Italy: the difficulty of reaching adulthood

Having one of the highest life expectancies in the world is a success a nation must be proud of, it is necessary to understand the reasons why the fertility rate in Italy is one of the lowest in the world. The historic research of IARD (IARD, 1988, 1993, 1997, 2003, 2007) can help. IARD identified five key steps in young people's lives for their attainment of a life as an adult. These were the steps identified: 1) having completed the education path; 2) having found a job; 3) having left the parental home; 4) forming a stable relationship; 5) (eventually) having children. Getting the n.5 goal is probably conditioned by the previous four. However, these are undoubtedly five fundamental steps in life - and they are of particular interest in this context, as the official data available to us can explain much of why in Italy there fewer and fewer children over the years are. Table 1 illustrates the different situation between Italy and EU-27 in years 2008, 2014, and 2020 (Istat, 2022c).

[3] Economic prosperity and an industrial and/or tertiary type economy are usually the socio-cultural terrain where the processes of secularization (characterized by phenomena such as the increase in divorces and the decrease in births) develop (Impicciatore and Billari, 2012)

[4] The choice of Germany as a term of comparison is due to the need to highlight the Italian disadvantage in Europe. It would not have been clear enough if the authors had chosen a country like, e.g., Sweden or Denmark - whose welfare aimed at aimed at the quality of life of young people is well known. Furthermore, the authors did not choose a European country already successfully implementing policies for the recovery of the birth rate, such as France. A useful website for comparing all age pyramids using UN data is https://www.populationpyramid.net/.

Table 1 – *People aged 30-34 who graduated and people aged 18-24 who left their studies early. Comparison Italy - EU-27. Years 2008, 2014 and 2020 (Per 100 individuals with the same characteristics).*

Education	Italy			EU-27
	2008	2014	2020	2020
People aged 30-34 with university degree	19.2	23.9	27.8	41.0
People aged 18-24 who early left the education and training system	19.6	15.0	13.1	9.9

Source of data: Istat Report on education levels in Italy (published on October 25th 2022)

Italy still lags behind the European average. Furthermore, school dropouts increased during the Covid-19 pandemic (Istat, 2022a, 2022b). Considering master's degrees, the average age at graduation in Italy is 27.1 in 2022 (AlmaLaurea, 2023). A separate chapter deserves the phenomenon of NEETs (Alfieri *et al*., 2015). In 2022, the data on the total number of NEETs aged 15-29 is equal to 19.0% of people in the same age class - a better percentage respect to year 2020 (23.7%), and better than the 23.1% of year 2021 (Istat, 2023b). Although the data show an improvement, the percentage of NEETs in Italy is still higher than the EU average (11.7%) in 2022[5]. With regard to the labour market, according to the latest Istat Annual Report (Istat, 2023c), Italy is the European country with the oldest workforce. The share of 15–34-year-olds in the total labour force aged 15-64 decreased even more significantly respect to the reduction observed for the population, while it was stronger the increase in the weight of the 50-64 age group. A very complex situation (especially with respect to female employment), on which policies should focus, especially in light of the possible opportunities offered by the NextGenerationEU (MEF, 2022). Even considering the abandonment of the parental home, young Italians are in a different condition than the European average. In the EU, on average, boys left their parental household at the age of 27.4 and girls at 25.5 in 2021. Italian data indicate an average of 29.9 years, which, however, rises to 30.9 for males. Women leave home shortly before with an average of 29.8 years. There is a reasonable correlation

[5]About EU statistics on NEETs, see: https://ec.europa.eu/eurostat/statistics-explainde/index.php?title=Statistics-explained/index.php?title=Statistics_on_young_people_neither_in_employment_nor_in_education_or_training

between the step of abandonment of parental home and formation of a new couple, due to a traditional marriage or a more uxorio cohabitation. In 2019, there were 184,088 weddings celebrated in Italy. The percentage of marriages is declining, especially for the first weddings. Cohabitations more uxorio quadrupled from 1998-1999 to 2018-2019, passing from about 340 thousand to 1 million 370 thousand. The average age at first marriage is 33.9 for men, and 31.7 for women (Istat, 2021a). These last data are crucial for our analysis. If none the circumstances that should help young people form a couple (and therefore have children) is conducive to this aim (Ambrosi and Rosina, 2009; Del Boca and Rosina, 2009), a couple forms relatively late and have the first child late (Golini and Lo Prete, 2019). In Italy, the average age of women having their first child was 31.3 in 2019, while the average age in Europe was 29.4. Having the first child late means having subsequent difficulties in having others. Italy becomes an increasingly old country. The Dynamic Factor Analysis with which the indicators will be analysed below has, in the intentions of the authors, the advantage of enriching the analysis of demographic aging with a temporal element, thanks to which the evolution of the multivariate phenomenon described by some selected well-being indicators is much clearer.

3. Materials, methods, and the results

In order to address the changes of the above-discussed scenario over the last years, a descriptive multiway analysis (Dynamic Factor Analysis DFA) has been performed considering the demographic and labour market indicators available for the Italian regions (NUTS2[6]) in years 2004-2022, from ISTAT data warehouse[7]

The Dynamic Factor Analysis (hereinafter DFA), proposed firstly in 1970s and improved since 1990s (Coppi & Zannella, 1979; Corazziari, 1999; Facioni, 2019) models multivariate data collected over time[8]. It consists in the joint application of factorial methods (Principal Component Analysis) and time regression models to specific covariance matrices addressing particular source of data variation. No inferential assumptions are included about the process generating data, the method aims to describe data variability. The three sources of variation refer respectively to the average over time relationships between Italian regions and indicators (*static*), the average interaction between the indicators and time (*dynamic of centers*), the Italian regions dynamic over time considered as *differential* or net compared to the indicators' average dynamic. The *differential dynamic* can strengthen the mean

[6] https://ec.europa.eu/eurostat/web/nuts/background

[7] Descriptions of the indicators and corresponding source can be asked to the authors. Details about data processing and imputation for missing data/periods can be asked to the authors.

[8] DFA works with complete data matrices. Missing data have to be imputed before performing DFA.

indicators' trend over time or move in other directions, weakening or even contrasting the overall dynamics.

In this work, we use the first DFA model[9], which focuses on the variability of the indicators and their interaction with the other two dimensions. In the first DFA model the *dynamic of centers* is described by ordinary least squares time regression models (OLS), the *static* variability by a PCA of the corresponding covariance matrix and on the same PCA space the unit trajectories over time allow evaluating the *differential dynamic*. Trajectories can be clustered according to their closeness and shape using Cluster Analysis on specific distance matrices (Blanco et al. 1999). If one region or cluster moves toward the centre of the PCs' space (which characterizes the overall dynamic of the system of data) homogeneity among Italian regions is increasing. By contrast, if the unit or cluster move away from the centre of the axes, its differential dynamic is to be interpreted accordingly to its direction. Finally indexes of the goodness of fit of each source of variation are also provided as the ratio between the modelled total variability (trace of the covariance matrix) and the corresponding observed one.

The following table shows the goodness of fit indices related to each source of variation analyzed in the first DFA model.

Table 2 – *Index of goodness of fit.*

Quality index	Description	Value
Overall index	overall covariance matrix	0.75
Regression analysis	dynamic of centers	0.66
Factorial analysis (PCA)		
	static	0.87
	differential dynamic over time of Italian Regions	0.53

The DFA describes the 75% of the overall variability, the factorial analysis account for the 87% of the static variability, while the regression describes the 66% of the centres dynamic. Indices of fit for each year can also be provided[10].

Considering regression results, few variables show nonlinear variability over time or are constant; the others present good regression results[11].

Main results are that the birth rate is linearly decreasing, the population is slightly ageing and life expectancies at 0 and 65 years for both males and females increasing. Consequently, Elders' burden on Youngers and on labour force is linearly increasing. Marriage rates are decreasing over time, with the exception of second or

[9] The DFA consists of four models, each of which defines a specific strategy to address the three sources of variation (Corazziari 1999, Facioni et al 2019).

[10] Values can be asked to the authors

[11] Details can be asked to the authors.

more marriages of older people realistically less interested in reproductive behaviours. The Labour Force (LF) is ageing as over 55 are increasing and the proportion of 15-34 on the total LF decreasing. Couples with both partners, or only the woman employed are slightly increasing, so as the proportion of couples with no partner employed and no employment pension. Within the described ageing Labour Force market, the proportion of NEET young people, not studying and not looking for a job, is increasing for both males and females. Also, the proportion of young people aged 30-34 with at least a university degree increased since 2004, while decreased those of young people aged 18-24 who left the studies. In the same period, the proportion of people still living at home with parents especially if looking for a job or students increased in the group of people aged 18-34.

The PCA of the *static* variability provided two principal components (PCs). Following a circle counterclockwise on the bottom, we start from a situation characterized by more elderly people, graduates, women with little children possibly employed, couples with both partners employed. Then, we found higher birth rates and number of children by woman. On the right, we find higher proportion of young people on the total LF and consequently higher rates of Youngers who left the study early, higher marriage rates for Youngers and first marriages.

Then worst situations follow higher proportion of couples without a job nor an employment pension or only the man employed, higher NEET rates and young people (18-34) still living with parents looking for a job. Finally higher proportions of workers over 50 on the total LF and students (18-34) still living with parents.

Projecting the mean of the Italian regions over time, the PCA's plane contrasts the Northern on the left of the plane versus the Southern ones on the right, with the former characterized by better employment opportunities with respect to the latter representing a more traditional society with more marriages and the man providing economic resources for the family[12]. A hierarchical cluster analysis of the regions' time trajectories to describe the regions' differential dynamic over time provided a partition of 4 clusters, the first two involving respectively Northern and Central regions (Liguria and Abruzzo the exceptions) and the last two clusters splitting Southern regions and Islands.

[12] Details can be asked to the authors.

Table 5 – *Cluster analysis of trajectories: size of each cluster.*

Cluster	Cluster size	NUTS2 regions
1	8	Piemonte, V.d'Aosta, Lombardia, Bolzano, Trento, Veneto, Friuli VG, Emilia-Rom.
2	6	Liguria, Toscana, Umbria, Marche, Lazio, Abruzzo
3	3	Molise, Basilicata, Sardegna
4	4	Campania, Puglia, Calabria, Sicilia

The second and third clusters show low dynamic over time. The Cluster 2 dynamic shows an improvement for the employment rates and an ageing of the population. Cluster 3, referring to Sardinia, Basilicata and Molise, after a slight worsening for Younger's opportunities seems to reduce distances from, he others.

The first and fourth cluster, show a high dynamic over time, but in opposite directions and consequences. Cluster 1, referring to the Northern regions except Liguria, increases employment rates in years 2009-2019 for people aged more than 25 and especially for males. Finally, the cluster 4 shows a worsening dynamic, moving away from the centre towards the worst situation for Youngers in terms of labour opportunities and of trust in the future (NEET young people).

4. Conclusions

This work attempted to provide a multidimensional - albeit synthetic and necessarily partial - vision of the problem, enriching the picture with a temporal dynamic of the phenomenon, and outlining the picture of a (present, and maybe future) crisis. Anyway, this data can be precious for future policies if put in connection with other indicators: for example, with the labour market's indicators, with the school system's indicators, with other aspects linked to satisfaction (e.g., the propensity to marriage among young people), etcetera. Suffice it to consider, among the many results that have emerged, how young Italians can no longer consider (as happened in the past) the possession of a high qualification as a sure springboard for a better and more qualified job. In Italy, the lack of work affects family life plans: this affects lower marriage rates, and the average age of first child, which is increasing. One of the effects of this complex situation is the collapse of births in Italy, with the consequent aging due to the combined very high life expectancy. The analysis of data tell us about the crisis of many fundamental aspects for the quality of life of Italian citizens. The results provide a picture that illustrates a progressive worsening of some dimensions, which decision makers must put a stop to in time. It needs policies related to welfare state (Baldacci and Lugaresi, 1996). The crisis was predictable. Eleonora Barbieri Masini, theorist and founder of Futures

Studies, did it in the 1970s (Facioni, 2019). Italy has lost years in which it could have done a serious job of anticipation by preventing the crisis. In 2022, Gian Carlo Blangiardo (president of Istat at the time), underlined how a "demographic winter" could have consequences on GDP in the future, but a possible collapse could be not only of an economic kind, but related to overall well-being. An aged nation is destined too no longer be competitive. It is necessary to improve the lives of young people, to allow them to enter a dignified adult life. For example, giving them a dignified and guaranteed work, in particular to women. Policies that support young couples' lives should be high on the public agenda. From a completely different point of view, policies should review the concept of "elderly" in Italy. The hypothesis that, in the next few years, a small group of young workers will maintain an army of retired people (born in the years of the economic boom) is not tenable. Italy needs a change in terms of demographic rejuvenation. It affects everyone's quality of life in the coming years. Italy needs to anticipate its own future.

References

ALMALAUREA 2023, XXV Rapporto Almalaurea sul Profilo e sulla Condizione Occupazionale dei Laureati. On: https://www.almalaurea.it (June 12th 2023).

ALFIERI, S., SIRONI, E., MARTA, E., ROSINA, A., MARZANA, D. 2015. Young Italian NEETs (Not in Employment, Education, or Training) and the influence of their family background. *Europe's Journal of Psychology*, Vol.11 No.2, 311.

AMBROSI E., ROSINA A. 2009, *Non è un paese per giovani. L'anomalia italiana: una generazione senza voce*, Venice: Marsilio..

BALDACCI, E., LUGARESI, S. 1996. Assessing the impact of demographic ageing on the welfare state in Italy. *Statistical Journal of the United Nations Economic Commission for Europe*, Vol. 13 No. 3, pp.255-273.

BLANGIARDO G.C., 2022, *Interview given to the newspaper "Il Sole 24 Ore" on November 19th, 2022*. On: https://www.ilsole24ore.com/art/blangiardo-istat-l-inverno-demografico-brucera-terzo-pil-AEet0UIC.

COPPI R., ZANNELLA F. 1979. L'analisi fattoriale di una serie temporale multipla relativa allo stesso insieme di unità statistiche, Atti della XXIX Riunione della SIS.

CORAZZIARI, I. 1999. *Dynamic Factor Analysis*. In: Vichi, M., Opitz, O. (Eds) *Classification and Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-60126-2_22.

DEL BOCA D., ROSINA A. 2009, *Famiglie sole. Sopravvivere con un welfare inefficiente*, Bologna: Il Mulino.

EUROSTAT, 2022, *Population projections by country*, on: https://ec.europa.eu/eurostat/statistics-explained/index.php?oldid=497115#Population_projections_by_country (November 23th, 2022).

FACIONI C., 2016. Anatomia dell'incertezza. Il futuro nella voce dei demografi italiani, *Futuri*, No. 7, pp. 6-37.

FACIONI C., 2019. Per una sociologia dei futuri: il contributo di Eleonora Barbieri Masini alla fondazione dei futures studies, *Futuri*, No. 11, pp. 65-86.

FACIONI C., CORAZZIARI I., MAGGINO F., 2019. Measuring uncertainties: a theoretical approach, *International Journal of Computational Economics and Econometrics*, vol. 9 (1/2), pp. 5-28, DOI: 10.1504/IJCEE.2019.097797.

FACIONI C., 2022. Scenari demografici per l'Italia: le criticità che attendono un paese che invecchia, *Futuri*, No. 17, pp. 13-22.

GOLINI, A., 1997. Demographic trends and ageing in Europe. Prospects, problems and policies. *Genus*, 33-74.

GOLINI A., LO PRETE M.V.2019, *Italiani poca gente. Il paese ai tempi del malessere demografico*, Rome: Luiss University Press

GOLINI, A., 2001. Demographic trends and population policies. Futures, Vol.33, No.1, pp.27-41.

IARD, *Rapporto sulla condizione giovanile* – Anni 1987, 1988, 1993, 1997, 2002, 2007. Bologna: Il Mulino.

IMPICCIATORE R, BILLARI F.C. 2012, Secularization, Union Formation Practices, and Marital Stability: Evidence from Italy. *Eur J Popul*. 2012 May; Vol.28, No.2, pp.119-138. Doi: 10.1007/s10680-012-9255-4.

ISTAT, 2021a, Statistiche Report – Matrimoni, unioni civili, separazioni e divorzi – Anno 2019. On https://www.ISTAT.it (February 18th 2021).

ISTAT, 2021b, *Statistiche Report – Households and population projection. Base 1/1/2020*. On https://www.ISTAT.it (November 26th 2021).

ISTAT, 2022a, 2023b *Rapporto BES 2021 e 2022: il benessere equo e sostenibile in Italia* on: https://www.ISTAT.it.

ISTAT, 2022b, *Statistiche Report – Livelli di istruzione e ritorni occupazionali*. Anno 2021 (October 25th, 2022).

ISTAT, 2022c, *Statistiche Report – Previsioni della popolazione residente e delle famiglie. Base 1/1/2021*, on: https://www.ISTAT.it (September 22th, 2022).

ISTAT, 2023a, *Statistiche Report – Indicatori demografici – Anno 2022*, on https://www.ISTAT.it (April 7th, 2023).

ISTAT, 2023c, Rapporto Annuale 2023. La situazione del Paese on: https://www.istat.it (July 7th, 2023).

MAGGINO F. 2022, *Il Regno della Qualità: Propositi, concetti e strumenti per un nuovo Umanesimo*, Edizioni Accademiche Italiane.

MEF, 2022, La condizione dei giovani in Italia e il potenziale contributo del Piano Nazionale di Ripresa e Resilienza per migliorarla. On: https://www.rgs.mef.gov.it/ (published on March 7th 2022).

POLI R. 2017, *Introduction to Anticipation Studies*, Berlin, Heidelberg: Springer.

REYNAUD C., MICCOLI, S., 2019. Population ageing in Italy after the 2008 economic crisis: A demographic approach. *Futures*, Vol.105, pp.17-26.

ROSINA A., Sorgi S., 2016, *Il futuro che (non) c'è. Costruire un domani migliore con la demografia*, Milan: Università Bocconi Editore.

ROSINA A. 2021, *Crisi demografica, Politiche per un paese che ha smesso di crescere*, Rome: Carocci.

ROSINA A., Impicciatore R. 2022, *Storia demografica d'Italia. Crescita, crisi e sfide*, Rome: Carocci.

UNITED NATIONS, Department of Economic and Social Affairs, Population Division, 2022, *World Population Prospects: The 2022 Revision*. On: https://population.un.org/wpp/.

WORLD HEALTH ORGANIZATION, (2020), *Life table by countries*.
On: https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-ghe-life-tables-by.

Carolina FACIONI, Istat, facioni@istat.it
Isabella CORAZZIARI, Istat, corazzia@istat.it

SKILLS AND WAGE DIFFERENCES BETWEEN NATIVE AND FOREIGN WORKERS IN ITALY

Camilla Pangallo, Oliviero Casacchia, Corrado Polli

Abstract. For several years, immigration has been one of the topics at the centre of the Italian media and political debate and has become a structural component of society and the economy. This paper aims to make a contribution to analysing the wage differences between Italians and foreigners in the Italian labour market in terms of the content of occupations. Several studies in the literature suggest that natives respond to immigration by specialising in occupations in which they have a comparative advantage over foreigners. Through the construction of a series of composite indicators, representing the content of occupations, the wage gap between Italian and foreign workers was analysed in order to identify the factors that most determine the observed gap. In this analysis, the Blinder-Oaxaca Decomposition was used as a statistical model and the study was conducted through statistical matching between the Labour Force Survey (research file, year 2019) of ISTAT and the Sample Survey on Occupations (year 2013) carried out jointly by ISTAT and INAPP.

1. Background and Objective

Over the last decade, the presence of foreigners in Italy has increased significantly. In fact, the number of legally resident foreigners has risen from 1.3 million in 2001 to about 5.5 million as of 1 January 2023 and their share in the total number of residents has increased from 2.3 per cent to 8.6 per cent (ISTAT, 2023). Moreover, the structure of the foreign population is particularly complex and rapidly changing, and the composition of foreigners is also very heterogeneous. Today, legally present foreign citizens are also an established reality in Italy, although to a lesser extent than in many other European countries. An important aspect for the realisation of the migration project is the wage received by foreign workers as well as their social and economic integration in the host country. The objective of this study is to identify the factors that affect workers' wages in Italy and to observe how the possession of non-Italian citizenship affects wages. Furthermore, the aim is also to identify some determinants associated with the wage differences found between

Italians and foreigners in terms of manual, communicative and quantitative content of occupations, as suggested by Peri and Sparber (2009). Specifically, *manual* content refers to those occupations in which the worker employs greater physical effort in carrying out work activities; while *communicative* content refers to that type of occupation in which the worker uses purely cognitive skills connected to the sphere of communication and, finally, *quantitative* content concerns that type of occupation in which the worker is skilled in mathematical reasoning and is familiar with numbers. Several studies in the literature have analysed the impact of immigration on native-born employees and wages. At the international level, and particularly in the US context, some authors have shown that foreign-born workers specialise in physical-manual labour-intensive occupations, while natives perform language- and communication-intensive jobs; they also show that less-educated foreign workers specialise in different productive occupations than natives, at the same level of education, due to their different skills. Thus, immigration reallocates the labour supply of natives, reducing downward wage pressure (Peri, Sparber, 2009). Among the studies in the literature analysing the dynamics in the Italian territory, an interesting contribution has been provided by Strozza, Inglese and Baldacci (1999), who analysed the determinants of immigrants' wages in some areas of Lazio and Campania, where the share of clandestine immigration is relatively high, using data from the period 1993-1994. According to the empirical results, income differences appear to be relatively high among immigrants; differences in labour market integration between genders and areas of origin also emerge. Finally, legal status plays an important role in explaining the wage gap between regular and non-regular immigrants, also due to the different structure of the labour sector in the two groups. The existence of a wide wage gap between immigrants and natives in Italy has also been highlighted by, among others, Coppola, Di Laurea and Gerosa (2013), who attempted to explain it using the characteristics observed in both populations. More recently, Pieroni, D'Agostino, and Lanari (2019) pointed out that, in Italy, language problems among foreigners reduce the employment rate by about 30 per cent, with even larger point estimates when assessing employment discrimination. The authors suggest that knowledge of the Italian language also significantly affects immigrants' wages.

2. Data and Methods

The data used for the analysis are from the Survey on Italian Occupations (carried out jointly by ISTAT and INAPP) with reference year 2013 and from the 2019 data of the Labour Force Survey carried out by ISTAT. The database was created by integrating the data in the Labour Force Survey with those in the Survey of Italian

Occupations. It was possible to create this integrated database by means of a variable present in both databases, namely the 4-digit classification of occupations. Since there is no data source that collects both specific information on the characteristics and requirements of the worker and the occupation and information on the labour market and the socio-demographic data of the workers, the wealth of information provided by the Survey on Italian Occupations and the Continuous Labour Force Survey was exploited. In addition, composite indicators were constructed to assess the occupations in terms of attitudes, skills and activities performed. We used the methodology proposed in Peri and Sparber's work (2009), in particular we analysed three types of indicators expressing the content of dexterity, communicativeness and the quantitative aspect of the profession. These composite indicators represent a weighted average value for each profession, and they were obtained using the information contained in the Survey on Italian Occupations database. In particular, the *manual* indicator was constructed using nineteen skills that cover information on dexterity of limbs, hands and fingers, body coordination and flexibility, speed and strength. The *communicative*[1] indicator was constructed using data on four cognitive skills such as comprehension or self-expression in oral and written form. Finally, the *quantitative* indicator summarises two pieces of information concerning the ability to perform mathematical calculations or to apply the correct mathematical methods to solve a problem.

To analyse the reasons for the wage gap between Italian and foreign[2] workers, we applied the Blinder-Oaxaca Decomposition (Oaxaca, 1973; Blinder, 1973), both the threefold and twofold decomposition, to the data. This technique is widely used to study average differences in outcome between groups. In the literature, this methodology is often used to analyse wage gaps by sex or race.

According to the *threefold decomposition*, the difference in outcome can be obtained as the sum of the following three components:

- the "Endowment Effect" which is equal to the part of the differential that is due to group differences in predictors;
- the "Coefficient Effect" which measures the contribution of differences in coefficients (including differences in the intercept);
- the "Interaction term" which takes into account the fact that between-group differences in explanatory variables and coefficients occur simultaneously.

An alternative decomposition, called the *twofold decomposition*, which is very popular in the discrimination literature, is based on the notion that there is a vector

[1] For example, four cognitive abilities belonging to section D of the ICP questionnaire, from question number 1 to 4, were selected for the construction of the *communicative* indicator. The abilities used are as follows: Understanding communication in oral form, Understanding communication in written form, Expressing oneself in oral form, Expressing oneself in written form.

[2] Employed persons with foreign citizenship resident in Italy also include EU citizens.

of non-discriminatory coefficients that should be used to determine the contribution of differences in predictors. This is the vector of regression coefficients that would emerge in a world without discrimination in the labour market.

Thus, this decomposition divides the difference in mean results into two parts:

- "explained" is the component explained by the differences between the groups in the explanatory variables.
- "unexplained" is the remaining part that is not explained by differences in the explanatory variables. It is often attributed to discrimination but can also result from the influence of unobserved variables.

The unexplained part can be further broken down into two sub-parts representing discrimination in favour of the first group and against the second group, respectively.

In this study, in order to apply the Oaxaca Decomposition, we used the last average monthly salary as the dependent variable and as independent variables: the manual, communicative and quantitative content of the occupations, some socio-demographic characteristics of the workers (gender, age and educational qualification) and some occupational characteristics (fixed-term or permanent employment, full-time or part-time[3], size of the company and economic sector). The categorical variables were transformed into dummy variables and a chosen category was omitted from the model for each categorical variable. The three realised composite indicators representing the manual, communicative and quantitative content of the occupations were standardised with their maximum value[4].

3. Results

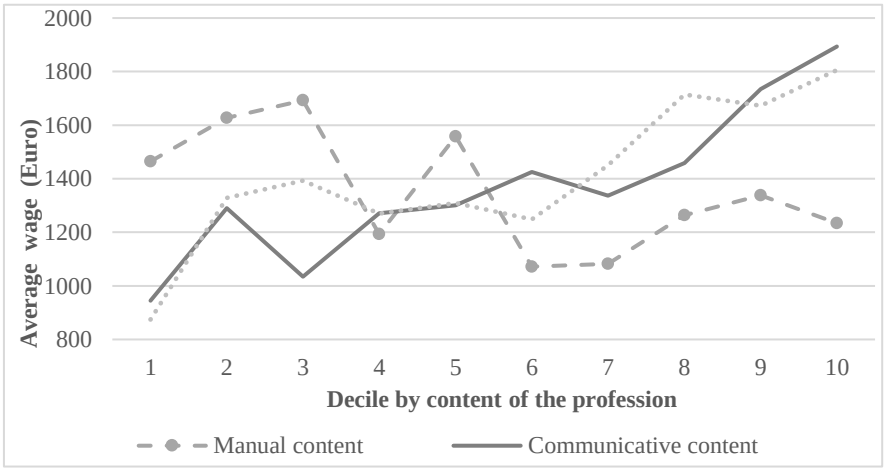
The analysis was developed from the observation of the relationship between the average wage of the employed and the content of the occupations. Figure 1 shows the curve of average wages for the total employed population as the intensity of the manual, communicative and quantitative content of the occupation changes. Wages clearly tend to decrease as the intensity of the manual content of the work performed increases and, on the contrary, to increase as the intensity of the communicative and quantitative content of the occupations increases. For example, a worker in a profession with the least communication content earns an average of 950 euros, while a worker in a profession with the most communication content earns an

[3] The data on the type of part-time or full-time work were already available in the Labour Force Survey dataset, the name of the variable taken into consideration is “PIEPAR”, it is a variable reconstructed by Istat. Therefore, wages were not converted into full-time equivalents, but this PIEPAR variable was used in the model, which distinguishes full-time from part-time work.

[4] The maximum value of the manual indicator is 64, the maximum value of the communicative indicator is 89 and the maximum value of the quantitative indicator is 85.

average of 1,900 euros. The trend of the three curves is not so linear but reflects in general terms what has just been stated.

Figure 1 - *Average wage of workers by manual, communicative and quantitative content decile in the occupation. Year 2019 (RCFL) and 2013 (ICP).*



Source: Own elaboration on ISTAT-INAPP data.

From this consideration, we wanted to investigate the topic further by applying the Blinder-Oaxaca decomposition in order to understand, above all, how much of the wage gap between Italian and foreign workers in Italy in 2019 can be attributed to working in a profession with a higher manual, communicative or quantitative content. All analyses were only conducted on employed persons due to data availability. As shown in Table 1, the sample consists of 152,115 employed persons (n), of whom 134,306 are natives (n_A) and 17,809 are foreigners (n_B).

Table 1 - *Size of groups.*

Employed Italian Citizenship (n_A)	Employed Foreign Citizenship (n_B)	Employed Pooled Regression (n)
134,306	17,809	152,115

Source: Own elaboration on ISTAT-INAPP data.

The first results (Table 2) show that the average monthly wage[5] is 1,386.27 euros for Italians (group A) while that of foreign workers (group B) is 1,076.65 euros. So,

[5] No hourly wage data are available to apply this type of analysis. Monthly wage values were estimated by extracting a sample of employed persons with a pooled regression.

the wage difference is about 309.63 euros and it is this gap that we tried to explain with the Blinder-Oaxaca Decomposition.

Table 2 – *The average wage values by groups.*

Average Wage of Italians (Y_A)	Average Wage of Foreigners (Y_B)	Wage Difference (Y_A-Y_B)
1,386.27 euro	1,076.65 euro	309.63 euro

Source: Own elaboration on ISTAT-INAPP data.

First, we examined the results of the threefold Blinder-Oaxaca decomposition, which suggest that the 309.63 euros wage difference can be explained by three different effects (Table 3). The endowment effect represents that part of the differential due to the average differences in the composition of the two groups observed in the predictors (age, gender, education, economic sector of employment) and is about 234.55 euros. The coefficients effect is that part of the differential due to the differences between the two groups in the coefficients, generally attributed to the discrimination present in the labour market or to potential effects in unobserved variables; this effect represents 49.83 euros and finally the interaction effect is that term which takes into account the differences in the two groups in both endowments and coefficients simultaneously.

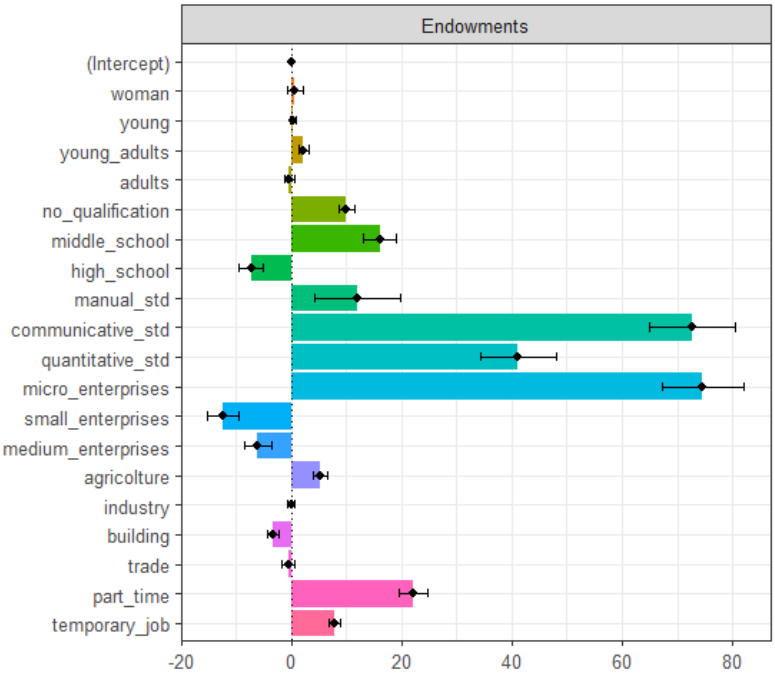
Table 3 – *Threefold Decomposition results.*

coef(endowments)	coef(coefficients)	coef(interaction)
234.55 euro	25.24 euro	49.83 euro

Source: Own elaboration on ISTAT-INAPP data.

Next, we examined the components of the endowments and coefficients of the triple decomposition, variable by variable. Figure 2 shows the estimation results for each variable of the threefold decomposition, together with the error bars indicating the 95% confidence intervals. In the endowment's component, the most statistically significant variables are: the communicative and quantitative content of occupations and working in micro enterprises. It thus appears that a significant part of the wage gap between natives and foreigners is determined by group differences in performing a more communicative and quantitative occupation and in the proportion of individuals working in micro-enterprises.

Figure 2 - *Endowment components of a threefold Blinder-Oaxaca decomposition to explain the wage gap between natives and immigrants.*



Source: Own elaboration on ISTAT-INAPP data.

Table 4 – *The estimates of the terms constituting the endowments effect for the three most significant variables.*

	β_B	$E(X_A) - E(X_B)$	*Endowment effect:* $\beta_B \{E(X_A) - E(X_B)\}$'
communicative_std	502.547	0.145	72.869
micro_enterprises	-281.239	-0.265	74.528
quantitative_std	320.721	0.128	41.052

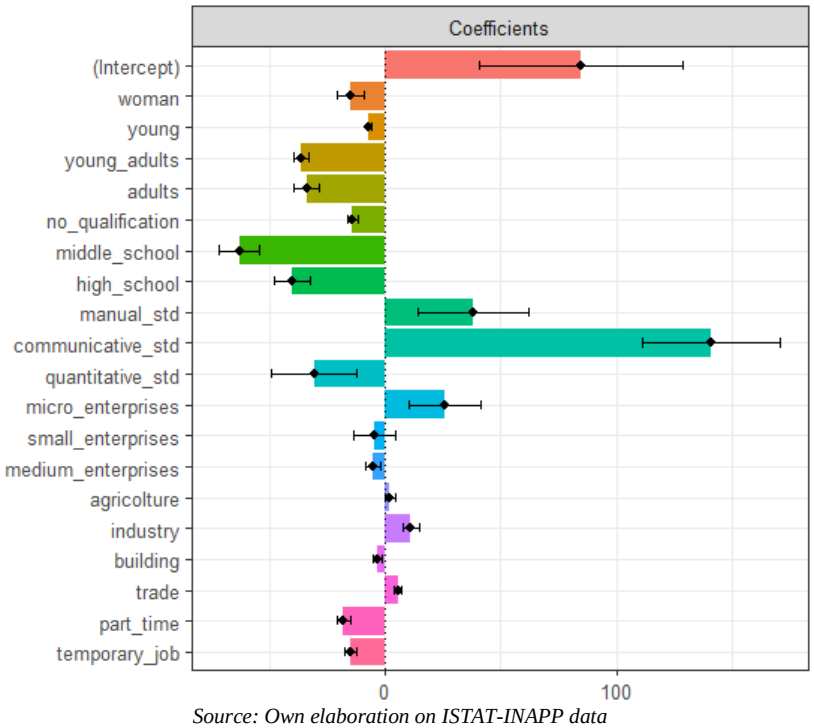
Source: Own elaboration on ISTAT-INAPP data.

As can be seen in Table 4, work with high communication content contributes strongly to the wage difference between the two groups. In fact, natives work in occupations with a higher communication content (0.145) than foreigners and at the same time the coefficient associated with communication occupations is strongly positive (502.547), indicating that among foreigners, those in occupations with a higher communication content tend to earn more. Thus, the product between the two explains the strong contribution of this variable in explaining the wage differences

resulting from the different composition between the groups. The variable associated with workers in micro-enterprises explains another large part of the wage differences between natives and foreigners. In fact, the coefficient of the regression indicates that foreign workers employed in micro enterprises earn less than foreign workers employed in large enterprises (-281.239). A reading of the differences in the endowment shows that foreigners work more in micro-enterprises (26.5%) than Italians. Therefore, the product between the two factors explains a substantial part of the wage differences. Work with a high quantitative content also helps to explain part of the wage difference between the two groups. Italians work in occupations with a higher quantitative content (0.128) than foreigners and at the same time the coefficient associated with quantitative occupations is positive (320.721). This indicates that foreigners employed in a purely quantitative activity receive a higher salary. Thus, the product between the two explains the strong contribution of this variable in explaining the wage differences resulting from the different composition between the groups.

Similarly, some variables are significant in the coefficient component (Figure 3).

Figure 3 - *Components of the coefficients of a threefold Blinder-Oaxaca decomposition to explain the wage gap between natives and immigrants.*



Source: Own elaboration on ISTAT-INAPP data

Table 5 - *Estimates of the terms constituting the coefficient effect for the three most significant variables.*

	$E(X_B)$	$\beta_A - \beta_B$	*Coefficient effect:* $E(X_B)(\beta_A - \beta_B)$
communicative_std	0.496	284.274	150.000
middle_school	0.430	-146.449	-62.973
young_adults	0.208	-173.518	- 36.092

Source: Own elaboration on ISTAT-INAPP data.

The most significant variables are, for example, the communicative content of professions, possession of a secondary school leaving certificate or being aged between 25 and 34.

As can be seen in Table 5 (*the coefficient effect*), the differences in the regression coefficients on the communicative aspect of occupations account for the decisive part of the wage gap, where the wage gain in performing a more communicative occupation is 284 euros higher for Italian workers than for foreigners. The average communicative content of the occupations performed by the group of foreigners is very low at 0.49.

Possession of a secondary school leaving certificate and being aged 25-34 also seem to influence wage differences. In particular, wage differences seem to be attenuated in favour of foreigners considering occupations with these characteristics, i.e. with low educational qualifications compared to higher qualifications and considering younger ages compared to older ages. This effect is essentially due to the "$\beta_A - \beta_B$" component, which returns strongly negative values. In the group of foreigners about 43% have a secondary school leaving certificate and 21% are young adults.

The Twofold Decomposition was then applied, which decomposes the difference in average results into two parts: the part explained by the differences between the groups in the explanatory variables and the unexplained part which is often attributed to labour market discrimination or the effect of unobserved variables.

We examined the results with group weights of "-1", which indicates that the reference coefficients come from pooled regressions in which the group indicator variable relating to citizenship (with value "1" for foreign citizenship and "0" for Italian citizenship) is omitted as a covariate.

Table 6 - *Twofold Decomposition Results.*

Group Weight	coef(explained)	coef(unexplained)	coef(unexplained A)	coef(unexplained B)
-1	284.99 euros	24.64 euros	2.89 euros	21.75 euros

Source: Own elaboration on ISTAT-INAPP data.

The results of the aggregate decomposition (Table 6) indicate that the wage gap of about 310 euros between Italian and foreign workers can be decomposed into 285 euros, which can be explained by group differences in the explanatory variables, and about 25 euros, which is the unexplained part.

We assume that the unexplained component of the wage gap is due to labour market discrimination and that the aggregate regression coefficients are non-discriminatory. The Blinder-Oaxaca decomposition would indicate that about 2.89 euros of the unexplained part comes from discrimination in favour of Italian workers ("unexplained A" component), while about 21.75 euros comes from discrimination against employed persons with foreign citizenship ("unexplained B" component).

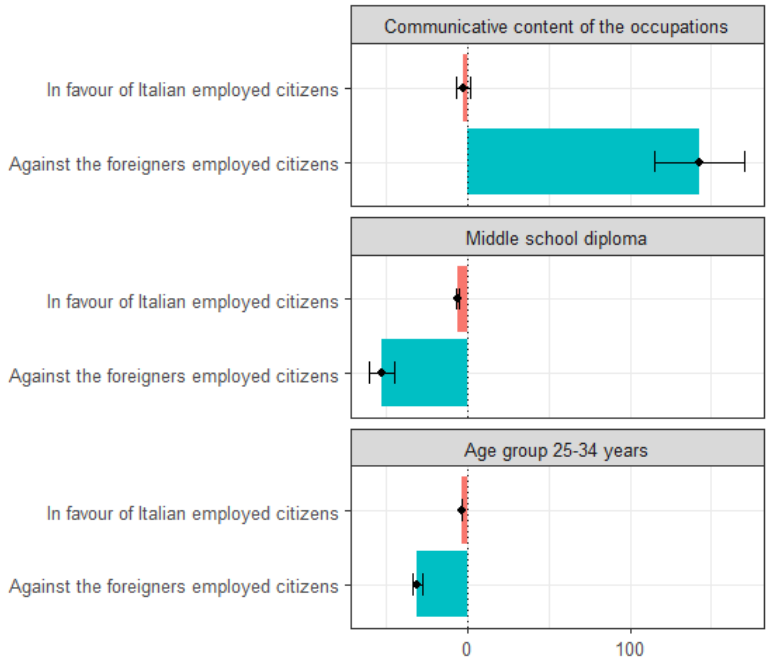
The contribution of each variable in the explained and unexplained part of the two-way decomposition was analysed. It turns out that the wage gap is largely determined by the fact that Italian workers are employed in more communicative and quantitative occupations and by a higher proportion of foreign workers working in small businesses. Furthermore, when analysing the unexplained part, it emerges that the wage differences between Italians and foreigners clearly benefit the natives when considering employment with a communicative content. It can be observed that the wage differences seem to be attenuated in favour of foreigners when considering those employed with a middle school degree as opposed to university degrees or an age between 25-34 rather than older ages.

We further explored the unexplained component. In Figure 5, we have examined the three variables that make a greater contribution to the breakdown: communicative content of occupations, average school-leaving certificate, age group 25-34. We therefore want to visualise how much of the unexplained wage gap can be attributed to discrimination in favour of natives and how much is due to discrimination against foreigners.

It seems that only the three components of discrimination against foreigners are significant. In fact, the unexplained part of the wage gap can be attributed to discrimination against foreigners in the exercise of a profession with a communicative content, whereas discrimination against foreigners seems to be attenuated for those with low educational qualifications (middle school certificate) compared to those with higher educational qualifications (university degrees) and for young adults, those between the ages of 25 and 34 compared to older ages.

Table 7 shows the specific numerical values of the point estimates of the unexplained discrimination components.

Figure 5 - *The sub-components of discrimination of the unexplained part in a Twofold decomposition of the wage gap between natives and immigrants.*



Source: Own elaboration on ISTAT-INAPP data.

Table 7 - *Estimates of the two components of unexplained discrimination.*

	Group Weight	coef(unexplained A)	coef(unexplained B)
communicative_std	-1	4.079	102.8048
middle_school	-1	-4.329	-41.587
young_adults	-1	-3.432	-32.1952

Source: Own elaboration on ISTAT-INAPP data.

4. Conclusions and Future Developments

We used the Blinder-Oaxaca Decomposition to examine the wage gap between Italian and foreign workers in Italy in 2019. The results of the analysis suggest that the wage gap can be explained by several factors. Mainly by the fact that Italians on average hold professions with a higher communication content than foreigners. Workers in non-communicative professions generally tend to receive lower wages in the Italian labour market. Consequently, the group of foreign workers in uncommunicative professions compared to Italians will, on average, earn lower

wages than their native counterparts. But also, that Italian employees work in professions with a higher quantitative content than foreigners, as well as a higher proportion of foreign employees working in small enterprises than natives. Moreover, the wage difference between natives and foreigners is reduced by considering low compared to high educational qualifications or youthful compared to older ages. Indeed, it would seem that foreigners compared to Italians have an advantage in terms of wages when considering low educational qualifications or younger ages compared to their native counterparts.

References

AMUEDO-DORANTES C., DE LA RICA S. 2011. Complements or substitutes? Task specialization by gender and nativity in Spain. *Labour Economics*, Vol.18, No.5, pp. 697-707.

BALDACCI E., INGLESE L., STROZZA S. 1999. Determinants of foreign workers' wages in two Italian regions with high illegal immigration. *Labour*, Vol. 13, No.3, pp. 675-709.

BLINDER A. S. 1973. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*, pp. 436-455.

COPPOLA L., DI LAUREA D., GEROSA S. 2013, The immigrant's wage gap in Italy. *Available at SSRN 2514714*.

ISTAT. 2023. *Statistiche report. Demographic Indicators, year 2022*.

OAXACA R. 1973. Male-female wage differentials in urban labor markets. *International economic review*, pp.693-709.

PERI G., SPARBER C. 2009. Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, Vol.1, No.3, pp. 135-169.

PIERONI L., D'AGOSTINO G., LANARI D. 2019. The effects of language skills on immigrant employment and wages in Italy. *MPRA Paper No. 91725*,

STROZZA S. 2003. Il reddito da lavoro degli stranieri: consistenza, determinanti, differenze di genere e per condizione giuridica di soggiorno, in N. Acocella e S. Sonnino (cur.), *Movimenti di persone e movimenti di capitale in Europa*, Il Mulino, Bologna.

Camilla PANGALLO, Sapienza University of Rome, camilla.pangallo@uniroma1.it
Oliviero CASACCHIA, Sapienza University of Rome, oliviero.casacchia@uniroma1.it
Corrado POLLI, National Institute for the Analysis of Public Policies (INAPP), c.polli@inapp.org

CARBON SUBSIDIES TO GREENING EDUCATIONAL AND BUSINESS ACTIVITIES[1]

Marina Albanese, Francesco Busato, Gianluigi Cisco

Abstract. This paper proposes a comparative framework to identify the impacts of alternative uses of carbon revenue and their redistribution. Although different studies have focused on the optimal carbon price path, the role of carbon revenue recycling is understudied. On this basis, we examine the macroeconomic impact of two carbon recycling strategies: i) encouraging investment in low-carbon projects at the firm level; and ii) subsidies to pursue SDGs, promoting the greening of human capital in the educational sector. To this end, this document presents an original Dynamic Stochastic General Equilibrium (DSGE) model that includes the higher education sector, low-carbon firms, and climate change. Our analysis shows that both strategies increase green human capital through two different channels. However, funding academics is crucial to improving the quality of education and promoting sustainable development in the long term. In contrast, firm-level green subsidies have a significant impact on short-term low-carbon productivity and mainly affect hiring and investment decisions. In both cases, these measures improve households' welfare.

1. Introduction

This study aims to investigate the effectiveness of carbon taxation and its potential macroeconomic implications in relation to green human capital. Although several studies have demonstrated the optimal level and dynamic for carbon prices (Acemoglu *et al.*, 2012), the most effective method for returning carbon tax revenues to the public has not been adequately determined. In recent years, global carbon pricing revenue increased significantly, to US$84 billion in 2021, representing a noteworthy additional source to fund climate mitigation, industry competitiveness, and economic and development goals (Agnolucci *et al.*, 2023). Global carbon pricing

[1] For research articles with several authors, a short sentence specifying their individual contributions can be provided here. The affiliations and contacts of the authors should NOT be reported here, but at the bottom of the last page.

revenue rose by almost 60% compared to 2020 levels, with emissions trading system revenues surpassing carbon tax revenues for the first time due to higher emissions trading system prices. Therefore, policymakers are paying increasing attention to how carbon revenues can be allocated most effectively.

However, while carbon revenues have been mostly used to fund measures that reduce emissions by promoting investment in low-carbon technologies. On the contrary, these funds could also be used to pursue broader objectives, such as health and human capital, i.e., direct funding to higher education institutions to improve education and research activities, particularly in climate change area, could be a potential strategy (World Bank, 2019). Nonetheless, the international climate policy debate has mainly focused on supply-side incentives in recent years. Although several studies have examined the adverse effects of carbon emissions increases and their policy implications in the Dynamic Stochastic General Equilibrium (DSGE) context (Fisher and Springborn, 2011; Heutel, 2012; Annicchiarico *et al.*, 2015), the role of human capital and educational policies has not been adequately considered. To fill this gap, this paper provides a comparative framework to identify the impacts of alternative uses of carbon revenue, mainly focusing on the *education sector*. In detail, this paper seeks to answer the following research questions*: What are the macroeconomic implications of employing a carbon revenues strategy to fund sustainable activities at the university level? How does this strategy differ from utilizing carbon revenues to subsidize low-carbon firms?* To answer this question, we extend a standard Environmental Dynamic Stochastic General Equilibrium (E-DSGE) model to include the higher education sector, high-carbon and low-carbon firms, and the climate module. Our research contribution extends the existing literature that employs DSGE models to investigate environmental issues (known as E-DSGE). Previous studies in this domain primarily focus on the supply side of the economy when addressing environmental implications (Fischer and Springborn, 2011; Heutel, 2012; Annicchiarico and Di Dio, 2015; Annicchiarico and Di Dio, 2017). Despite these contributions, two significant research gaps exist. First, our study innovates in its research topic, with few studies analysing carbon revenue recycling in a general equilibrium framework. Previous works neglect the potential impact of using carbon revenues for educational goals. Thus, our study analyses the effects of using carbon revenue to finance tuition subsidies. Second, our work innovates in its modelling approach. In a general equilibrium framework, no previous studies have examined the interaction between education supply, macroeconomic outcomes, and their impact on environmental issues This paper is structured as follows. Section 2 describes the DSGE model. In section 3, we present the model calibration. Section 4 discusses model performances and policy experiments. Finally, section 5 concludes.

2. The Model

The economy is populated by households, academic departments, academic institutions, green and traditional human capital, final and intermediate goods-producing firms, a government, and the climate system. The structure of the model consists of a standard TANK (Two Agents New Keynesian) model augmented to include the academic institution sector and a climate module. The model presents two types of households differing with respect to their ability to access financial and labor markets, namely Ricardian (Skilled) and Non-Ricardian (Unskilled). Ricardian households offer labor services in the green and dirty sectors and the green and traditional departments as skilled workers and teachers, respectively. In addition, Ricardian Households offers capital for both firms. On the contrary, non-Ricardian households offer labor services in the green and dirty sectors as unskilled workers, and they can spend part of their time in education activities (green or traditional). The production sector produces a unique final good combining two intermediate goods, dirty and green. This study provides two alternative types of human capital accumulation: green and traditional. In the following, we discuss the behavior of HEIs and departments, the remaining part of the model corresponds to the standard E-DSGE modeling.

2.1. Higher Education Institutions

The educational sector is defined by two hierarchic levels: academic departments and institutions. Departments hire teachers in perfectly competitive factor markets to convert time spent in education, from the non-Ricardian households into the new human capital F_t.The representative academic department employs teachers to provide two kinds of education, namely traditional and green. Academic institutions aggregate green and traditional courses to provide the total educational supply, considering costs related to each course implementation.

2.1.1 Green Departments

The intermediate green educational sector is dominated by a continuum of monopolistically competitive green departments indexed by $j \in [0,1]$ facing a demand function from the higher education sector. Notably, green departments produce new green human capital by combining time to green education ($E_{G,t}$) with teaching production hours ($N^R_{HG,t}$), as in the follows:

$$F(j)_{G,t} = A_{HG,t}\left[E(j)_{G,t}\right]^{\theta_G}\left[N(j)^R_{HG,t}\right]^{\theta_{NG'}} \in (0,1)\,, \tag{1}$$

where θ_G defines the marginal return to time spent in green education to former green human capital, because of the diminishing return to education, it takes values in the interval $(0,1)$; θ_{NG} defines the share of traditional teachers in the human capital formation; $A_{HG,t}$ measures the efficiency of human capital production technology. In addition, following The European University Association (EUA) survey, we consider quadratic adjustment costs of green sector teacher recruitment that reveal that tertiary institutions face several barriers to greening and environmental sustainability. They found a lack of employee engagement, coordination of activities, and strategic support among them. These types of challenges are formalized as follows:

$$\Gamma_t(N(j)^R_{HG}) = \frac{\gamma_N}{2}\left(\frac{N(j)^R_{HG,t}}{N(j)^R_{HG,t-1}} - 1\right)^2 N(j)^R_{HG,t}, \quad (2)$$

where γ_N represent the green teachers adjustment cost parameter. Accordingly, to its production function, input costs and adjustment costs on new green teachers, green Departments maximizes the following profits:

$$\max_{N(j)^R_{HG,t}} \Pi(j)^{AI}_{G,t} = \left(1+\tau^h_t\right)\frac{P^E_{G,t}}{P_t}F(j)_{G,t} - \frac{W(j)_{HG,t}N(j)^R_{HG,t}}{P_t} - \Gamma_t(N(j)^R_{HG}), \quad (3)$$

τ^h_t defines a green education subsidy for each of new green additionally green course implemented. From the maximization problem, we derive the input price equation given by:

$$\frac{W(j)_{HG,t}}{P_t} = \left(1+\tau^h_t\right)\frac{P^E_{G,t}}{P_t}\theta_{NG}\frac{F(j)_{G,t}}{N(j)^R_{GT,t}} - \left[\gamma_N\left(\frac{N(j)^R_{HG,t}}{N(j)^R_{HG,t-1}} - 1\right)\frac{N(j)^R_{HG,t}}{N(j)^R_{HG,t-1}} + \frac{\gamma_N}{2}\left(\frac{N(j)^R_{HG,t}}{N(j)^R_{HG,t-1}} - 1\right)^2\right] + \gamma_N\left(\frac{N(j)^R_{HG,t+1}}{N(j)^R_{HG,t}} - 1\right)\left(\frac{N(j)^R_{HG,t+1}}{N(j)^R_{HG,t}}\right)^2, \quad (4)$$

In our model, the rate of return per unit of green teaching labor $W_{HG,t}$ is defined as marginal products of $N^R_{GT,t}$.

2.1.2 Traditional Departments

The intermediate traditional educational sector is dominated by a continuum of monopolistically competitive traditional departments indexed by $j \in [0,1]$ facing a

demand function from the higher education sector. Traditional Departments produce standard human capital by combining time to traditional education $\left(E_{T,t}\right)$ with teaching production hours ($N_{HG,t}^{R}$), as in the follows:

$$F(j)_{T,t} = A_{HT,t}\left[E(j)_{T,t}\right]^{\theta_{GT}}\left[N(j)_{HT,t}^{R}\right]^{\theta_{GNT,}} \in (0,1)\,, \tag{5}$$

where θ_T defines the marginal return to time spent in education green in human capital formation, because of the diminishing return to education, it takes values in the interval (0,1); θ_{NT} defines the share of traditional teachers in the human capital formation; $A_{HG,t}$ measures the efficiency of human capital production technology. Traditional academic departments maximize the following profit:

$$\max_{N_{HT,t}^{R}} \Pi_{G,t}^{AI} = \frac{P_{T,t}^{E}}{P_t} F(j)_{T,t} \frac{W_{HT,t} N(j)_{HT,t}^{R}}{P_t}, \tag{6}$$

From the maximization problem, we derive the input price equation given by:

$$\frac{W(j)_{HT,t}}{P_t} = \theta_{NT} \frac{P_{T,t}^{E}}{P_t} \frac{F(j)_{T,t}}{N(j)_{HT,t}^{R}} \tag{7}$$

In our model, the rate of return per unit of traditional teaching labor $W_{HT,t}$ is defined as marginal products of $N_{HT,t}^{R}$.

2.1.3 Academic Institutions

The final educational supply is produced according to the following production function:

$$F_t(j) = \left\{(\gamma^e)^{1/\sigma_e}\left[F(j)_{G,t}\right]^{\frac{\sigma_e-1}{\sigma_e}} + (1-\gamma^e)^{1/\sigma_e}\left[F(j)_{T,t}\right]^{\frac{\sigma_e-1}{\sigma_e}}\right\}^{\frac{\sigma_e}{\sigma_e-1}},\ \gamma^e \in (0,1)\,, \tag{8}$$

where γ^e represents the share of intermediate green course used in the defining the final educational supply and $\sigma_e > 0$ is the elasticity of substitution between green and traditional intermediate course; γ^e represents the degree of environmental sustainability in the academic institution. Similarly, the demand curves for the two courses can be derived as follows:

$$F_{G,t} = \gamma^e \left(\frac{P_{G,t}^{E}}{P_t^{E}}\right)^{-\sigma_e} F_t, \tag{9}$$

$$F_{T,t} = (1-\gamma^e)\left(\frac{P^E_{T,t}}{P^E_t}\right)^{-\sigma_e} F_t, \tag{10}$$

where $F_{i,t}$ consist of a continuum of intermediate varieties $F_{i,t}(j)$ with $j \in (0,1)$and$j \in \{G,T\}$ as follows:

$$F_{i,t} = \left(\int_0^1 F(j)_{i,t}^{(\theta_e-1)/\theta_e}\, dj\right)^{\theta_e/(\theta_e-1)} \quad \theta_e > 1 \tag{11}$$

$$P^E_{i,t} = \left(\int_0^1 P(j)_{i,t}^{E(\theta_e-1)}\, dj\right)^{1/(\theta_e-1)} \quad \theta_e > 1 \tag{12}$$

Notably, non-Ricardian Households pay an equal fee for both green and traditional courses, P^E_t. However, for simplicity, we set this price equal to one.

3. Calibration

This section summarizes the calibration of the model presented in this study. The model is calibrated for the Euro area. Accordingly, we refer to previous studies focused on Euro area models to calibrate the households and production sections of the model. Macroeconomic parameters are set in accordance with those used in the calibration of a basic New Keynesian model (See Table 1).

Table 1 – *Model Calibration- Macroeconomic Parameters.*

Parameters	Descriptions	Values	Source
s_{NR}	Non-Ricardian (Unskilled) share	0.57	Eurostat
γ_N	Green Teaching Adjustment Cost	71	Annicchiarico *et al.* (2017)
γ	Share of Green Goods	0.2	Giovanardi *et al.* (2023)
γ^e	Share of Green Courses	0.2	Assumed equal to γ
σ	Elasticity of Substitution Green and Dirty goods	2	Giovanardi *et al.* (2023)
σ_e	Elasticity of Substitution Green and Dirty courses	2	Giovanardi *et al.* (2023)
θ	Elasticity of Substitution- Production	3.58	NAWM-II
β^j	Discount Factor Ricardian	0.98	Standard in Literature
ψ_j^i	Inverse of Frish Elasticity	1.00	Annicchiarico and Di Dio (2015)
δ_j	Depreciation Rate Sectoral Capital	0.1	NAWM-II
α_D	Share of Dirty Capital	0.3	NAWM-II
α_D^R	Share of High Skilled Workers -Dirty Sector	0.30	Calibration based on Eurostat
α_G	Share of Green Capital	0.30	NAWM-II
α_D^R	Share of High Skilled Workers -Dirty Sector	0.30	Calibration based on Eurostat
ξ_D	Emission per unit of dirty output	0.38	Annicchiarico and Diluiso (2019)
δ_m	Emissions Decay rate	0.0035	Gibson and Heutel (2020)

In contrast to the economy sector calibration, there is relatively little econometric evidence on the parameters in human capital formation. In detail, to calibrate the Higher Education Institution Sector, we refer to previous studies. Heckman (1976) estimates the parameters of the human capital production function, relative to time spent in educational activity, in the range from 0.51 to 0.80. More precisely, Kim and Lee (2007) set θ_G and θ_T equal to 0.49 and θ_{NG} and θ_{NT} such that$\theta_j + \theta_{Nj}$ is equal to 0.8. Turning to the depreciation rate of human capital, Heckman's (1976) estimates range from 4 to 9 % per year. In this study we set $\delta_j^h = 0.04$.

4. Results

In this section we examine the impacts of two carbon revenue recycling strategies: (i) subsidies for firms-level low-carbon investments; and (ii) subsidies for sustainable development goals, including the promotion of green human capital via academic and research programs. [2] Figure 1 displays the impact of a 10% carbon tax shock on the academic sector under both carbon revenue recycling strategies. When the focus is on promoting other sustainable development goals, the shock leads to an increase in revenues used to incentivize green initiatives in academia. In response, academic institutions adjust their course offerings to promote interdisciplinary approaches to climate change. This leads to an increase in demand for teachers with expertise in sustainability topics, resulting in a 0.12% increase in the number of green teachers. The subsidy also encourages households to allocate more of their time towards acquiring skills and knowledge relevant to the low-carbon sector, resulting in a 0.03% increase in green human capital. This mechanism contributes to an increase in the availability of green human capital in the economy. However, the impact on educational variables is limited due to staff adjustment costs and decreasing marginal returns of departments' production functions. Additionally, rigidities associated with staff turnover make the impact of this fiscal policy measure on the higher education sector more persistent, even after changes in educational offerings. In contrast, traditional courses, teachers, education, and human capital experience slight decreases. Figure 2 illustrates the impact of measures to promote the provision of green courses on the goods sector. The overall dynamics of this economy are shaped by two primary factors. First, a carbon tax impacts the competitiveness of carbon-intensive firms, favoring the low-carbon industry.

[2] The simulations have been obtained using numerical analysis and perturbation methods to simulate the economy and compute the equilibrium conditions outside the steady state. We solve the model using a second-order Taylor approximation around its steady state. The model has been solved by using the Dynare, a software platform for handling a wide class of economic models, in particular dynamic stochastic general equilibrium (DSGE) and overlapping generations (OLG) models. For details, see http:// www. cepre map. cnrs. fr/ dynare/ and Adjemian *et al.* (2011).

Second, the introduction of a "green" education subsidy alters the skillset available in the workforce. The available human capital acts as the conduit between education and production sectors. A hike in the carbon tax renders the 'dirty' industry less profitable than the 'green' one (as indicated by a decrease in dirty output by roughly 0.03%). Ricardian households, anticipating future growth in the green sector, adjust their preferences accordingly. They choose to divest from the dirty sector and increase their investment in low-carbon production. This investment decision results in an increase in green output, thereby escalating the demand for low- and highly skilled workers in the low-carbon sector (0.12% and 0.03%, respectively). However, low-skilled workers experience more significant effects. An educational grant influences low-skilled abilities in the green sector, enhancing its productivity. In addition, a portion of high-skilled resources is derived from the educational sectors to instruct in the new "sustainable" courses. A second scenario explores a carbon revenue recycling strategy designed to stimulate low-carbon production. Specifically, we examine the macroeconomic implications of a 10% carbon tax shock, with carbon revenues being directly used to support low carbon firms. After implementing this measure, green output, capital, and labor increase considerably more than with measures to subsidize green education (as depicted by the red line versus blue line in Figure 2). This type of measure has ripple effects in the educational sector.

Figure 1- *Impulse response functions to a 10 % Carbon Tax Shock under: "green firms" and "green education" recycle strategies: Education Variables.*

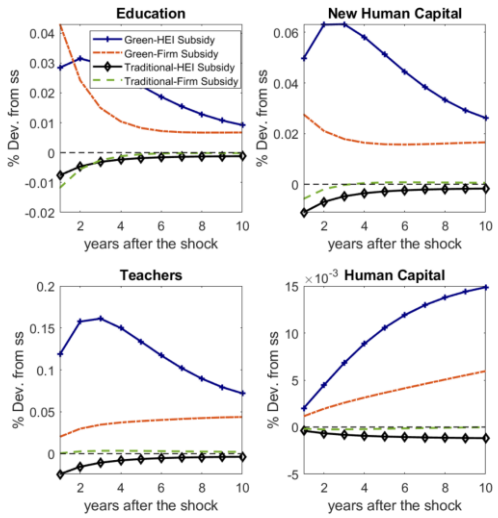
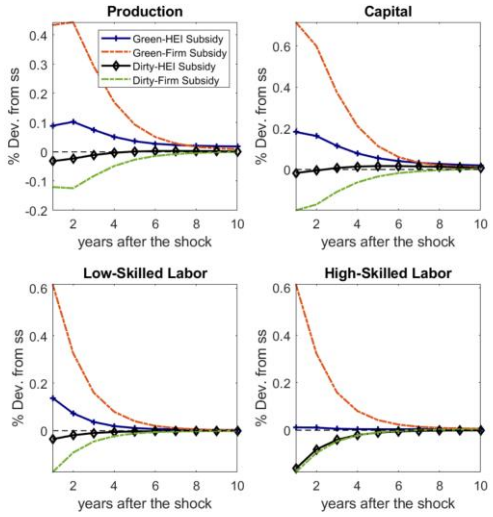
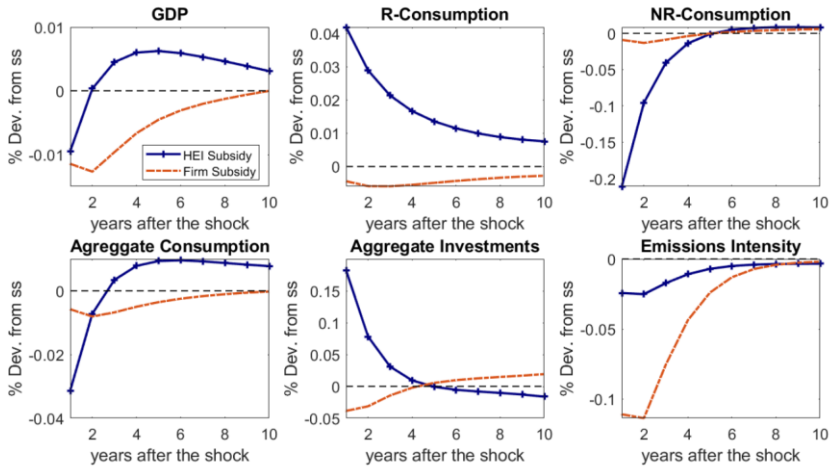


Figure 2- *Impulse response functions to a 10 % Carbon Tax Shock under: "green firms" and "green education" recycle strategies: Sectoral variables*



At the moment of the shock, green technology push measures significantly impact the demand for green education more than higher educational institution (HEI) subsidies. In essence, the rise in low-carbon production is forcing non-Ricardian households to pursue education that meets evolving market needs and acquire skills applicable in the green sector. Higher education institutions respond by tailoring their educational offerings to the demand for education, boosting the proportion of sustainable courses by 0.03%. This mechanism results in an increase in green teachers and green human capital. Conversely, traditional higher education experiences a decrease in demand and supply for conventional courses. With this comprehensive information, we can evaluate the effectiveness of both carbon tax recycling measures at a sectoral level. First, though a green production subsidy significantly influences the demand for green education, higher education institutions are less inclined to modify their educational offerings in response to this measure. Changing curricula is expensive, and only a green education grant measure can substantially influence academic decisions. However, a grant measure on green education has a lesser positive effect on low-carbon production at the shock impact (0.1% compared to 0.4% in the green production subsidy). In contrast, the significant positive effects of green human capital permit the maintenance of green output above their initial level. After examining the intersectoral dynamics, we can consider the performance of both carbon tax recycling measures in terms of aggregate variables (see Figure 3).

Figure 3- *Impulse response functions to a 10 % Carbon Tax Shock under: "green firms" and "green education" recycle strategies: Aggregate Variables.*



Measures encouraging green education increase aggregate investment and Ricardian consumption. However, they negatively affect non-Ricardian consumption and overall production. Specifically, non-Ricardian households reduce their consumption as they opt to invest more time in education, which decreases their disposable income. Nonetheless, enhancing their educational status enables to increase consumption after two years persistently. After about two years, the latter affects aggregate demand and pushes production above its original level. On the other hand, a measure of green production subsidy results in a reduction of both Ricardian and non-Ricardian consumption. Additionally, this fiscal policy has a negative impact on aggregate investment and production, with both variables decreasing by approximately 0.05 and 0.01, respectively. Unlike the previous scenario, the impact on human capital is less significant, and this mechanism prevents the aggregate variables from increasing persistently after shock exposure. Despite this, the effect on emission intensity is positive in both cases. However, in the case of subsidies for green technologies, the reduction in intensity is mainly due to a slowdown in production. In contrast, in the case of green education, the reduction occurs because the proportion of low-carbon output in the total output increases.

5. Conclusions

This paper presents a new DSGE model setup that extends the standard TANK framework to include the higher education sector, green and traditional human

capital, to analyzes the impact of two different carbon recycling policies This study finds that both subsidy measures positively influence HEIs to offer sustainable education, allowing for the greening of human capital. A subsidy for educational institutions is crucial to increasing the quality of education by allowing HEIs to hire more faculty members, even though non-Ricardian households spend less time on education compared to the green firm subsidy scenario. Furthermore, a grant measure for green education allows for maintaining new investments in green human capital above the initial level for a more extended period. However, a green stimulus measure at the firm level has a more significant impact on green production. This latter occurs because green firm subsidies significantly affect the labor market and skill acquisitions, allowing for an increase in the market share of green firms. Although both measures see a decrease in GDP at the impact level, the significant positive consequences of a sustainable education push measure on green human capital enable sustainable development in the long term. Additionally, both measures contribute to improving environmental efficiency and reducing emissions intensity. The study highlights the need for more research on the interaction between different areas of environmental policy, particularly when educational and cultural issues come into play.

References

ACEMOGLU D., AGHION P., BURSZTYN L., HEMOUS D. 2012. The environment and directed technical change, *American Economic Review*, Vol. 102, No. 1, pp. 131-166.

ADJEMIAN S., BASTANI H., JUILLARD M., MIHOUBI F., PERENDIA G., RATTO M., VILLEMOT S. 2011. Dynare: Reference manual, version 4.

ANNICCHIARICO B., DI DIO F. 2015. Environmental policy and macroeconomic dynamics in a new Keynesian model, *Journal of Environmental Economics and Management*, Vol. 69, pp. 1-21.

ANNICCHIARICO B., DI DIO F. 2017. GHG emissions control and monetary policy, *Environmental and Resource Economics*, Vol. 67, pp. 823-851.

AGNOLUCCI P., FISCHER C., HEINE D., MONTES DE OCA LEON M., PRYOR J., PATRONI K., HALLEGATTE S. 2023. Measuring Total Carbon Pricing, *The World Bank Research Observer*, lkad009.

FISCHER C., SPRINGBORN M. 2011. Emissions targets and the real business cycle: Intensity targets versus caps or taxes, *Journal of Environmental Economics and Management*, Vol. 62, No. 3, pp. 352-366.

GIOVANARDI F., KALDORF M., RADKE L., WICKNIG F. 2023. The preferential treatment of green bonds, Review of Economic Dynamics.

HECKMAN J. J. 1976. A life-cycle model of earnings, learning, and consumption, *Journal of Political Economy*, Vol. 84, No. 4, Part 2, S9-S44.

HEUTEL G. 2012. How should environmental policy respond to business cycles? Optimal policy under persistent productivity shocks, *Review of Economic Dynamics*, Vol. 15, No. 2, pp. 244-264.

KIM D., LEE C. 2007. On-the-job human capital accumulation in a real business cycle model: Implications for intertemporal substitution elasticity and labor hoarding, *Review of Economic Dynamics*, Vol. 10, No. 3, pp. 494-518.

WORLD BANK. 2019. "The World Bank Group Action Plan on Climate Change Adaptation and Resilience."

Marina ALBANESE, University of "Naples Federico II" and Centro interdipartimentale di ricerca Laboratorio di Urbanistica e di Pianificazione Territoriale "Raffaele d'Ambrosio" (L.U.P.T.). e-mail: albanese@unina.it.

Francesco, BUSATO, University of Naples "Parthenope," Department of Economic and Legal Studies. e-mail: francesco.busato@uniparthenope.it.

Gianluigi CISCO University of Naples "Parthenope," Department of Economic and Legal Studies. e-mail: gianluigi.cisco@uniparthenope.it.

THE EFFECTS OF THE LOCKDOWN ON CHILDREN: AN EMPIRICAL ANALYSIS [1]

Pietro Iaquinta, Federica Abbruzzese

Abstract. The two years of the pandemic, with a series of alternating lockdowns, have certainly influenced the neuro-cognitive development of each of us; even more so the growth of the children who experienced this sad period in a delicate and important moment of their lives. The resulting consequences will be able to be observed and studied in the future, when these now grown children will be able to express their fears, insecurities and everything that has been conditioned by a past that will only be a memory, perhaps to joke about or remember fondly. Meanwhile, in this work we wanted to outline a micro-investigation at a local level to test a possible cognitive-behavioral questionnaire on the evolution of educational and pedagogical processes. The numbers are small, but this moment could be useful to imagine a larger scale investigation using a much larger panel.

1. Introduction

The effects of the pandemic have not yet been fully explored, neither from an economic nor from a social point of view. In particular, children aged 0 to 3 years may have had a distorted view of the world, with consequences that will influence their current and future growth (Itard, 2007). Always attentive to the growth and education of the child, the Montessori Method pays particular attention to the environment and the effects that surrounding events can have on the child's education (Montessori, 2017a). Education, according to this scientific method, should start from the maternal womb so that the new-born's consciousness can develop in a complete and balanced way (Haines A., 2021, Eagleman, 2021). In particular, the possibility of a pilot project, carried out in an institute in Southern Italy where the Montessori method is applied, represented the determining element in the drafting of this work, not so much due to the small amount of data collected (only 44 interviews) but, above all, for evolutionary perspectives; an event that could count

[1] This article is the result of the coordinated work of both authors. In particular, however, paragraphs 1 and 3 are attributed to Iaquinta; to Abbruzzese paragraphs 2 and 4.

on an audience of around 7,000 children (and their families) spread across the entire national territory. Social analysis, although developed on small samples, represents a good method for investigating the evolution of people's behaviour, and in this case the opportunity is interesting because it has rarely been attempted to investigate how children react when faced with such changes. sudden, like the one suffered due to the Covid-19 restrictions. The sample size is typical of pilot surveys (Del Vecchio, 1995) and does not claim to have any statistical significance, but as already mentioned previously, has the sole purpose of probing into a study opportunity that could bring into its fold approximately 7,000 possible contacts.

2. The Montessori Method

Maria Montessori was: doctor, scientist, active feminist, anthropologist, educator, pacifist, traveller and tireless scholar. Presenting her as an educator would reduce her life and work, in honor of children, to a simple educational style; but she did much more. She loved to define herself as a "rigorous scientific researcher" intent on observing children in various occasions of daily life, without intervening, but only observing them in their "natural" movements, in their interests and needs (Mecocci 2019). Montessori's aim was neither to become an educator nor to invent an educational method, but to consider the child an integral part of the family, endowed with great potential which the adult must awaken through silent guidance, and which respects the principles of freedom and independence (Malloy, 2017).

His Method is considered revolutionary for three main reasons: the conception of the child (placed at the center of education and considered a "thinking being"); the importance given to the prepared and "tailored" environment; the role of the teacher, understood as a "guide", whose main role is to prepare the environment based on her observations (Gordon, 1995). In 1896 Maria Montessori graduated in medicine, with a thesis in psychiatry concerning the problem of hallucinations. Two years later, in 1898, at the Pedagogical Congress of Turin, she brought attention back to the education of children then defined as "abnormal" or "frenasthenic", for whom she took care of as a doctor. When the first "Children's House" was born, on 6 January 1907, in the infamous neighborhood of San Lorenzo in Rome, Maria Montessori was already known in Italy for her education, for her feminist struggles and for her commitment to children slows us down; so she decided to extend the experiment to able-bodied children too (Montessori, 2021).

In fact, the intent of the project was to redevelop the poor neighborhood of San Lorenzo, creating apartments equipped with toilets, electricity and water; to create a "home school" by entrusting the task of directing these special schools to Montessori,

given the commitment shown over the years in favor of children and, in general, the socially weak and inferior groups.

Montessori, with her Method, indirectly implemented a revolution in the educational field; in particular with the concepts of "absorbent mind" (Montessori, 2021) and sensitive periods (Montessori, 2017b) make a solid contribution to pedagogy. From 1910 M. Montessori decided to abandon medicine and dedicate herself completely to education, organizing various courses to prepare teachers, studying and expanding scientific research regarding the liberation of childhood. He traveled the world where, in the meantime, numerous nursery schools and Montessori schools sprang up.

Today Montessori nursery schools and nursery schools are now widespread (at a national and international level) in all continents, but there are conflicting opinions regarding the relevance of the Method and its effectiveness.

3. Territorial research

3.1. The fields of researcher

The opportunity for this research arises from the desire of a Montessori teacher, who, working in a Method school, decided to investigate the children's perception of the pandemic event. An event that unquestionably marked the experiential phase of growth of the typical users of these structures. As already mentioned previously, the prospect was to identify a sort of pilot survey in order to then be able to extend the data collection to the entire national territory, counting on an audience of over 7,000 possible users (D'Ovidio 2012). The questionnaire, however, had already been approved and therefore the summary results of this first analysis are affected by the choices made during the drafting of the questionnaire itself, being based on open-ended questions addressed to parents and/or those who take care of the children in the structure.

In this first phase, 44 questionnaires were collected, of which 41 were considered valid (Del Vecchio, 1995), aimed at understanding the perception that small users had during the period of forced closure imposed by the imminent pandemic that characterized 2020 and 2021.

The questionnaire, proposed with Google Forms, was made up of ten questions:

- eight multiple choice questions;
- two open-ended questions.

After generating the link, it was sent to parents of children aged 1 to 7 through social networks, such as Instagram and Facebook, and messaging apps such as WhatsApp and Telegram, starting from April 1, 2022, when the effects of the

closures were quite evident and the perception of young users was still very much alive, due to the various obligations that covered daily life in those times (masks, green passes, restrictions on access to shops, mandatory school closures, etc.).

After a few days of presence, 44 questionnaires were received, considered sufficient to incorporate some reports from the small users of the structure, covering approximately 85% of the users of the structure (52 children, of which 25 belonging to the nursery section and 27 to the school of childhood), although of these only 41 questionnaires were considered valid, therefore with a user coverage that falls just under 80% (Viola, 2005).

3.2. The structure of the questionnaire

The questionnaire, as already mentioned, consisted of 8 multiple-choice questions and 2 open-ended questions.

Even the multiple-choice questions, however, did not provide for a range of possible answers, but left users with the possibility of describing the proposed discriminations with their own adjectives and nouns. The result was the need to use a textual analysis technique with a spectrum of opinion in order to understand the quantitative meaning of the answers (Fabbris, 1997).

While keeping in mind the operational difficulties of this technique, it was thought that leaving the interviewees free to respond could have returned more pervading sensations of reality.

In any case, the questions posed to the users of the facility (their parents) were as follows:

1) Had the child sensed the spread of the virus?
2) What was the child's reaction to the news of the lockdown?
3) How did you explain the virus to the child?
4) How did you justify the school closure?
5) How long did it take for the child to want to go back to school?
6) How long did it take for the child to actually return to school?
7) During the lockdown how much time did the child spend in front of the TV?
8) During the lockdown, how much time did the child spend in front of tablets, PCs, smartphones?
9) In your opinion, what were the consequences deriving from the pandemic crisis in your child (short and long term)?
10) How did the child react to the news of going back to school?

3.3. First processing results

A first summary descriptive analysis (Iaquinta, 2018) highlights how over 50% of the children had understood the situation that was being created when the pandemic spread and only ¼ of them had arrived unaware of everything at the time of the imposition of the total block which took place at the beginning of March 2020.

It was not easy to find adequate and suitable words to explain the virus to children, many families tried to make it concrete such as:

- An invisible monster;
- A big bad germ;
- A little being that if touched or inhaled could cause discomfort to our body.

Unfortunately, the lockdown (imposed in March 2020) forced both families who decided to reassure their children and those who preferred to keep them in the dark to become aware of the reality.

Figure 1 – *Did your child sense the spread of the virus?*

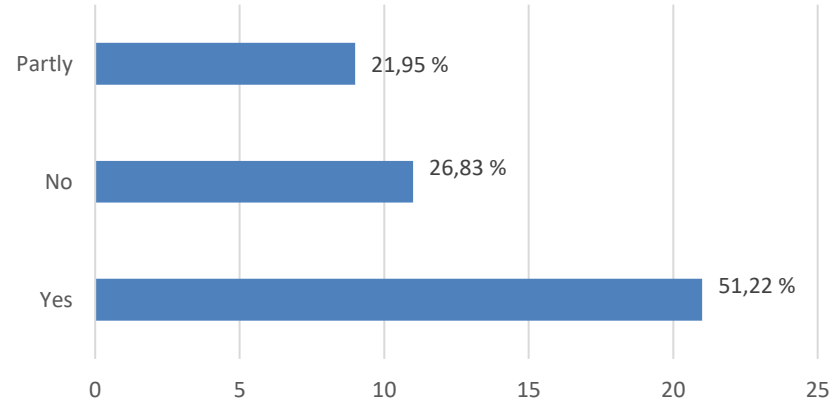


Figure notes. our elaborations on questionnaire data

That "home" environment has become the fortress of each of us, a protection from the virus, a "sacrifice" to be able to "feel well" and return to school soon. Thus, in a "fertile" moment of growth and interest, children were forced to stop and, indirectly, to give up moments of socialization, play and sharing. Upon hearing the news "today the school will be closed" and when faced with "why", many families, as emerged from the investigation, simply told the truth. When schools were completely closed,

children were initially happy but soon this illusion of freedom from school constraints turned into a “prison”.

Once they understood the situation, the children reacted differently. In order to classify and have a minimum amount of feedback regarding the responses received to the questionnaires (which presented the most disparate responses) we operated manually by classifying those responses that could be defined as positive (spending more time with the family, being able to play more freely, etc.) from those that presented a negative impact (not being able to go to school, worrying about illnesses, not being able to play with classmates, etc.), and, finally, by those that denoted an absolute sense of indifference to the situation.

Figure 2 – *What was your child's reaction to the news of the lockdown?*

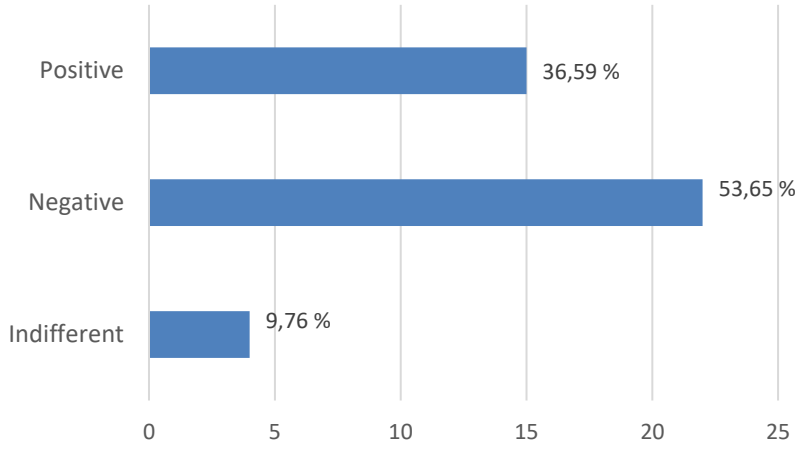


Figure notes. our elaborations on questionnaire data

As can be seen from the data in Figure 2, over half of the children showed a negative attitude towards the lockdown event.

Once they understood that the lockdown would somehow affect their habits, it took the children some time before they began to show the typical intolerance of childhood. An interesting measure of the child's distress can be identified in the time between the start of the total closure and the request to return to school.

Also, in this case, given the heterogeneous quantity of responses, a reclassification was carried out, obtaining a grid of possible responses that varies from a few days to a few weeks and is monthly in size.

From the results of the responses it emerges that almost ¾ of the interviewees expressed the desire to return to school within a few weeks.

Figure 3 – *After how long did the child want to go back to school?*

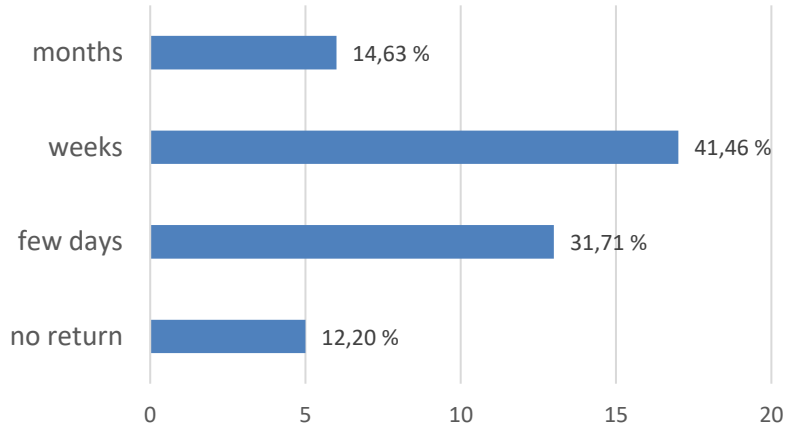


Figure notes. our elaborations on questionnaire data

Among the most interesting topics that can be explored in more detail from the proposed questionnaire is, certainly, the management of the time spent by children in the new domestic dimension imposed by the lockdown.

One of the most recent pedagogical problems is given by the excessive use of electronic devices (PCs, Tablets, Smartphones) and television.

The most frequent mistake found in families is the belief that the use of technology in childhood is sufficient for growth (Fogassi, Raniero, 2019). In reality, disproportionate use can be counterproductive if not even limiting in their growth. Time is one of the aspects underestimated by parents, in fact the research showed that approximately ¾ of the children spent more than two hours a day in front of the TV and more than 3 hours using electronic tools, a precious and important learning time in this delicate growth phase. The parents themselves underlined the abuse and the repercussions, as will be seen better later, that the pandemic crisis has had on the educational path of the interviewees. This is not intended to imply a refusal to use technology in education, but rather an invitation to reflect on the educational purpose and the appropriate age to introduce these devices. The brain, in early childhood, is not ready to welcome certain stimuli, consequently they interfere with brain development. There are very conflicting opinions regarding the appropriate age for their introduction, scholars recommend around 6-8 years of life (Valle, 2017). In this regard, Montessori considered the first six years of life fundamental for balanced and

healthy growth, free from negative influences but open to exploring the world through real and concrete experiences. Hence the importance given to the use of the hands and the development of the senses.

Figure 4 – *In your opinion, what were the consequences deriving from the pandemic crisis in your child (short and long term)?*

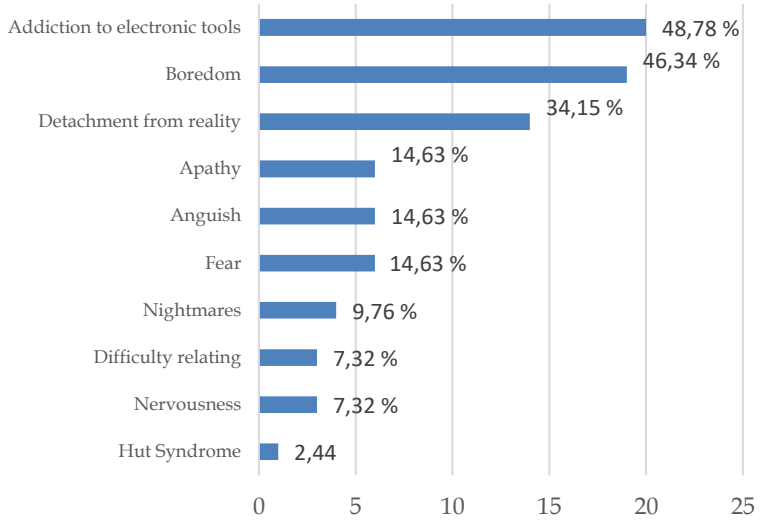


Figure notes. our elaborations on questionnaire data

The consequences of the pandemic have been multiple, both in the short and long term. The greatest enemy, as mentioned previously, for approximately 46% of the subjects interviewed was boredom, which they attempted to overcome by resorting to the almost spasmodic use of electronic instruments, which however resulted, as a consequence, in the acquisition of period of dependence.

Fears follow, with almost 40% (the children have started to be afraid of being alone, as well as moving independently) and around 35%, they have suffered a detachment from reality. This is followed by a lower percentage, around 10-13%, but this should not be underestimated, with consequences such as: anguish; nightmares; apathy; eating disorders. In fact, around half of the subjects interviewed developed negative situations in experiencing the closure period, an event which certainly had a strong influence on the growth of the children (especially the youngest) and which will almost certainly have repercussions on their future educational path, when they become teenagers.

Figure 5 – *How did the child react to the news of going back to school?*

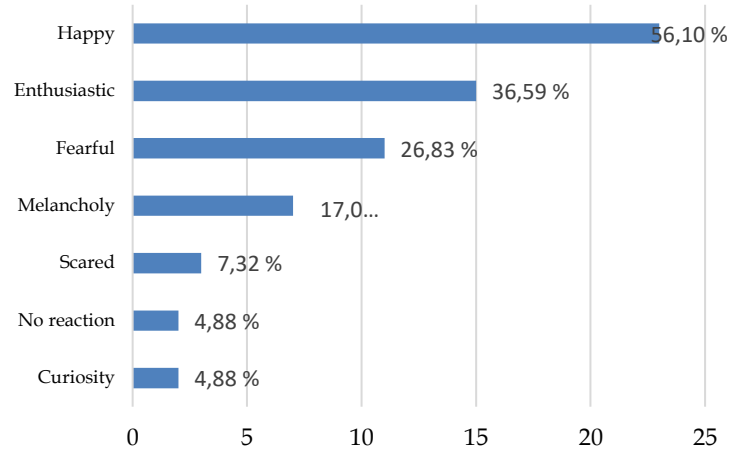


Figure notes. our elaborations on questionnaire data

The end of the lockdown obviously triggered the reaction of the young school users, who in almost 90% of cases expressed positive feelings towards the reopening of schools, (Isaacs, 1982) also linked to the end of isolation and the possibility of seeing their classmates again. game. Only a quarter of the children showed moments of negativity, most likely due to the previous school situation, thus demonstrating a pre-existing discomfort.

4. Conclusion

The last twenty years have been characterized by a slow abandonment of childhood; the most alarming fact is that this topic seems to be of less and less interest to the institutions, no one deals with it and no one cares to do so. Not only does the State appear indifferent to the education of children (especially under 3 years old), but the recently enacted laws seem to have been created specifically to further limit their growth. Just think of the October 2017 circular in which the Ministry of Education obliged school staff to hand over children and young people to their parents or a responsible adult. Especially for children aged 11 to 14, this security has become a real limit since they have been denied the possibility of becoming autonomous, responsible and independent.

Today, at least in Italy, we can say that, in addition to childhood, we have also lost children, that is, children have left the collective imagination, losing their social relevance. Children are seen as a burden and a "management difficulty" to worry about, parents are not helped, nor supported and supported in the education of their children, furthermore society does not seem to notice the absence of children in public places. Pedagogical support, which educates them and informs them about their child's growth, would probably make the parent's difficult task less complex.

The state, on the other hand, further burdens parents with even greater burdens, creating laws tailored to every situation in which parents could endanger their children. With the result of making them doubt even more - as if there was any need today - of their own abilities. Think of the law on life-saving child seats born following various episodes in which children were forgotten and helped by their parents, busy with their occupations and work worries. The State has therefore decided to make these seats mandatory so as to be able to intervene in the event of involuntary abandonment, through a connection with an App programmed to send a message in the event of "forgetting" your child in the seat.

All of this is an alarming sign that indicates a fracture in the relationship between child and parent, a fracture that is also evident in school. Educators and teachers, in fact, complain about requests from parents that border on the absurd:

- “try not to paint it today, otherwise it will get dirty!”;
- “don't let it come out if you use chalk!”;
- “avoid washing, otherwise you'll get wet!”;
- "don't give him the salty pasta, who knows he'll eat it!"
- “don't let him run today, otherwise he will sweat”.

Requests that are unlikely to be fulfilled, given that the protagonist in question is always the child, whose aim is to grow and learn in freedom!

References

DEL VECCHIO F., 1995. *Scale di misura e indicatori sociali*. Bari: Cacucci Editore

D’OVIDIO F. D., 2012. *Elementi di statistica per la valutazione dei servizi*. Padova: CLUEP.

EAGLEMAN D., 2021. *L’intelligenza dinamica.* Milano: Garzanti.

FABBRIS L., 1997. *Statistica multivariata, Analisi esplorativa dei dati*. Milano: Mac-Grow-Hill.

FOGASSI L., RANIERO R., 2019. *Maria Montessori e le neuroscienze. Cervello, mente, educazione*. Roma: Fefé editore.

GORDON T., 1995. *Insegnanti efficaci.* Firenze: Giunti & Lisciani editori.

HAINES A. 2021. *Maria Montessori Lezioni da Londra 1946*. Perugia: Il leone verde.
IAQUINTA P., 2018. *Esercizi di statistica descrittiva*. Bari: L'Arco e la Corte.
ITARD J., 2007. *Il fanciullo selvaggio dell'Aveyron… cresciuto nei boschi come un animale selvatico*. Roma: Armando Editore.
ISAACS S., 1982. *Il valore educativo della scuola materna*. Brescia: La scuola.
MALLOY T., 2017. *Montessori e il vostro bambino*. Roma: Edizioni Opera Nazionale Montessori.
MECOCCI M., 2019. *Narrare il vero.* Perugia: Terra nuova edizioni.
MONTESSORI M., 2017a. *Educare alla libertà (1949)*. Milano: Mondadori.
MONTESSORI M., 2017b. *La mente del bambino (1949)*. Milano: Garzanti.
MONTESSORI M., 2018. *Come educare il potenziale umano (1947).* Milano: Garzanti.
MONTESSORI M., 2021. *Educazione per un mondo nuovo (1946).* Milano: Garzanti.
VALLE M., 2017. *La pedagogia Montessori e le nuove tecnologia*. Torino: Il Leone Verde.
VIOLA D., 2005. *Valutazione della qualità e customer satisfaction nei servizi di pubblica utilità e nelle imprese*. Bari: Cacucci.

Pietro IAQUINTA, Calabria University, Pietro.iaquinta@unical.it
Federica ABBRUZZESE, MoM Montessori Montalto, federicabbru@gmail.com

GEO-REFERENCED SENTIMENT ANALYSIS FOR TOURISTS' POINTS OF INTEREST: THE CASE OF MATERA EUROPEAN CAPITAL OF CULTURE

Luigi Celardo, Michelangelo Misuraca, Maria Spano

Abstract. Known as the 'city of Sassi', Matera underwent a renewal process involving all of the Basilicata regions in the last few years. The city's Tourism resources were almost unknown at a national and international level, although the Sassi were included in the UNESCO World Heritage List since 1993. The European Capital of Culture 2019 nomination triggered an intense regeneration, opening the city to global tourism and revealing a high resilience. Tourists' experiences and opinions have been valuable resources for designing tourism activities and creating a new symbolic identity for the city, especially in the Web 2.0 era. Here we propose to compute the reviews' polarity scores and use them with other characteristics (e.g., price, offered services and type of tourist facilities) to build spatial clusters according to the logic of Local Spatial Association Indicators (LISA). The geo-referenced semantic orientation of reviews concerning a particular activity or attraction represents a useful quantitative feature for further analyses and the production of territorial statistics. The proposal can be extended to other cases to monitor the change of sentiment towards specific areas of interest and plan possible intervention policies.

1. Introduction

Cyberspace has opened new opportunities to improve communication and develop customer acquisition and retention approaches. A crucial aspect is the phenomenon of online interpersonal influence (Senecal and Nantel, 2004), which can be referred to as *electronic word-of-mouth* (eWOM), based on the previous Westbrook (1987) definition of word-of-mouth (WOM). The eWOM is an informal communication of consumers about the characteristics and the mode of consumption of particular goods and services when they need to make a purchase decision. Among the different sectors, such opinions became increasingly significant in hospitality and tourism, where the intangible nature of products makes it difficult for an assessment and interpersonal confidence in products is extremely important.

The reputation generated by digital platforms is used by diverse actors in the tourism industry. A good eWOM about facilities and attractions helps potential visitors find information and reinforces images and opinions, as it contributes to creating the symbolic image of tourist destinations. On the other hand, eWOM takes on a strategic role in destination management and operators begin to integrate traditional contextual analysis with tourist review data. In this framework, monitoring tourist sentiment is crucial for analysing the impact of tourism growth in specific areas, considering external and internal factors that influence the image. Here we propose to compute polarity scores of reviews about touristic points of interest (attractions, restaurants, accommodation) on social media platforms such as TripAdvisor, to have both a spatial representation of the spread of sentiment and to construct spatial clusters, according to the logic of *Local Spatial Association Indicators* (LISA), to identify hot or cold spots where positive and negative sentiments are more concentrated (Anselin, 1995). Some authors have already used LISA to classify tweets into different topics and group them spatially to show the relationship between content and place characteristics (Hu *et al.*, 2020) or to find the distribution of topics associated with urban features (Zhong *et al.*, 2018). The proposal effectiveness is shown through a case study about the city of Matera, which recently experienced a regeneration process involving, more generally, the entire Basilicata region. The city's tourism resources were almost unknown, both nationally and internationally. Although in 1993, the Sassi became the first site in southern Italy to be included in the UNESCO World Heritage List as a 'cultural landscape', the real change came with its nomination as the European Capital of Culture for 2019. The nomination triggered an intense development, opening the city to global tourism and demonstrating a high resilience (Ivona *et al*., 2019). The Internet played a key role in reinterpreting the local urban reality, where tourists want to have a unique immersive experience in the Lucanian community rather than being mere onlookers. Our strategy is developed within the framework of the so-called *Ambient Geographic Information* (AGI) system of Stefanidis *et al.* (2013), and it can be extended to other cases to monitor the evolution of sentiment towards specific areas of interest and plan possible intervention policies. The geo-referenced semantic orientation of tourists' opinions represents a useful quantitative feature for further analysis and the production of spatial statistics.

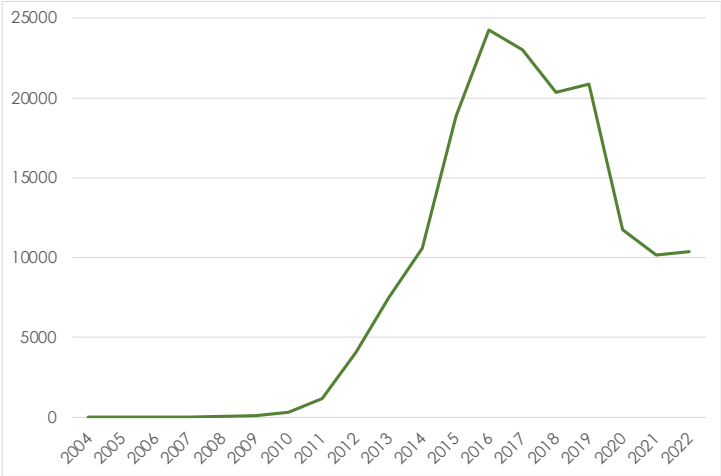
2. Data and methods

The so-called *experiential* tourism, which stems from the "desire to see life as it is lived and to get in touch with the locals" (Lemmi, 2009), is a concept of tourism strongly inherent to the European Capital of Culture event promoted by the European

Union (Guala, 2002). This initiative gives the winning city 365 days to enable its environmental and cultural heritage, involving a broader territorial context than the purely urban area to create virtuous synergies. Matera's strategic project has enabled the Lucanian city to offer alternatives and specific accommodation to around 700,000 temporary citizens in an urban area that, after having been condemned to painful isolation along with the entire regional system of Basilicata since the Second World War, is now undergoing an intense internal regeneration, opening up to competition from glocal tourism and showing a high degree of resilience (Pollice and Urso, 2014). The dossier submitted to the jury for the 2014 European Capital of Culture nomination was entitled *Open Future*, highlighting the possibility of combining science and technology with an exceptionally creative streak to consolidate the city (and Basilicata) position in the creative sector at a European level. The report also mentions the creation of an online cultural platform to provide tourist information at all spatial levels and in real-time to reach and be reached directly by potential visitors (Pollice, 2010) through eWOM. Embracing the same idea, we developed a dataset including the opinions of tourists that visited Matera in the last years. All the restaurant, accommodation, and attraction reviews used in this study have been extracted from the Italian TripAdvisor website, using `MATERA` in the query and restaurants, accommodation and attractions as the main category. Reviews posted between 2004 and 2022 (up to March 23, 2023) have been stored in a local repository together with some metadata, like *address*, *latitude*, and *longitude* (validated with the corresponding Google Maps ID_place), *rating*, # *number of reviews*. We considered the total number of activities and filtered the facilities by latitude and longitude values not geolocated outside the sub-areas defined by the Municipality of Matera[1]. The data extraction procedure belongs to the so-called *web scraping*. Generally, web scraping can be defined as a systematic process of identification and retrieval of content of interest from the Web. A software agent mimics the browsing interaction between the web servers and a human user. Step by step, the agent accesses as many web pages as necessary, analysing the content to extract the data and structure them in the desired form. In our case, two scripts were built using *Python* and *SeleniumLibrary* to extract information on the different POIs and the related on individual reviews. We obtained in this way 806 points of interest with 164,651 reviews.

[1] http://dati.comune.matera.it/dataset/aree-sub-comunali-comune-di-matera

Figure 1 - *Reviews in Italian on Tripadvisor (2004-2022).*



Analysing the evolution over the years is useful to assess the overall trend since the number of reviews published over time is an indicator of the consolidation of eWOM about Matera. Figure 1 shows how the number of reviews tends to increase over the years, reaching a peak in 2014, when Matera officially presented its bid to become the European Capital of Culture. The subsequent years of preparation for the event increased until reaching a peak in 2019. The year Matera became the Capital of Culture was crucial, but it should be remembered that the spread of the Covid-19 pandemic had a decreasing effect on the number of visits. It is worth noting that the number of reviews returns to the years before 2015, indicating that even though the number of visitors decreases, the number of reviews remains at around 10,000.

To acquire the sentiment scores associated with POIs, we employed an original customised lexicon of Italian terms to perform a lexicon-based sentiment analysis of the reviews. Most resources in the sentiment research area, like lexicons, labelled supplies and NLP tools, are mainly available in English. The lack of linguistic resources is critical in most studies, producing a so-called *lexical gap* (Chiavetta *et al.*, 2016). Thus, we built an Italian lexicon by merging the resources developed in the Opener project (Russo *et al.*, 2016) and other selected studies (e.g., Bolasco and Della Ratta, 2004). The resulting set contains 12,400 polarised lemmas with a value of +1 if positives and -1 if negatives. The reviews were lightly pre-processed. Non-alphabetic characters and symbols - such as numbers or emoticons - were removed to include only content-bearing terms. Polarity scores were calculated using a sentence-level logic (Balbi *et al.*, 2018). Given a review r_i (i = 1, …, n), its a_i sentences $\{s_{i1}, ..., s_{ik}, ..., s_{ia_i}\}$ are identified by considering as separators strong punctuation marks like full stops, question marks and exclamation marks. The k-th

sentence s_{ik} is a sequence of its p_k terms $\{t_{ik1}, ..., t_{ikj}, ..., t_{ikp_k}\}$. Each term t_{ikj} in the k-th sentence of the i-th review is compared with the terms in the lexicon, assigning a -1 to negative terms and a +1 to positive terms, respectively. Terms not listed in the lexicon are scored with a null value. The polarity of each term is then weighted considering negators (e.g., *mai*, *nessuno, nessuna*), amplifiers and de-amplifiers (e.g., *poco, molto, pochissimo*), adversative and contrasting terms (e.g., *ma*, *tuttavia*). This weighting scheme allows for emphasising or dampening the negativity or positivity of each polarised term, leading to a more effective measure of semantic orientation (Vechtomova, 2017). The polarity score of each sentence $\mathrm{PS}_{s_{ij}}$ is obtained as the sum of weighted term scores $\mathrm{PS}_{w_{ikj}}$ on the sentence length:

$$\mathrm{PS}_{s_{ij}} = \frac{\sum_{k=1}^{p_j} \mathrm{PS}^{*}_{t_{ijk}}}{\sqrt{p_j}} \tag{1}$$

Since we are interested in obtaining a polarity score at a review level, we calculated an overall score PS_{r_i} for each text by an average of sentence polarities:

$$\mathrm{PS}_{d_i} = \frac{\sum_{j=1}^{q_i} \mathrm{PS}_{s_{ij}}}{\#(q_i^{-}) + \#(q_i^{+}) + \#(q_i^{0})} \tag{2}$$

where q_i^{-}, q_i^{+} and q_i^{0} are the numbers of sentences in r_i with a negative, positive, or neutral polarity, respectively.

In the last step, we used a global autocorrelation indicator called Global Moran's I to relate the sentiment score to the spatial location of the points of interest. The basic idea is to assess the absence of spatial randomness in the distribution of a variable. We relate the global value of sentiment to its importance in neighbouring areas to summarise the degree of spatial similarity between neighbouring observations across the study area (Getis *et al.*, 1992)

$$I = \frac{n}{\Sigma_i \Sigma_j w_{ij}} \frac{\Sigma_i \Sigma_j w_{ij} z_i z_j}{\Sigma_i z_i^2} \tag{3}$$

where z_i, is the deviation of sentiment score for each i observation from its means, w_{ij} is the spatial weight for the pair of observations i and j, and n is the number of observations. A positive value of Global Moran's I represent a situation where similar values are found in an adjacent area.

Nevertheless, we are interested in the global spatial representation of sentiment and in identifying regions, i.e. the sub-areas defined by the municipality of Matera, which influence the average positive sentiment at the local level due to their spatial proximity or remoteness. For this reason, local spatial autocorrelation could reflect

sentiment spread between scores in neighbouring locations, representing scenarios where scores in one place influence other sites or where scores in one location causally affect other areas. Local measures of spatial autocorrelation focus on the relationships between each observation and its environment rather than providing a single summary of these relationships across the map. In this sense, they are not summary statistics but scores that allow us to learn more about the spatial structure of our data. A Local Moran's I_i identifies cases where the value of an observation and the average of its surroundings are more similar (H-H or L-L) or dissimilar (H-L, L-H) than we would expect randomly. The index is applied to every observation. In this way, there are as many statistics as original observations. The formal representation of the statistic can be written as follows:

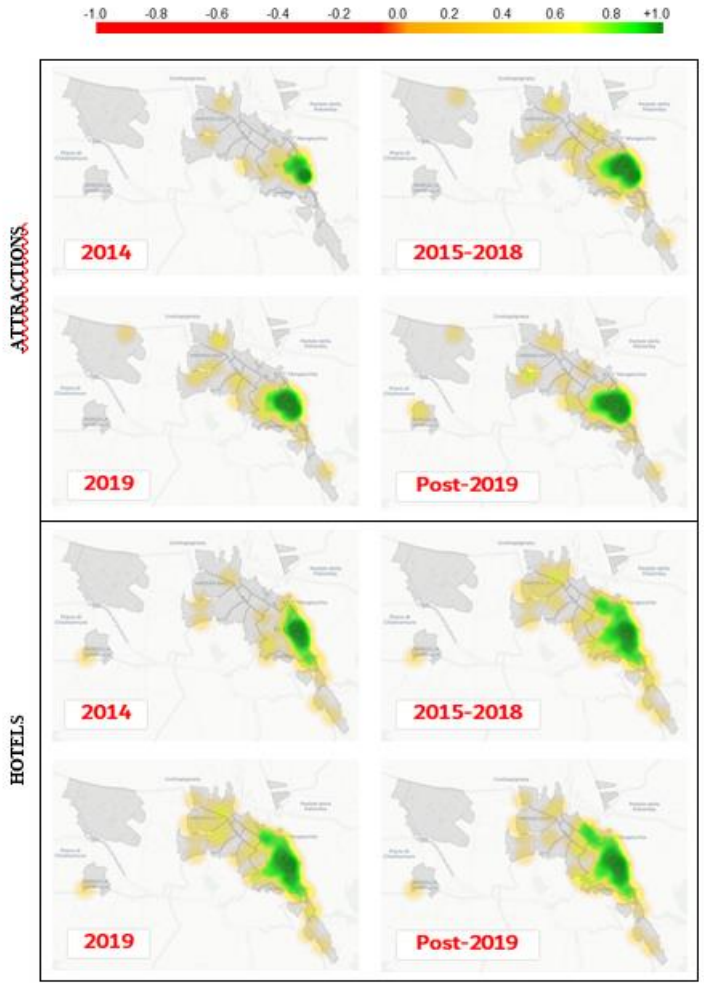
$$I_i = \frac{n(y_i - \bar{y})}{\Sigma_i (y_i - \bar{y})^2} \sum_j w_{ij} \left(y_j - \bar{y}\right) \tag{4}$$

where y_i is the value of sentiment score for each *i* observation in the distribution of data, w_{ij} is the spatial weight for the pair of observations *i* and *j*, and *n* is the number of observations.

3. Empirical results

Given the scale of the investment, the question of long-term legacy is essential for all ECOCs (European Capitals of Culture). While all cities that were ECOC in 1995-2004 set long-term objectives for their year of glory, only half set up funds or bodies to pursue them. Therefore, monitoring the widespread opinion is a good way to avoid overlooking the issue of project sustainability over time when planning the event (Corinto and Nicosia, 2016). A first graphical output of the analytical strategy proposed here allows for displaying attractions, hotels, and restaurants' sentiment orientation by means of a geographical heatmap, where the red colour represents a negative sentiment score, whereas the green colour represents a positive score. In Figure 1, the different periods are compared in order to appreciate the temporal evolution of the tourists' opinions about the POIs. In the year of the nomination (2014), a wide green area can be seen around the Sassi districts and the historical city centre for all the POIs categories. In the other analysed periods, the positiveness gradually expanded to reach new areas, even in some districts of the new city centre, such as San Pardo, San Giacomo and Villa Longo, particularly for hotels and restaurants. Thus, an effective spread of positive sentiment aligns with the main objectives for promoting the city of Matera included in the dossier submitted for the Cultural Capital bid.

Figure 2 - *Sentiment Heatmap of attractions, hotels and restaurants (2004-2022).*



Before performing the spatial analysis, we calculated the Global Moran's *I* from the sentiment scores, obtaining a 0.568 value. To give additional substance to the hypothesis of non-randomness in POIs distribution, by means of a Monte Carlo simulation with 999 permutations, we carried out a statistical test and obtained a <<0.001 *p*-value. Figure 3 depicts the four clusters of POIs in the suburbs, with each area associated with one of the results typically generated by the LISA model (high-

high in red, high-low in orange, low-high in light blue and low-low in dark blue). The output can be read again with respect to an overall improvement in tourists' perception from the nomination as the Capital of Culture to the present day.

As in the previous output, we observed for the LISA model a general spread of positiveness from the Sassi area to other Matera suburbs, such as the historic city centre, San Pardo and San Giacomo districts. The non-significant clusters for all the tourist attractions tend to diminish over time, showing high-high clusters (in red) from the Sassi districts to more peripheral neighbourhoods such as Agnara and Serra Rifusa. It is interesting to note how for attractions, the different areas of the city flowed across time in the high-high cluster (in red), giving evidence of the success of Matera promotion and of the triggering effect of the nomination as Capital of Culture. As concerns accommodations, the COVID-19 pandemic influenced the spatial distribution of POIs, highlighting how not all the structures were able to maintain a high standard in the tourists' view.

Figure 3 - *Hot-spots and cold-spots of attractions, hotels and restaurants (2004-2022). (segue)*

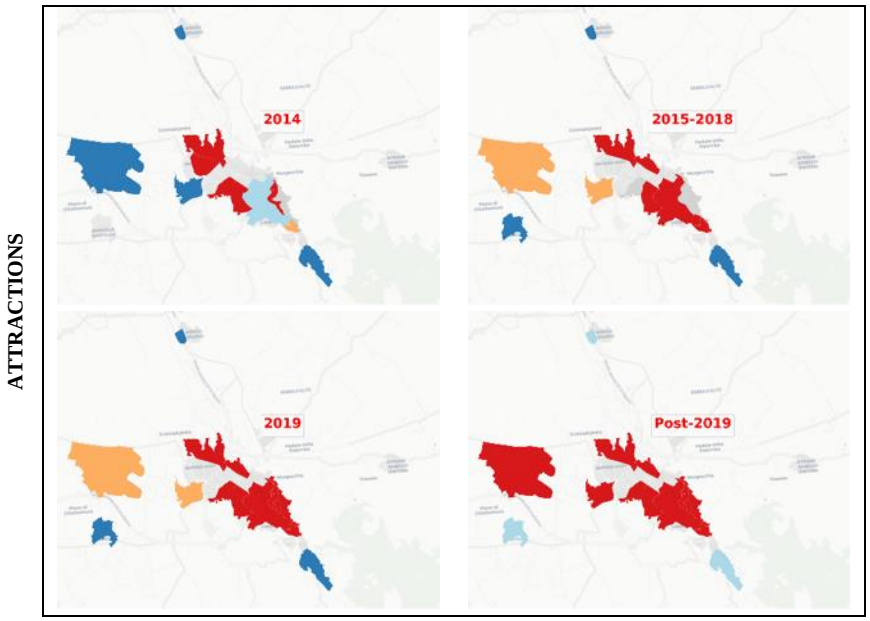
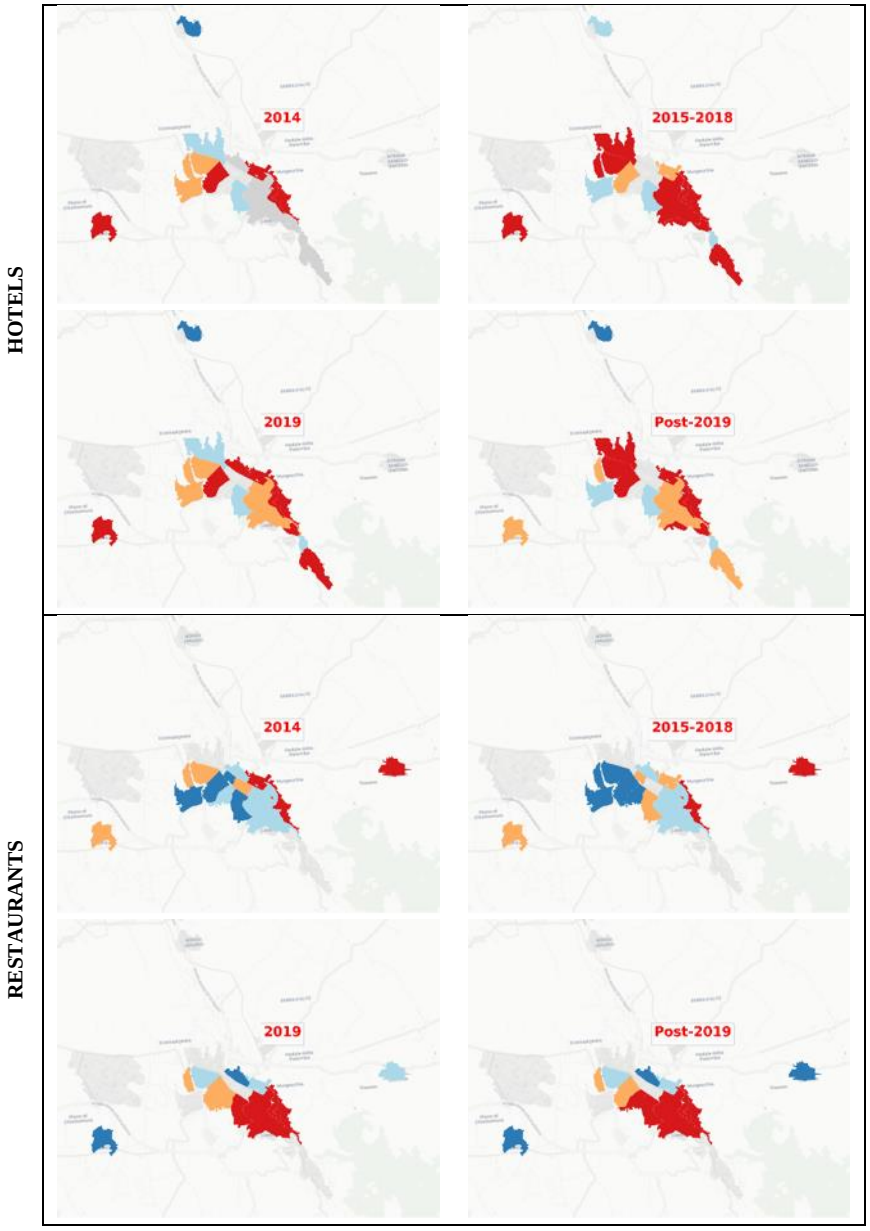


Figure 3 - *Hot-spots and cold-spots of attractions, hotels and restaurants (2004-2022). (continua)*



4. Conclusion and final remarks

Following Matera's example, many cities can benefit from the positive effects of enhancing social and cultural assets. The process of urban regeneration is a valuable case study to illustrate the success and importance of coopetition between actors in a territory (Nguyen *et al.*, 2022).

The integrated analysis of polarity scores and geographical data can help explore tourists' opinions and perceptions, providing all stakeholders with a powerful tool to assess and monitor the attractiveness of a given territory and develop actions or policies to maintain or improve its image. Thanks to sentiment analysis, tourism operators and public administrations can study the orientation of reviews published by tourists after a trip (considering all facets of this experience) or after visiting a tourist attraction. The possibility of geo-referencing tourist sentiment increases the informative value of this type of data. The ability to manage these data flows also provides the opportunity to integrate traditional contextual analysis with data-driven strategies by incorporating the mutable opinions of tourists into planning. In this sense, creating interactive maps where the territory can be explored at different levels is possible instead of using a static representation of sentiment. Since the monitoring process included in the Matera OPEN FUTURE dossier requires real-time feedback on the impact of the city's renewal and regeneration processes, it could be useful not only to display static photographs of the various LISA clusters in the area but also to consider the possibility of identifying dynamic clusters. In this sense, other LISA techniques, such as its spatio-temporal version, could be tested. Closely related to the comparative static analysis of LISA statistics, the dynamic performance could provide additional insights into the shape and direction of cold and hot spots of sentiment in a given time interval (Rey, 2019). Sentiment could also be linked to other intrinsic characteristics of POIs and their geographical location. For example, the restaurant's price or the rating associated with the services, cleanliness, or type of cuisine influences tourists' opinions.

Finally, a positive trend in sentiment indicates an optimal reception of the services offered in the area. In promoting these processes, the European Commission is interested in observing and assessing civic engagement when selecting European Capitals of Culture (ECoCs) applications and implementing effective evaluation and long-term monitoring mechanisms. Synergies of this kind are useful to verify that the participation of all local energies remains constant to support the public administration in maintaining the real involvement of the local population in the long term and not only in the short period (Demartini *et al.*, 2020).

References

ANSELIN L., 1995. Local Indicators of Spatial Association - LISA, *Geographical Analysis*, Vol. 27, No. 2, pp. 93–115.

BALBI S., MISURACA M., SCEPI G. 2018. Combining different evaluation systems on social media for measuring user satisfaction, *Information Processing & Management*, Vol. 54, No. 4, pp. 674–685.

BOLASCO S., DELLA RATTA F. 2004. Experiments on semantic categorisation of texts: analysis of positive and negative dimension. In PURNELLE G, FAIRON C. and DISTER A. (Eds.) *Le poid des mots. Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles*, UCL Presses Universitaires de Louvain, pp. 202–210.

CHIAVETTA F., LO BOSCO G., and PILATO G. 2016. A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language. In MAJCHRZAK T.A, TRAVERSO P., MONFORT V. and KREMPELS K.-H. (Eds.), In *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, Scitepress, pp.159–170.

CORINTO G. L., NICOSIA E. 2016. The European Capital of Culture. Will Matera 2019 be a successful example? *In development, art(s) and culture UNEEC forum*, Vol. 8, Lucian Blaga University of Sibiu Press, pp. 19–35.

DEMARTINI P., MARCHEGIANI L., MARCHIORI M. 2020. Culture Invites Participation. An Inquiry on Matera as European Capital of Culture 2019. In PIBER M. (Ed.) *Management, Participation and Entrepreneurship in the Cultural and Creative Sector*. Berlin: Springer, pp. 161–186.

GETIS A., ORD J. K. 1992. The Analysis of Spatial Association by Use of Distance Statistics, *Geographical Analysis*, Vol. 24, No. 3, pp. 189–206.

GUALA C. 2002. Towards A typology of Mega-Event, *Bollettino della Società Geografica Italiana*, Vol. 7, No. 4, pp. 743–755.

HU T., SHE B., DUAN L., YUE H., CLUNIS J. 2020. A Systematic Spatial and Temporal Sentiment Analysis on Geo-Tweets, *IEEE Access*, Vol. 8, No. 1, pp. 8658–8667.

IVONA A., RINELLA A., RINELLA F. 2019. Glocal Tourism and Resilient Cities: The Case of Matera "European Capital of Culture 2019, *Sustainability*, Vol. 11, No. 15, pp. 4118–4130.

LEMMI E. 2009. *Dallo «spazio consumato» ai luoghi ritrovati. In Verso una Geografia del Turismo Sostenibile*. Milan: Franco Angeli.

NGUYEN T. Q. T., JOHNSON P., YOUNG T. 2022. Networking, coopetition and sustainability of tourism destinations, *Journal of Hospitality and Tourism Management*, Vol. 50, pp. 400–411.

POLLICE F. 2014. *Territori del turismo. Una Lettura Geografica delle Politiche del Turismo*. Milan: Franco Angeli.

POLLICE F., URSO G. 2014. Turismo vs. globaritarismo. In TURCO A. (Ed.), *Filiere etiche del turismo. In Territori della Vacanza tra Valori, Politiche e Mercati*, Unicopli, pp. 75–96.

REY S., 2019. Spatial Dynamics and Space-Time Data Analysis. In: FISHER M., and NIJKAMP P. (Eds.), *Handbook of Regional Science*, Berlin: Springer, pp. 1–18.

SENECAL S., NANTEL J. 2004. The influence of online product recommendations on consumers' online choices, *Journal of Retailing*, Vol. 80, No. 2, pp. 159–169.

STEFANIDIS A., CROOKS A., RADZIKOWSKI J. 2013. Harvesting ambient geospatial information from social media feeds, *GeoJournal*, Vol. 78, No. 2, pp. 319–338.

VECHTOMOVA O. 2017. Disambiguating context-dependent polarity of words: An information retrieval approach, *Information Processing & Management*, Vol. 53, pp. 1062–1079.

WESTBROOK R. A. 1987. Product/consumption-based affective responses and postpurchase processes, *Journal of Marketing Research*, Vol. 24, No. 3, pp. 258–270.

ZHONG C., ZENG S., TU W., YOSHIDA M. 2018. Profiling the Spatial Structure of London: From Individual Tweets to Aggregated Functional Zones, *International Journal of Geo-Information*, Vol.7, No.10, pp. 386–400.

Luigi CELARDO, University of Naples Federico II, luigi.celardo@unina.it
Michelangelo MISURACA, University of Calabria, michelangelo.misuraca@unical.it
Maria SPANO, University of Naples Federico II, maria.spano@unina.it

SOCIAL AND DEMOGRAPHIC PREDICTIONS IN INNER AREAS: A COMPARISON BETWEEN CAMPANIA AND SICILY[1]

Cira Acampora, Antonella La Faci,
Valeriana Leporanico, Massimo Potenzieri, Matteo Potenzieri

Abstract. Demographic forecasts aim to chart the likely future of a population in terms of size and structural indicators. For policy they are a useful tool to understand demographic trends and plan strategies of an area. The interest for the territorial dimension has always characterized the activity in the field of demographic forecasts so that, since November 2021, the Italian National Institute of Statistics (ISTAT) has published demographic forecasts at 10 years, by age groups and municipalities with a population of over 20,000. The study aims to compare the demographic structure and the population movement in 2031 in the municipalities of Campania and Sicily classified according to the National Strategy for Internal Areas (SNAI) which aims to develop and territorial cohesion (Barca *et al*., 2014). The choice of the regions of Campania and Sicily is determined by the fact that both are the most populous in the Mezzogiorno and have well characterized and different internal areas. Work objectives, given the real diversity in territorial behaviours and demographic characteristics, are the identification of primary determinants of marginalization compared to Italy, and a study on the ageing of the population and the decline in birth rates that will influence the contraction of working-age residents in the next decade.

1. Introduction

About four thousand municipalities for a population of over thirteen million people with an occupied area equal to fifty-nine percent of the national territory, these are the numbers of the Internal Areas in Italy so called because they are significantly distant from the centers of offer of essential services or those related to education, health, mobility. However, these are areas rich in significant environmental and cultural resources and therefore to be considered the subject of a policy of development and territorial cohesion with a view to combating marginalisation and the "demographic desert". Investments on the protection and

[1]Authors of the sections: 1. Matteo Potenzieri; 2.-2.1. Valeriana Leporanico; 3. Cira Acampora; 4.- 4.2. Antonella La Faci; 4.1 Massimo Potenzieri.

promotion of the territory and enhancement of environmental and cultural resources are therefore the cornerstones on which the national strategy for Inner areas (SNAI) was based - promoted in 2013 by the Agency for Territorial Cohesion - with a view to creating new job opportunities (De Rossi, 2019). The prediction of demographic phenomena has always been fertile ground for the analysis of territories. Tracing the likely future of a population on the basis of assumptions of mortality, fertility, migration and structural components is indeed of the utmost importance for the implementation of new policies in the transport network, in the health sector, social security, as well as urban, energy and environmental needs. (Dip. politiche di coesione. 2022a; 2022b).

Table 1 – *Population and percentage of the total population by Areas – Campania, Sicily, and Italy – 2021.*

Classification AI		A Urban poles	B Intermunicipal poles	C Belt	D Intermediate	E Peripheral	F Ultraperipheral	Total	A+B+C (Centers)	D+E+F (Inner Areas)
Number of municipalities	Italy	182	60	3,828	1,928	1,524	382	**7,904**	**4,070**	**3,834**
	Campania	11	4	245	125	123	42	**550**	**260**	**290**
	Sicily	12	3	66	119	157	34	**391**	**81**	**310**
Population	Italy	20,325,781	1,573,512	23,733,409	8,044,378	4,636,615	716,438	**59,030,133**	**45,632,702**	**13,397,431**
	Campania	1,537,749	186,509	2,931,439	510,749	377,773	80,201	**5,624,420**	**4,655,697**	**968,723**
	Sicily	1,666,404	91,868	763,426	1,151,585	1,058,683	101,739	**4,833,705**	**2,521,698**	**2,312,007**
%	Italy	34,4	2,7	40,2	13,6	7,9	1,2	**100**	**77,3**	**22,7**
	Campania	27,3	3,3	52,1	9,1	6,7	1,4	**100**	**82,8**	**17,2**
	Sicily	34,5	1,9	15,8	23,8	21,9	2,1	**100**	**82,8**	**17,2**
Surface	Italy	24,455,1	3,727,5	96,345,0	75,837,0	79,393,6	22,310,1	**302,068,0**	**124,528,0**	**177,541,0**
	Campania	662.0	171,5	3,757.8	3,811.9	3,990.7	1,276.6	**13,670.6**	**4,591.0**	**9,079.0**
	Sicily	2,561.8	320.5	3,409.2	6,584.5	10,749.8	2,206.7	**25,832.5**	**6,291.5**	**19,541,0**
%	Italy	8,1	1,2	31,9	25,1	26,3	7,4	**100.0**	**41,2**	**58,8**
	Campania	4,8	1,3	27,5	27,9	29,2	9,3	**100.0**	**33,6**	**66,4**
	Sicily	9,9	1,2	13,2	25,5	41,6	8,6	**100.0**	**33,6**	**66,4**

Source: Our data processing from Istat

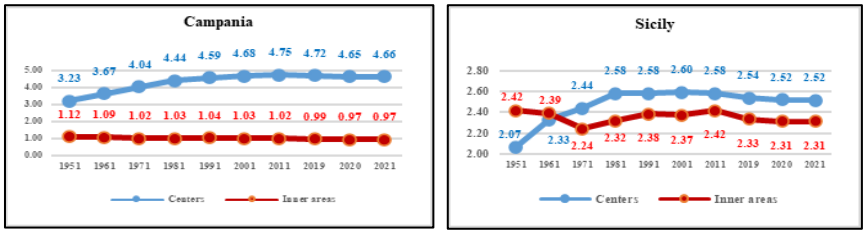
In this perspective and given the growing interest in the forecasts applied to the most "micro" territorial levels, we wanted to photograph the population structure of the inner areas and centers, both in 2021 and 2031, two regions of the Mezzogiorno of particular interest both for their demographic size and for the significant differences that emerge from the comparison between their inland areas: Campania and Sicily. The assessment of the main determinants of marginalisation in contexts of ageing populations and depopulation of inland areas, albeit similar to the national framework, highlights marked differences - see, for example, structural indices or indicators of migration flows - which make the study particularly attractive and indicative for cohesion policies which, taking their cue from the heterogeneity of territories; aim to reduce the gaps between declining inland areas and population risk areas and Centres. In a discussion based on demographic and social variables, it was chosen to operate - not casually - on short-term time forecasts (2031) mainly because the emphasis was placed on spatial disaggregation at the municipal level. Notoriously, in fact, in the forecast field, the more the temporal horizon is amplified the greater is the uncertainty of the information. The area of Belt is prevalent in

Campania and Italy, although with significantly different percentages (52% and 40%). The Urban Pole stands out in Sicily (34%). Very similar percentage population (between 1 and 3%) in the outermost areas and the inter-municipal Poles for all the territories considered. Demographic breadth comparable between Intermediate and Peripheral area of Campania and Sicily although with very different percentages (7% and 9% Campania, 22 and 24% Sicily). The Inner areas in Italy account for 23% (with predominant Intermediate area - 14%), against 48% in Sicily (predominant Intermediate area - 24%) and higher than 5% in Campania (prevailing Intermediate area: 9%) (Table 1).

2. Population dynamics

The dynamics of the resident population in the last 10 Censuses show how the difference between the two regions is particularly significant, as you can imagine, between 1951 and 2021 the resident population decreases in the Inner Areas and increases in the Centers, in particular in Campania, it decreases of -13.3% since 1951 and in Sicily of -4.4% but it is to notice that until 1961 the resident population in the Inner areas was greater in the Centers (Figure 1).

Figure 1 – *Resident Population at Permanent population census from 1951 to 2021, by area (millions of people).*

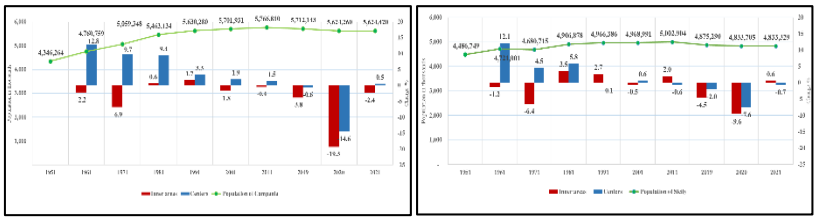


Source: General population Censuses 1951-2011 and permanent 2018-2021.

Observing the trend of the population on the individual decades and how much the population changes in annual average, we can say that in Campania in the last 70 years the resident population has increased, especially between 1951 and 2011 the residents in Campania increased by over 1 million and 400 thousand units, with an average annual growth rate (+4.7%). The decreasing trend starts in 2019 when between 2020 and 2019, in the Centers, there is an average annual decrease rate of -14.6 per thousand residents, 4.7 points lower than the equivalent of the Inner areas (-19.3 per thousand) which corresponds to the period of greatest demographic decline. In Sicily in the last 70 years the resident population is almost stationary but

from 2019 that corresponds to the period of greatest demographic decrease, is significant the reduction of the resident population in the Centers (-7.6 per thousand residents) that in the Inner areas (-9.6 per thousand). In 2021, although slight, there is a population growth in Inner areas, equal to 0.6 per thousand residents while in the Centers there is a decrease in population (-0.7 per thousand residents) (Figure 2).

Figure 2 – *Resident population in Campania and Sicily (left axis) and annual average changes[2] for 1,000 residents in the Inner Areas and Centers (right axis).*

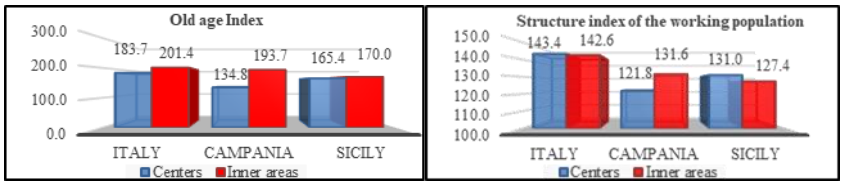


Source: General population Censuses 1951-2011 and permanent 2018-2021.

2.1. Structure of population: permanent censuses

The socio-demographic structure indicators show significant differences between Centers and Inner areas of Campania and Sicily compared to the national territory. In particular, the old age index shows the degree of ageing of the population especially in Inner areas but while in Sicily the difference compared to the Centers is less marked and is equal to 170 elderly per 100 young people, registering a difference of 4,6 percentage points compared to the Centers, in Campania is particularly evident and in the Inner areas there are well over 193 over-65s per 100 young people compared with 134.8 in the Centers (a difference of 58.9 percentage points).

Figure 3 – *Old age index and Structure index of the working population (100 inhabitants).*



Source: Istat, Permanent population Census, year 2021.

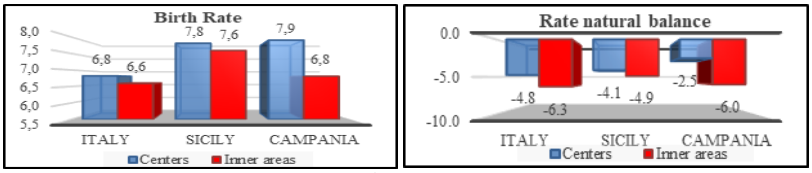
[2] Annual geometric (or compound) mean change.

The ageing of the Italian population is even more evident in comparison with past censuses: The old age index in 70 years increased from 33.5 in 1951 to 187.6% in 2021, of which 201 elderly people only in Inner areas. The Structure of the Working Population Index represents the degree of ageing of the working-age population, in relation to the older (still active) generations that will replace them, it is always above 100. An indicator higher than 100 means that the population of working age is very old and as a must in Figure 3 is greater in the Centers than the Inner areas in Italy and Sicily while in Campania, the index is higher in the Inner areas with a difference of almost 10 percentage points compared to the Centers.

3. Population dynamics: the main demographic balance sheet indicators

Both regions have a much higher birth rate than the national average (7.7 against 6.8) and this characteristic is reflected almost exclusively in the areas of the Centers, more significant, however, is the difference between the Inner areas, where in Sicily the birth rate exceeds by 0.8 percentage points that of Campania.

Figure 4 – *Birth rate and Rate natural balance (1000 inhabitants).*

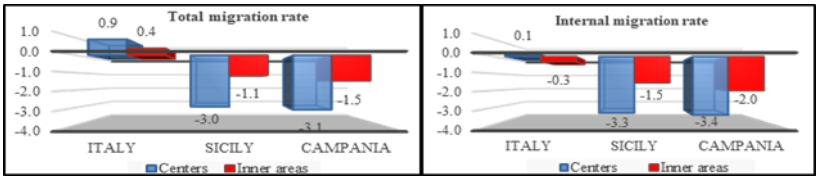


Source: Istat, Movement and calculation of annual resident population, year2021.

Strong is the demographic crisis throughout Italy, the number of births is much lower than deaths. The natural increase in population is particularly low in the Inner areas and Sicily and Campania, although it is better than the national average, due to a higher birth rate, have, however, very low values and the difference is particularly significant in the Inner areas of Campania (Figure 4). Campania and Sicily have a negative total migration rate, a sign of a lack of dynamism in both regions, which differ considerably from the national average, which, for the same indicator, positive values with a slightly higher trend in the centers. Specifically, it emerges that the number of cancellations resulting from the transfer of residence is greater in the areas of the center than in the internal areas and this difference is more marked in Campania with a difference of 1,7 percentage points from Inner areas. The same dynamic emerges for the internal migration rate that for both regions is negative with a prevalence in the centers compared to the Inner areas, especially in Campania. The migratory rate with foreign countries is positive everywhere, but both Campania and

Sicily continue to show a lower attraction capacity than the national figure. In Campania the migratory rate with foreign countries is lower than in Sicily with a difference of 0.5 percentage points and is much more accentuated in Inner areas than in the Centers; Sicily, on the other hand, is much more homogeneous internally and is closer to the national average, with a very slight accentuation in the centers (Figure 5).

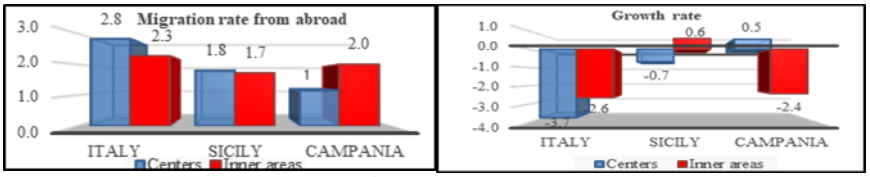
Figure 5 –*Total migration rate and Internal migration rate (1000 inhabitants).*



Source: Istat, Movement and calculation of annual resident population, year2021.

As regards the total growth rate, the two regions are rather uneven and diverge from the development of the Italian territory that has a sharp decrease in population, especially in the central areas, albeit slightly offset by a positive migration rate. For Sicily it is positive in the internal areas and negative in the centers, for Campania, instead, it is the opposite, it is positive in the Centers and strongly negative in the internal areas but in line with the national one (Figure 6).

Figure 6 – *Migration rate from abroad and Growth rate (100 inhabitants).*



Source: Istat, Movement and calculation of annual resident population, year2021.

4. Demographic forecasts: 1 January 2021-2031

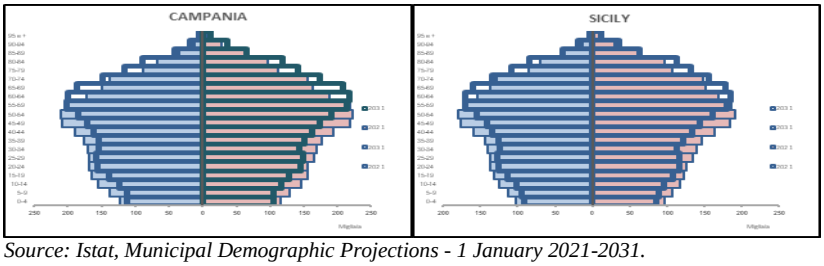
Demographic forecasts aim to chart the likely future of a population in terms of total size and structural components (Marsili, 2007). Many and diversified uses of this data by policy makers: from the field of health planning to social security, from the study of urban needs to energy-environment, from the organization of school facilities to the transport network. Being able to produce detailed spatial information is essential to analyze and understand the same national demography. However, the level of regional detail is insufficient when the planning of a territory needs micro-

based information. This led Istat to develop a new product in the field of demographic forecasting that shifts the details to the level of the municipality. The published data concern the provincial capital municipalities and all municipalities that exceed 20 thousand inhabitants. Data for consolidated supra-municipal aggregations such as SLL and SNAI are available for all other municipalities. As for this study, the forecasts are experimental and start from 2021 until 2031, as in the field of forecasting, the wider the time horizon, the more uncertain the information. The evolutionary hypotheses about the future demographic trend in the municipalities concerning fertility, survival and migration in the different municipalities are derived from regional forecasts, according to a probabilistic and redistribution approach from top to bottom. The comparisons that are presented in this paragraph have been made both using the data of the demographic database "Demo - Demography in figures", and those offered by the Experimental Statistics. The updates concern the replacement of the population per base year with the latest yield available from the Census and a partial revision of the evolutionary assumptions of the model, in line with what was the revision produced for regional and national forecasts.

4.1 Campania and Sicily, a future in comparison

The new forecasts on the demographic future of the two regions under analysis confirm the presence of a potential crisis as in the rest of the country. In the decade 2021-2031, the total population decreased: in Campania from 5,624,420 to 5,400,000 inhabitants and in Sicily from 4,833,329 to 4,600,000. The new forecasts highlight the progressive aging of the population that decreases particularly until the age of 59 years increasing from 60 years and up, as can be seen from the age pyramid (Figure 7). The population of 65 years and more today represents 19.9% of the total in Campania and will reach, according to Istat forecasts, 25.2% in 2031 (2 percentage points less than the average in Italy); the population up to 14 years is 14.3% and will be 12.5% in 2031; in the 15-64 age group from 65.8% will rise in 2031 to 62.2%.

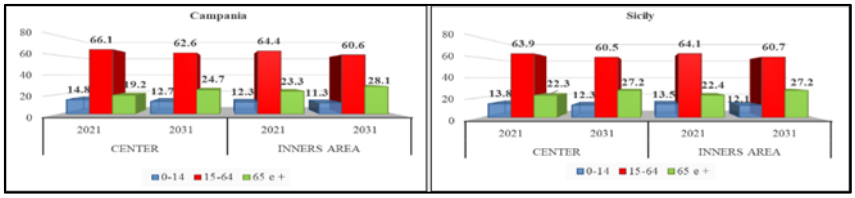
Figure 7 – *Campania and Sicily. Pyramid of Age. Comparison 2021-2031.*



Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

In the scenary of 2031, in Campania, the population over 65 years will be concentrated mainly in the Inner Areas with a difference of 3.4 percentage points more than the Centers, while the under 14 and the working age population more in the Centers (+1.4 percentage points). In Sicily the population will be equally distributed (27.2%). In the temporal comparison, the same band of population of Campania will increase more in the centers (+5.5) than the internal areas (4.8%); for Sicily the difference between Centers and Inner Areas is minimal (0.1 percentage points), while in 2031, the population under 14 years will decrease from 13.8% to 12.3% in the Centers, and that in working age will rise from 64%, to 60.5% (Figure 8).

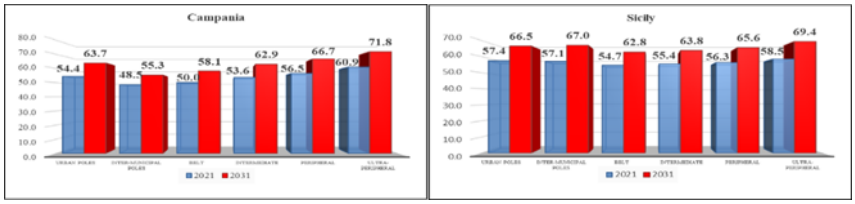
Figure 8 - *Percentage composition of population in SNAI areas by age group. Period 2021-2031.*



Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

The structural dependency ratio (Figure 9) expresses the theoretical social and economic burden of the population of working age; values above 50% indicate a situation of generational imbalance which in Campania, are present in all areas except for inter-municipal areas that are also those where this value increases to a lesser extent (+6.8%). The highest percentage is found in the outermost areas which are also those where the IDS increases the most (+10.9%). In Sicily, values above 50% are present in all SNAI areas with a lower ratio in Centres (54.7%) and a peak in Ultra-peripheral areas (58.5%) where there will be, in 2031, the most significant increases.

Figure 9 – *Structural dependence index in SNAI areas -period 2021-2031.*



Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

4.2. Some hypotheses observing the variations on the differences

Observing the variations in relation to the differences, in terms of increase and decrease (Table 2), both on the average age (this calculated both for 2021 and for 2031 only through the experimental data) and in the comparison by age groups, we immediately notice some differences. Looking at the distribution of the population from 15 to 84 years, the ageing is evident, in particular comparing the negative sign of the classes 0-64 with that of the over 65 years: while for the class 0-14 the decrease is greater in the Centers, the phenomenon is increasing in the peripheral areas for the class 15-64 years. The positive sign, if one observes the group of people over 85 years old, testifies to the ageing, greater in urban areas: the answer to this phenomenon is evidently to be found in the possibility that one has, living in urban areas, access to health and care services of proximity and high professionalism.

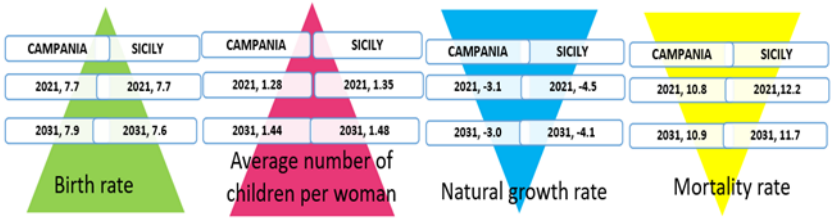
Table 2 – *Differences in regional comparison by Areas, years 2021-2031 (changes).*

	AVERANGE AGE		POPOLATION 0 -14 YEARS		POPOLATION 15-64 YEARS		POPOLATION 65 YEARS AND OVER		POPOLATION 85 YEARS AND OVER	
	Campania	Sicily	Campania	Sicily	Campania	Sicily	Campania	Sicily	Campania	Sicily
Urban poles	2.8	2.5	-1.5	-1.2	-3.7	-3.5	5.2	4.7	0.8	0.7
Inter-municipal poles	3.4	2.5	-2.3	-0.8	-3.1	-3.7	5.4	4.5	0.8	0.9
Belt	3.3	2.6	-1.8	-1.4	-3.5	-3.2	5.3	4.6	0.8	0.6
Intermediate	2.4	2.6	-0.9	-1.3	-3.7	-3.3	4.6	4.6	0.2	0.6
Peripheral	2.4	2.4	-0.9	-1.0	-3.9	-3.6	4.8	4.6	0.4	0.4
Ultra-peripheral	1.7	2.3	-0.3	-0.7	-3.8	-4.1	4.1	4.8	0.0	0.0
TOTAL	2.9	2.5	-1.5	-1.2	-3.6	-3.4	5.1	4.7	0.7	0.5

Source: Istat, Municipal Demographic Projections - 1 JANUARY 2021-2031.

Recalling the decrease in population of 200,000 units and observing instead the increase in the birth rate and the decrease in that of mortality in both regions, the decline of the population in absolute numbers in both is evidently also caused by other factors (Figure 10).

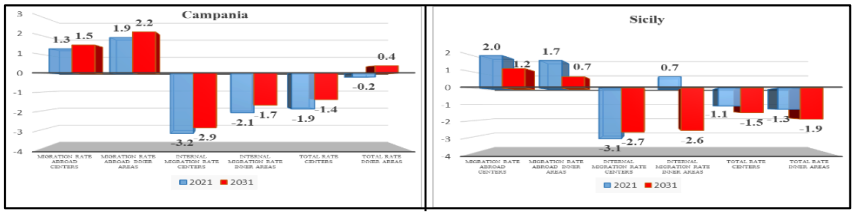
Figure 10 – *Regional comparison and rates, 2021-2031 (changes).*



Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

The rate of migration from abroad is always positive, especially for the Internal Areas, both in Campania and Sicily, with a growth trend, from 2021 to 2031, of 0.3 percentage points for Campania and a decrease for Sicily (-1.0). Both regions have a negative rate with regard to internal migration; in detail the phenomenon concerns more Centers than the Internal Areas both in 2021 and in 2031 where in Campania the value will be -2.9 while in Sicily -2.7 (Figure 11).

Figure 11 – *Migration flows 2021-2031 (percentages). Campania and Sicily.*



Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

5. Discussions and conclusions

The expected demographic decline in the coming years is mainly attributable to migratory flows, which in the forecasts show an increase in the exit from regional and state borders of a large part of the population and diversified by age group. However, a separate discourse must be made in the comparison between the real data of 2021, which are the result of a period of pandemic crisis, and the forecasts of both the 2021 and 2031 projections. In the case of Campania, in fact, considering only the projections, it can be seen that in the face of both internal (3.5‰) and foreign immigration (3.2‰), there is a much higher share than that, on the contrary, it leaves the region mainly to settle in other Italian regions (6.2‰). For Sicily, immigration from abroad decreases (-0.3), the rate of internal immigration increases by +0.1 as well as emigration abroad by 0.5 while that to other regions decreases -0.3‰.The decline of the population in 2031 compared to 2021, whether we look at the observed data or the forecast, is instead substantially "slowed down" because, in the face of migratory changes, there is a slight increase in the value of the birth rate and a moderate reduction in the death rate, albeit to a different extent between the two regions. One possible explanation may be that the elderly population leaves the inner areas more easily because it moves in search of more numerous and efficient services, especially from the social welfare point of view; young people, Instead, they look to areas that offer better and more opportunities for study and work and therefore opt for transfers outside regional and national borders (Table 3).

Table 3 – *Forecast and real data. Regional comparison, migration rates and total balance.*

	Data 2021	2021 data of forecast	2031 data of forecast	Difference between forecast data
Campania				
Popolazione*	*5,624,420*	*5,624,420*	*5,384,914*	*-4,3*
Birt rate	7.7	7.7	7.9	0.2
Mortality rate	10.8	11	10.9	-0.1
Natural growth rate	*-3.1*	*-3.3*	*-3*	*0.3*
Rate Immigration from abroad		2.8	3.2	0.4
Emigration rate for foreign countries		1.3	1.6	0.3
Net migration rate abroad	1.3	1.4	1.6	0.2
Rate Immigration from the inside		3.5	3.5	0.0
Emigration rate for the interior		6.5	6.2	-0.3
Net migration rate with the interior	-3.2	-3	-2.7	0.3
Total net migration rate	3.1	-6.1	-1.1	5.0
Total growth rate	*0.0*	*-4.2*	*-4.1*	*0.1*
Sicily				
Popolazione*	*4,833,329*	*4,833,329*	*4.583.054*	*-4,8*
Birt rate	7.7	7.7	7.6	-0.1
Mortality rate	12.2	12.6	11.7	-0.9
Natural growth rate	*-4.5*	*-4.8*	*-4.1*	*0.7*
Rate Immigration from abroad		3.7	3.4	-0.3
Emigration rate for foreign countries		1.9	2.4	0.5
Net migration rate abroad	1.8	1.8	1	-0.8
Rate Immigration from the inside		3.1	3.2	0.1
Emigration rate for the interior		6.1	5.8	-0.3
Net migration rate with the interior	-2.5	-3	-2.6	0.4
Total net migration rate	4.4	-1.2	-1.7	-0.5
Total growth rate	*-0.1*	*-6*	*-5.8*	*0.2*

** The data reported in the first two columns, under the heading "Population" coincide as they refer to the legal population in Campania and Sicily, defined on the basis of the Census, on 31 December 2021, as published in the G.U. on 03/03/2023.*
Source: Istat, Municipal Demographic Projections - 1 January 2021-2031.

Appendix

Old age index: ratio between the population aged 65 and over and the population aged 0-14, multiplied by 100.

Structure index of the working population: percentage ratio of population aged 40-64 to population aged 15-39, multiplied by 100.

Natural growth rate: difference between the birth rate and the death rate and is calculated by comparing the natural balance of the year to the population at 31/12 of the same year x 1,000).

Total growth rate: sum of the natural growth rate and the total migration rate and is calculated as the ratio of population change in a given year (difference between population at 31 December and 1 January) and the average population of that year per thousand individuals).

Migration rate with foreign countries: ratio between the migration balance with foreign countries of the year and the average amount of the resident population, per 1,000.

Internal migration rate: ratio between the internal migration balance of the year and the average amount of the resident population, multiplied by 1,000.

Total migration rate: ratio between the year's migration balance and the average amount of the resident population, multiplied by 1,000.

Birth rate: ratio between the number of live births of the year and the average amount of the resident population, multiplied by 1,000.

Acknowledgments

A special thanks for our colleague Istat of the Basilicata Office, Agata Maria Madia Carucci for her scientific contribution.

References

BARCA F., CASAVOLA P., LUCATELLI S. 2014. *A Strategy for Inner Areas in Italy: Definition, Objectives, Tools and Governance,* Materiali Uval Issue n. 31.

DE ROSSI, A. 2019. *Riabitare l'Italia: le aree interne tra abbandoni e riconquiste*. Roma: Donzelli editore.

DIPARTIMENTO PER LE POLITICHE DI COESIONE.2022a. *Aggiornamento 2020 della mappa delle aree interne. Nota tecnica NUVAP*.

DIPARTIMENTO PER LE POLITICHE DI COESIONE. 2022b. *Criteri per la Selezione delle Aree Interne da sostenere nel ciclo 2021 – 2027*.

ISTAT, Municipal Demographic Projections - 1 January 2021-2031 - Experimental Statistics, 8 August 2022, Roma, https://www.istat.it/en/archivio/273736.

MARSILI M. (2007), *Demographic projections: the impact of net international migration on population ageing in Italy*, Atti del Convegno Intermedio della SIS 2007 "Rischio e Previsione", Università Ca' Foscari, Venezia, 6-8 giugno. 47.

Cira ACAMPORA, Istat, acampora@istat.it
Antonella LA FACI, Istat, lafaci@istat.it
Valeriana LEPORANICO, Istat, valeriana.leporanico@istat.it
Massimo POTENZIERI, Istat, mapotenz@istat.it
Matteo POTENZIERI, Istat, potenzie@istat.it

A PROTOTYPE OF A VISUALIZATION SYSTEM TO EXPLOIT DATA ON THE POPULATION AGEING IN ITALIAN URBAN CONTEXTS [1]

Donatella Cangialosi, Giuseppe Lancioni, Debora Tronu, Daniela Vacca

Abstract. There is a strong request to improve the easy of interpretation of the official statistic information. The request is backed by a number of international organizations and is intended to the advantage of decision makers, as well as researchers and non-specialist users. Modern tool for graphical and geographical data visualization offer resources to implement systems suited for the scope. The aim of this work is to perform a comparative analysis between different technological platforms, and to find an effective trade-off between the completeness of the traditional, official statistics and the immediate awareness of a phenomenon given by the graphical and geographical data visualization technologies. The study is performed using two dashboard prototypes implemented with two benchmark platforms: *Tableau* and *ArcGIS Dashboards*. Prototypes consist of the very same graphical objects and share a similar layout, however showing a different look and feel, and cover complementary aspects of the statistic information. Both have been extensively tested and proved effective in giving an immediate and consistent comprehension of the phenomenon on study.

1. Introduction

International recommendations on communication and dissemination strategies underline the importance of ensuring and extending accessibility to data produced by official statistics and promoting an informed use of data for correct interpretation and analysis of phenomena (Eurostat, 2021a; Unece, 2021; Oecd, 2021). In this context, the use of data visualization tools is essential to improve the dissemination and promote statistics, facilitating their use by policy makers, researchers and non-specialist users.

Advances in the data visualization topic has been firstly explored and driven by research efforts in Business Intelligence (Rouhani (2012)), with the aim of

[1] Authors contributors: D. Cangialosi, D. Tronu, D. Vacca the revision of the whole article; summary to all the authors, § 1, 2 and 3 to G. Lancioni.

improving comprehension and efficiency of the business processes; see for example Dresner (2007), Eigner (2013), Kumar (2017). More in general, Ward *et al*. (2015), Sadiku *et al*. (2016), Qi *et al*. (2020), trace the lines of the field of data visualization and discuss methodologies and techniques. Diamond *et al*. (2017), Chapman (2018), Furmanova *et al*. (2020), provide a comparison among state of the art visualizations tools, including Tableau (Salesforce, 2023), Microsoft Power BI (Microsoft, 2023), Google Charts (Google, 2023), Infogram (Infogram, 2023), to limit the list to those with geographical mapping support ad whose usage does not require accurate programming skills. Müller (2003) is indeed devoted to time-series data representations. All of the cited tools are illustrated and assessed, and examples of their use are shown. Srivastava (2023) extends the discussion to various domains, such as business, social sciences, humanities, sports, environmental sciences, and healthcare, while Grainger *et al.* (2016) and Bujack (2020) analyse the visualization of environmental science data. Given the nature of such data, geographical representation is an essential asset; moreover, a particular attention is devoted to the fact that some of the involved actors could be outside the scientific community, so stressing the search of effectiveness and easy of comprehension for the discussed approaches. An increasing number of dashboards showing very effective and interactive data visualizations have been published in the web. Restricting the scope to demography and healthcare, maybe the most popular is the 'Covid-19 Dashboard' by the Johns Hopkins University reviewed by Kelly (2020), and developed with ArcGIS Dashboards (Esri, 2023). Other valuable efforts are 'Age Structure' by Ritchie *et al.* (2019), built with dedicated owner software; 'Ageing and Health in the Americas' by PAHO (2019), on Tableau platform; and 'Interactive map generator' by Eurostat (2021b), developed with dedicated software which requires IT skills to be reused. The idea of designing a data visualization tool stems from a previous work, in which an analysis was developed on the profiles of metropolitan cities (Istat, 2023). The main characteristics, diversities and common factors of these territories were identified by means of a set of indicators including some socio-demographic ones, so making it possible to measure ageing degree for populations residing in urban contexts. This set of indicators represents the basis on which he dashboard prototypes have been built. Indeed, the demographic trends of ageing are a global phenomenon, and Italy is one of the "oldest" countries in the world. Since the 1970s, demographic analysis has paid attention to the social and economic consequences of this phenomenon (Golini, 1997). This paper consists in a comparative study between geo-statistical tools for visualizing and reading data applied to the analysis of ageing in metropolitan cities through a selection of the aforementioned socio-demographic indicators. The study is carried out on two prototypes of dashboards developed with two different technologies for graphical and geographical data representation. In light of the tools comparisons discussed in the cited references, and driven by the

geographical nature of our data, and finally aiming at easy-to-understand and self-consistent representations, we choose two state of the art platforms: Tableau and ArcGIS Dashboards. Prototypes are not yet publicly accessible since final evaluation and validation is still underway. The aim is not to make a ranking of the different approaches, rather to find an effective trade-off between the completeness and precision of the traditional statistical information, and the immediate awareness allowed by the graphical data fusion and integration supported by the geo-statistical tools. The remainder of the paper is structured as follows: section 2 (Methods) is devoted to explaining the approach and to detail the layout and features of the two prototypes; section 3 (Results) outlines the results of the comparison between the dashboards and the technologies involved and assesses the general fulfilment of the targets; and section 4 (Summary) is an overall recap of this study.

2. Methods

The comparative analysis is conducted, experimentally, by comparing two different technologies for viewing statistical information, namely and ArcGIS Dashboards. They are exploited to realize systems for the representation of data that facilitate the process of knowledge and interpretation of phenomena, featuring:

- interactive visualizations that allow the exploration and reading of data, through the creation of customized graphic and cartographic objects, facilitating comparisons in the territorial domains considered;
- georeferenced statistical information storage and management systems, which allow the integration of traditional and spatial statistical analysis.

More in detail, the two mentioned technologies are leveraged to implement two different prototypes of dashboards for the visualization and integration of the statistical information. Considering the different look and feel of the two platforms, as well as their different data management features, the prototypes are designed from the beginning with two different targets:

- ArcGIS dashboard: analysis of the various aspects of a phenomenon, showing as many information as possible about it;
- Tableau dashboard: in-depth analysis of a selected, single aspect of a phenomenon, showing all its structural and territorial details.

The territorial domain used for experimentation are the 14 metropolitan cities (hence CM) in their territorial articulation in urban centres and urban belts. Statistical data consists in a selection of 33 socio-demographic indicators on the structure and dynamics of the population (Istat, 2023).

The aim of this study is to find an effective trade-off between the completeness of the traditional statistical information, represented by tabular data organized in

structured databases, and the instant data visualization and integration allowed by the graphical and geographical representation tools.

2.1. Overview

Both dashboards are one-page sized web applications, and only two graphical objects are used to represent statistical information: cartograms and histograms. This facilitates a fair comparison between the prototypes, as well as boosts the expected effect of immediate data integration and representation.

Underlying data, as well as a short guide, are available for download in excel and pdf format, and are accessible by means of web links.

Software versions are: ArcGIS Pro 2.8.0 (Maps creation and editing); ArcGIS Enterprise 10.9.1 (Dashboard creation and editing); Tableau Public 2022.3 (Dashboard creation and editing).

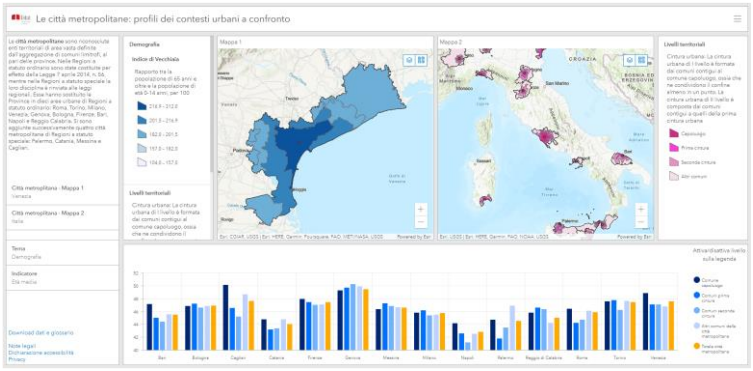
2.2. ArcGIS Dashboard

The design target of the ArcGIS dashboard is to analyse various, independent aspects of the phenomenon on study by showing simultaneously many information about it. Since all the aspects shown are virtually independent, they have the same relevance and contribute together to give to the user a complete and immediate awareness of the subject. For this reason, there is not a definite sequence of interactions, or a pattern, in the use of the dashboard.

It consists of two different cartograms on the top, based on the very same object, an *ArcGIS map* (Esri, 2019); and a histogram on the bottom. In addition, on the left side there is a short description and historical summary of the Italian CM, and a set of selectors to filter both the territory on the maps, and the indicators in the histogram. For an overview of the dashboard, see Figure 1.

Each of the cartograms shows a distinct pair of indicator and territorial detail. Also reported, on the side of the cartograms, is a short definition of the selected indicator, and a legend showing the range of its values classified in quintiles.

Figure 1 – *ArcGIS dashboard prototype.*

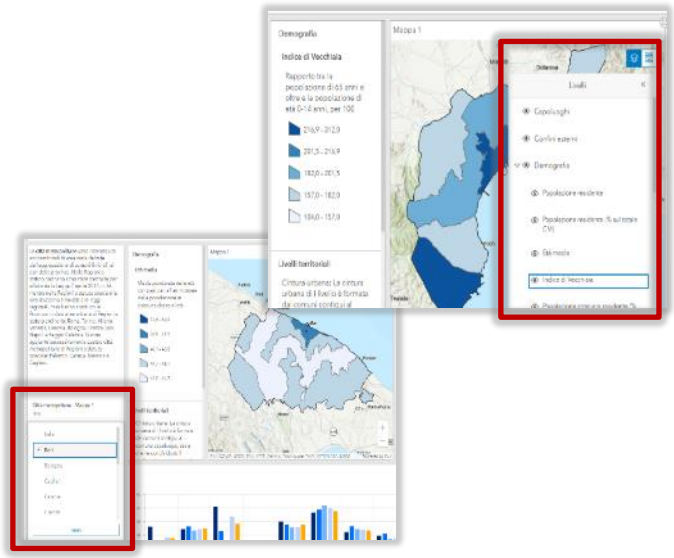


General layout.

The territorial detail, i.e. the metropolitan city, can be chosen by means of the selectors in the left side panel, see Figure 2; the default value is Italy with the full set of 14 CM. Details on the map include the urban centres and belts articulation.

The indicator to be shown is selected directly in the cartogram by clicking the *layers* icon, see Figure 2.

Figure 2 – *Selections on the map.*



Top-right: the list of the available indicators is accessible by the layers icon on the map. Clicking the eye on the left of each indicator toggles visible/not visible; only the top visible indicator is shown on the map.

Bottom-left: territorial detail is selected by means of the left side panel selectors.

Indicators are grouped by subject: demography, health, labour, and others. Each indicator is a *map layer* (Esri, 2023) that can be toggled visible or not; as per the GIS[2] tools standards and conventions, only the top visible layer is shown on the map.

By clicking a point in the map, a pop-up shows the punctual value of the indicator for that territorial detail.

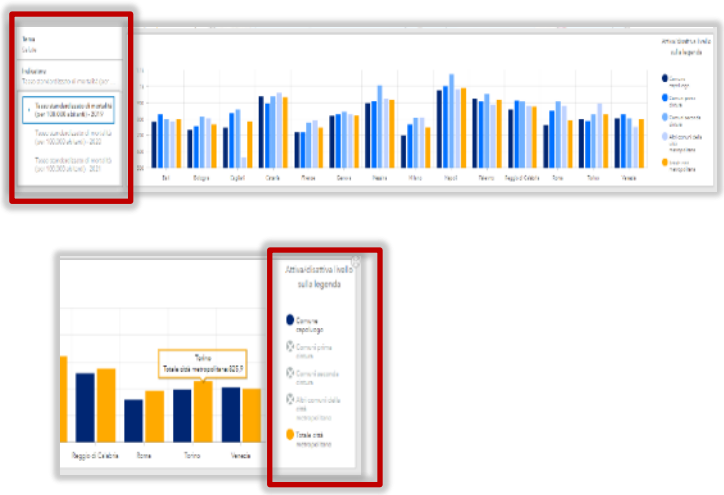
The presence of two cartograms supports the possibility of comparing, at a glance, the same indicator in two different CM; or two different indicators in the same territory; or whichever combination of (indicator, territory) pairs.

The histogram is leveraged to perform a simultaneous comparison between the 14 CM.

In this case, the indicator is chosen by means of the selectors on the left side panels, which filters them with respect to the subject they are grouped by: demography, health, labour, and others. Territorial structure of urban cities and belts is customizable by clicking the legend icons on the map, toggling visible or not the related detail, see Figure 3. Note that in the histogram the option "Total CM" is available, showing the value for the entire CM. This value cannot be visualized in the maps, due to aggregative structure of the belt levels.

Punctual value of the columns are available via pop-up, which activates on mouse moving.

Figure 3 – *Histogram.*



Top: the list of the available indicators is accessible by the left side panel selectors.
Bottom: clicking the icons in the legend toggles the related detail visible or not visible.

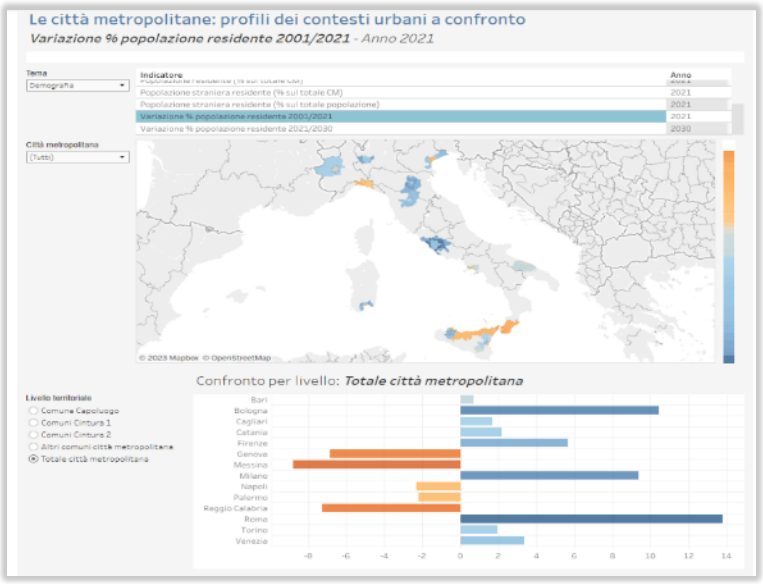
[2] Geographic Information System

2.3. Tableau

The rationale of the Tableau prototype is to carry out an in-depth analysis of a single, selectable aspect of the phenomenon on study. The whole page is therefore *indicator-driven*, and the choice of the indicator on top of the dashboard leads the interpretation of the whole dashboard. This means that there is pattern in the visualization: a sequence of interactions moving from the top to the bottom of the dashboard.

The general layout consists of a set of selectors on the top, which allows the choice of the indicator, filtered by the subject. Then there is a cartogram in the middle, and a histogram on the bottom. A view of the general layout of the dashboard is in Figure 4. All the graphical and interactive objects are conceptually identical to the ones employed in ArcGIS Dashboard prototype, although they have a different look and feel.

Figure 4 – *Tableau prototype - General layout*



As said, the main interaction is the selection of the indicator on top. Once selected, the particular aspect it describes is represented in both the cartogram and the histogram.

The geographical distribution is portrayed in the cartogram, or map. The territorial detail, i.e. the specific CM, can be selected by means of the selector on the left of the map itself. Default visualization is the map of Italy with the 14 CM. The territorial structure in urban cities and belts is also shown, as well as a legend in the right; the legend rescales to cover the values represented in the geographical detail (i.e., the CM) of the map.

In the histogram the indicator is represented simultaneously for all the 14 CM; by the options button on the left it is possible to choose one belt aggregation, or the urban cities, or the aggregate value for the entire area of the CM. In this case, differently from the ArcGIS prototype, the choice is exclusive and one only level detail can be selected.

On both map and histogram a pop-up shows the detailed information on mouse moving on it.

3. Results

Analysis of the results after an extensive session of tests on both dashboard prototypes proofs that they allow an integrated reading of the data through thematic and territorial representations of a selection of indicators.

The design choices taken in the planning phase proven crucial in obtaining the expected trade-off between completeness and instant comprehension of the phenomenon on study. In particular, the one-page layout and the use of only a limited number of graphical objects (namely cartograms and histograms) stresses the search for the ease of use and the immediacy of the information presented. On the other side, the general awareness of the phenomenon is granted by the accurate choice of the set of indicators, which covers the full spectrum of its aspects. Also to support the ease of understanding, only the last available versions of the indicators are employed; as a consequence time-series analysis, even if possible, is not supported here, and is considered out of the scope of the present study. The same project specifications prevented us to build more structured and ‘verbose’ prototypes such as the sites cited in the Introduction, see section 1. In those cases, priority is given to the completeness of information and the variety of data visualization tools; the loss is on the side of instant comprehension of the phenomenon.

The implementation and the comparative testing of the two prototypes highlights the different “look and feel” of the two technologies used:

- *ArcGIS* - It is the standard *de-facto* of Geographic Information Systems. As such, it requires the user to be aware of the usual GIS interaction standards (i.e. selection of the active layer in the map, see Figure 2). It allows implementing dashboards that are, essentially, built around one (or more)

cartogram, and in this case implementation is straightforward and spatial features can be easily filtered with selectors. However, the set of available graphical objects is limited, and interaction between them and with the cartograms is troublesome and not always possible.

- *Tableau* - A much more versatile platform, it emphasizes flexibility and ability to link data together. It makes available a plethora of graphic objects for data visualization, and it virtually enables a complete integration between them: each object can filter (and can be filtered by) any other object in the dashboard.

The prototypes are designed from the beginning to leverage these different features. The ArcGIS dashboard has the target of analyzing the various, possibly independent, aspects of the phenomenon, showing as many information as possible about it. On the other side the Tableau dashboard consists in an in-depth analysis of a selectable, single aspect of the phenomenon, integrating the information and showing all its structural and territorial details. Although based on the very same conceptual objects for visualization, and showing an overall similar layout, the two prototypes present a complementarity of the information.

References

BUJACK R., MIDDEL A. 2020. State of the Art in Flow Visualization in the Environmental Sciences. *Environmental Earth Sciences*, Vol. 79, No. 2, p. 65.

CHAPMAN C. 2018. *A Complete Overview of the Best Data Visualization Tools*, https://www.toptal.com/designers/data-visualization/data-visualization-tools; last visited 21/11/2023.

DIAMOND M., MATTIA A. 2017. Data visualization: An exploratory study into the software tools used by businesses. *Journal of Instructional Pedagogies*, Vol.18, pp. 1-7.

DRESNER H. 2007. *The Performance Management Revolution: Business Results Through Insight and Action*. Hoboken, New Jersey: John Wiley & Sons, Inc.

EIGNER W. 2013. Current Work Practice and Users' Perspectives on Visualization and Interactivity in Business Intelligence. In *17th International Conference on Information Visualisation*, London, UK, 2013, pp. 299-306, doi: 10.1109/IV.2013.38.

EUROSTAT. 2021a. *Eurostat Communication and Dissemination Strategy 2021 – 2024*, https://ec.europa.eu/eurostat/web/main/about-us/policies/communication-dissemination; Accessed, November 21, 2023.

EUROSTAT. 2021b (last available data). *IMAGE Interactive map generator*. On line Dashboard https://gisco-services.ec.europa.eu/image/screen/home; Accessed, November 21, 2023.

ESRI. 2019. Getting started with ArcGIS Pro. In ESRI *ArcGIS Pro: Essential Workflows*, Redlands, California, USA, https://www.esri.com/training/ , pp. 15-31.

ESRI. 2023. *ArcGIS Dashboards. Your information at a glance, designed for those who need it*, https://www.esri.com/en-us/arcgis/products/arcgis-dashboards/overview; Accessed, November 21, 2023.

FURMANOVA K., GRATZL S., STITZ H., ZICHNER T., JARESOVA M., LEX A., STREIT M. 2020. Taggle: Combining overview and details in tabular data visualizations. *Information Visualization*, Vol.19, No.2, pp.114-136. https://doi.org/10.1177/1473871619878085.

GOLINI A., BRUNO P., CALVANI P. 1997. *Aspetti e problemi dell'invecchiamento della popolazione*, CNR, Collana Monografie, 8.

GOOGLE. 2023. *Display live data on your site*, https://developers.google.com/chart?hl=en.

GRAINGER S., MAO F., BUYTAERT W. 2016. Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling & Software*. Vol.85, pp. 299-318, ISSN 1364-8152, https://doi.org/10.1016/j.envsoft.2016.09.004.

INFOGRAM. 2023. *Create engaging infographics and reports in minutes*, https://infogram.com/; last visited 21/11/2023.

ISTAT. 2023. *Profili delle Città Metropolitane. Molte fragilità ma anche potenzialità dei contesti urbani*, https://www.istat.it/it/archivio/280436; Accessed, November 21, 2023.

KELLY R. 2020. *Johns Hopkins Dashboard Maps Global Coronavirus Cases*, https://campustechnology.com/articles/2020/03/05/johns-hopkins-dashboard-maps-global-coronavirus-cases.aspx; Accessed, November 21, 2023.

KUMAR S. M., BELWAL M. 2017. Performance dashboard: Cutting-edge business intelligence and data visualization. In *2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, Bangalore, 2017, pp. 1201-1207, doi: 10.1109/SmartTechCon.2017.8358558.

MICROSOFT. 2023. *Uncover powerful insights and turn them into impact. Product Overview*, https://www.microsoft.com/en-us/power-platform/products/power-bi/

OECD. 2021. *OECD Report on Public Communication: The Global Context and the Way Forward*, OECD Publishing, Paris, https://doi.org/10.1787/22f8031c-en; Accessed, November 21, 2023.

MÜLLER W, SCHUMANN H. 2003. Visualization for Modeling and Simulation: Visualization Methods for Time-Dependent Data-An Overview. *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation*, pp. 737–745.

PAHO (PAN AMERICAN HEALTH ORGANIZATION). 2019 (last available data). *Data and Visualizations*. On line Dashboard: //https://www.paho.org/en/data-and-visualizations; Accessed, November 21, 2023.

QI L., LI Q. 2020. *Embodying Data: Chinese Aesthetics, Interactive Visualization and Gaming Technologies*. Singapore: Springer.

RITCHIE H., ROSER M. 2019. *Age Structure*, https://ourworldindata.org/age-structure; Accessed, November 21, 2023.

ROUHANI S., ASGARI S., SEYED VAHID MIRHOSSEINI S. V. 2012. Review study: Business intelligence concepts and approaches. *American Journal of Scientific Research* 50.1 (2012): 62-75.

SADIKU M. N. O., SHADARE A. E., MUSA S. M., AKUJUOBI C. M. 2016. Data Visualization. *International Journal of Engineering Research and Advanced Technology (IJERAT)*, vol. 2, no. 12, pp. 11-16.

SALESFORCE. 2023, *Tableau. A free platform to explore, create, and publicly share data visualizations online*, https://public.tableau.com/app/discover.

SRIVASTAVA D. 2023. An Introduction to Data Visualization Tools and Techniques in Various Domains. *International Journal of Computer Trends and Technology*. 71. 125-130. 10.14445/22312803/IJCTT-V71I4P116.

UNECE. 2021. *Strategic Communications Framework for Statistical Institutions*, United Nations, Geneva, https://statswiki.unece.org/display/DIS.

WARD M.O., GRINSTEIN G., KEIM D. 2015. *Interactive Data Visualization: Foundations, Techniques, and Applications*. New York: A K Peters/CRC Press.

Donatella CANGIALOSI, Istat, donatella.cangialosi@istat.it
Giuseppe LANCIONI, Istat, giuseppe.lancioni@istat.it
Debora TRONU, Istat, debora.tronu@istat.it
Daniela VACCA, Istat, daniela.vacca@istat.it

LIFE SATISFACTION AND LEISURE TIME AMONG THE ELDERLY POPULATION

Giovanni Busetta, Maria Gabriella Campolo, Antonino Di Pino Incognito

Abstract. The well-being of the elderly population has garnered significant attention as their numbers continue to grow. Life satisfaction serves as a key indicator to assess the quality of life among older adults. Among the factors that play a crucial role in shaping their overall life satisfaction, the allocation of time and how it is utilized has gained considerable attention. This paper aims to suggest a specification and estimation model of the relationship between time use patterns of the elderly and their life satisfaction, in Italy. Using data from the 2013 Use of Time Survey conducted in Italy by Istat, we estimate the relationship between life satisfaction and socio-demographic characteristics frequency of social interactions, identifying empirically the influence of time spent on unpaid work, media leisure, and non-media leisure activities of, respectively, young old and older old individuals. We found how engagement in unpaid work and media and non-media leisure activities, friendship, a good economic condition and belonging to a large family positively influenced life satisfaction, especially of the older old, while stress is negatively correlated with it.

1. Introduction

The evaluation of well-being and life satisfaction among the elderly population has emerged as a crucial area of research as the number of older adults continues to increase (United Nations, 2017). The increasing elderly population deserves to experience a life of dignity during the later stages of their lives (Smith, 2001). Life satisfaction represents an individual's subjective evaluation of their overall life conditions, encompassing various domains such as physical health, social relationships, and personal fulfilment. Life satisfaction, as defined by Diener (1984), refers to an individual's subjective evaluation of their own life, encompassing cognitive judgments and affective reactions. It is a global assessment that reflects the extent to which individuals perceive their lives as fulfilling and satisfying. Life satisfaction is influenced by various factors, including personal goals, values, expectations, and social comparisons. It is considered a key component of subjective

well-being and an indicator of overall happiness and contentment in life (Diener, 1984; Diener, Emmons, Larsen, and Griffin, 1985; Meggiolaro and Ongaro, 2015).

Many studies have found a U-shaped relationship between age and life satisfaction, indicating that younger and older individuals tend to be more satisfied with life compared to those in middle age (Frey and Stutzer, 2010). One of the main interpretations of this relationship suggests that it reflects the influence of key life events (such as marriage, childbirth, retirement) that are closely tied to specific ages. In this respect, the objective of this paper is to analyze whether levels of satisfaction increase or decrease among the young old and the older old categories as a consequence of the different impact of the use of the time of the elderly people belonging to these two categories. In particular, we want to determine the extent to which the change in life satisfaction depends not only on socio-demographic factors, but also on leisure activities. Previous studies have found that, gender alone may not have a substantial impact in explaining the differences in life satisfaction (e.g. Wallace, 2008). Other researchers have found distinct outcomes and trends for men and women depending on social and cultural environment (Meggiolaro and Ongaro, 2015) as, for example, in the case of the elderly person who lives alone.

The results presented in this study provide preliminary insights into the associations between various factors and life satisfaction among the elderly population. The findings underscore the importance of social connections, engagement in leisure activities, and unpaid work in shaping subjective well-being. These insights have implications for developing targeted interventions and policies to enhance the quality of life for older adults.

The remainder of the paper is structured as follows. Section 2 report a brief survey on previous studies regarding the influence of leisure time on older well-being; Section 3 describes the dataset and the methodological instruments employed to reach our aims; Section 4 presents the results; Section 5 provides some discussion and concludes.

2. Background

Several studies have found that time employed in leisure activities may be consistently identified as influential factors in promoting life satisfaction among older adults (Brajša-Žganec *et al.*, 2011). Engaging in leisure activities not only provides individuals with a sense of enjoyment and fulfilment but also contributes to the formation of social connections and the enhancement of overall quality of life (Smith and Kawachi, 2014; Dykstra and Fokkema, 2011). Leisure activities can range from physical exercises and hobbies to cultural and social pursuits. Several studies have specifically emphasized the benefits of leisure activities for the elderly

population (Chang *et al*., 2012; Cheung *et al*., 2018; Holt-Lunstad *et al*., 2010). Moreover, strong social networks and meaningful social interactions have been consistently associated with higher levels of life satisfaction (Smith and Kawachi, 2014; Dykstra and Fokkema, 2011).

Certain studies indicate that the influence of social interactions on life satisfaction differs between men and women. For instance, research by Cheng and Chan (2006) suggests that communication with friends and acquaintances has a more significant impact on women's life satisfaction compared to men. Additionally, Oshio (2011) suggests that women tend to experience a stronger association between family ties and life satisfaction, while men may exhibit lower levels of life satisfaction following divorce or widowhood. Family status is also identified as a crucial determinant of life satisfaction among older individuals in Europe, with research indicating that married or partnered individuals tend to report higher well-being compared to those who are single (Buber and Engelhardt, 2008).

It is important to note that the literature on life satisfaction among the elderly population also acknowledges certain limitations in the interpretative capacity of the phenomenon (Baltes and Smith, 2003). One of the main limits of the studies on this topic is that self-reported measures used in these studies may introduce biases, as individuals' perceptions and subjective interpretations of well-being can vary (see, among others, Graham, 2012). Following the Graham reasoning (Graham, 2012), for example, this bias is due to the circumstance that synthetic measures of perceived life satisfaction often fail to capture individual differences in the experience and evaluation of objectively similar situations. To reduce the effect of this concern, the current study seeks to investigate the influence, on life satisfaction, of "at-least-in-part" exogenous predictors given by the individual measures of the use of the time, such as leisure time and unpaid or domestic activities using the time diary method. To accomplish this, data from the 2013 Use of Time Survey conducted by ISTAT in Italy will be utilized.

3. Data and Methodology

The data used in this work have been taken from the Time Use Survey, provided by the Italian National Institute of Statistics (ISTAT), a cross national database that collects information from a representative sample of households across Italy. The aim of the survey is to capture a comprehensive picture of individuals' daily routines and the distribution of time across various domains, examining aspects of daily life that are otherwise unobservable (Robinson and Godbey, 1997) with an extremely high level of accuracy (Raley, 2014).

Participants are asked to keep a self-administered time diary, recording their activities and the duration of each activity over a designated period. The 24 hours of each day were divided into 144 intervals of ten minutes each, and for each time interval, the respondents note their primary and secondary activities in their diaries, indicating where and with whom these activities were performed.

The data collected through the Time Use Survey allows researchers, policymakers, and social scientists to analyse trends, patterns, and disparities in time use within the population. It provides a basis for understanding how individuals allocate their time, the impact of gender roles and socio-economic factors on time use, and the relationship between time use and well-being.

In Italy, this survey, conducted through a sampling method, is carried out every five years and covers the entire calendar year. The last edition was conducted from November 1, 2013, to October 31, 2014, on a sample of approximately 24,000 households (44,866 individuals), using the PAPI technique (PAper and Pencil Interview) and involves the use of a questionnaire and two paper diaries.

For the empirical analysis, we selected a sample of 9,581 older individuals, and we categorized them into two groups: the young old (aged 65-74 years; 50.39%) and the older old (aged >74 years; 49.61%), following previous research (Vilhelmson *et al.*, 2022).

Considering the demographics of the sample, 43.21% were women, 57% had a low level of education, a percentage that decreased to 44% when considering only the group of young elderly, 57% were married, 37% lived in the south or islands. Considering the overall sample, the average minutes spent in unpaid work was 221.01 minutes (SD 164.21; range= 0-930), in media leisure 199.18 (SD 134.07; range= 0-850 minutes) and 208.23 in non-media leisure (SD 134.95; range= 0-930 minutes). When considering the two subsamples, it was found that the younger elderly group spends more time, on average, on unpaid work and non-media leisure activities compared to the older group. Specifically, regarding unpaid work, the young old group reported an average 248.46, while the older old group reported 193.13 minutes ($t = -16.73$, p-value = 0.0000), indicating a statistically significant difference. As for media leisure, the young old group reported an average of186.10 minutes, while the older old group reported 212.46 minutes ($t = 9.67$, p-value = 0.0000), also showing a statistically significant difference. However, in the case of non-media leisure, the mean difference between the young old group (209.43 minutes) and the older old group was not statistically significant.

In the following table (Table 1) we show the descriptive statistics of the sample by age.

Table 1 – *Characteristics of the sample by age*

	Overall obs. 9581		Young Old obs. 4828		Older Old obs. 4753	
	Mean or %	Std. Dev	Mean or %	Std. Dev	Mean or %	Std. Dev
Life satisfaction	6.72	1.90	6.96	1.74	6.47	2.02
Woman (yes)	0.43		0.47		0.39	
Years of schooling	7.35	4.23	8.42	4.34	6.27	3.82
South and Islands	0.37		0.36		0.37	
Marital status (ref Single)	0.06		0.06		0.06	
Married/coahabiting	0.57		0.69		0.45	
Sepeareted/divorced	0.05		0.8		0.03	
Widowed	0.32		0.17		0.46	
N° household members	2.02	1.00	2.20	1.01	1.84	0.95
Economic satisfaction:						
Not at all	0.14		0.14		0.14	
A little	0.33		0.31		0.35	
Quite a bit	0.50		0.52		0.48	
Very much	0.03		0.03		0.03	
Frequency seeing friends:						
No friends	0.03		0.02		0.05	
Never	0.11		0.06		0.16	
A few times a year	0.12		0.10		0.15	
A few times a month (<4)	0.19		0.20		0.18	
Once a week	0.16		0.18		0.13	
More than once a week	0.23		0.26		0.19	
Daily	0.16		0.17		0.15	
Feeling stressed:						
Never	0.40		0.37		0.44	
Sometimes	0.43		0.47		0.38	
Often	0.11		0.10		0.11	
Always	0.07		0.06		0.07	
Rely on other people (Yes)	0.85		0.83		0.86	
Trust in others: Most people are trustworthy	0.18		0.21		0.16	
Satisfaction with the time dedicated to oneself:						
Yes	0.82		0.80		0.85	
No, too much	0.04		0.03		0.05	
No, too little	0.14		0.17		0.10	
Log Media Leisure	199.18	134.07	186.10	125.14	212.46	141.36
Log Non-Media Leisure	208.23	134.95	209.43	138.52	207.02	131.23
Log Unpaid Work	221.01	164.21	248.46	166.23	193.13	157.32

Notes: Standard Deviations are reported only for continuous values

The dependent variable, "Life satisfaction" was assessed using the following question: "How satisfied are you with your life overall?". Participants rated their satisfaction on a scale of 0 to 10, where 0 represented "not satisfied at all" and 10 represented "very satisfied". The mean score of the overall sample was 6.72 ± 1.90. Considering the two subsamples the level of life satisfaction was 6.96 ± 1.74 and 6.47 ± 2.02 for young old and older old, respectively (t = -12.7433, p-value = 0.0000).

The socio-demographic variables used in our analysis include gender (Woman: 1 = yes; 0 = otherwise), the education level of the subject expressed in years of schooling, a dummy variable that indicates whether he/she lives in Southern Italy/or islands (South and Islands: 1 = yes, 0 = otherwise).

Family structure is analysed by the following variables: the numbers of household members and the marital status. For this last variable, we consider the following four categories: single as reference, cohabiting or married, separated/divorced, widowed.

The Italian time use survey did not include information about the wages of adult members or family income. Therefore, we include as proxy of these variables the level of the economic satisfaction a categorical variable ranging from 1 to 4, where 1 (reference) represents “not at all” and 4 represents “very much”.

To evaluate the relations with neighbours and the territorial context, we considered three variables indicating: i) the frequency of contact with friends; ii) if he/she can rely on people (excluding cohabiting family members) in case of need; iii) the level of trust in most people assessed using the following question “Do you generally think that you can trust most people? (Dummy: 1 =Most people are trustworthy; 0= otherwise).

Since there were no direct information about health status or self-perception of health, we considered the self-reported stress level as a proxy (feeling stressed: 1=No, never as reference; 2=Yes, sometimes; 3= Yes, often; 4= always).

Finally, we report the time spent by the subjects on the following activities:

- Unpaid work (including housework and family care);
- Media Leisure (TV, radio, reading books and magazines);
- Non-Media Leisure (informal leisure activities that involve exchanging feelings and opinions, such as socializing, conversations, sports, hobbies, as well as formal activities that contribute to the cultural development of individual, such as voluntary work, aid to other families, social participation, religious activities, adult education, and cultural events (Robinson and Godbey, 2010).

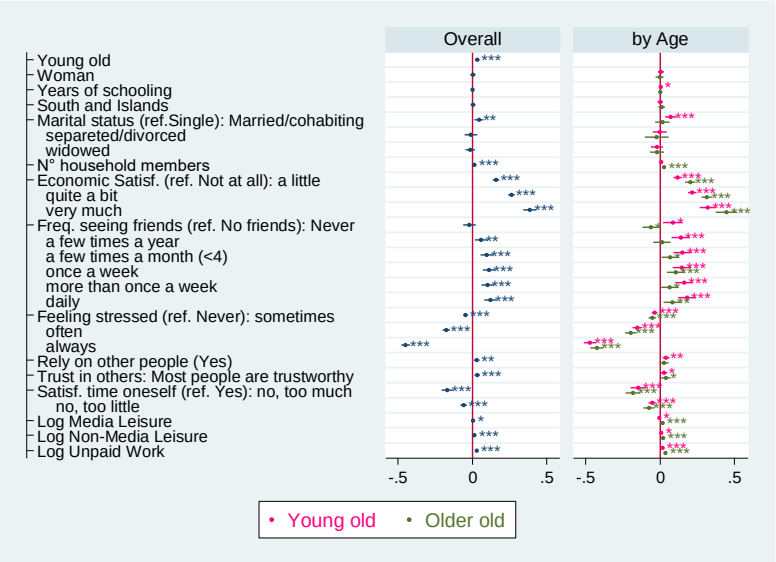
We employed three Ordinary Least Squares (OLS) regression models: one for the overall sample and two for each age subgroup. Acknowledging potential criticism regarding the treatment of life satisfaction scores as cardinal, we have opted to estimate a cross-section model using OLS regressions, consistent with other studies (e.g., Cerci and Dumludag, 2019; FitzRoy et al., 2014). This decision stems from the

observation that studies of this nature often yield similar results, whether life satisfaction is treated as a cardinal variable in OLS or as an ordinal variable in an ordered response model (Ferrer-i-Carbonell and Frijters, 2004). Our preference for OLS is motivated by its simplicity and the ease of result interpretation, given that estimates from cardinal models are typically more intuitive. All analyses were conducted using Stata (SE15).

4. Results

When considering the full sample (Overall), life satisfaction increases for young old, marital status (married/cohabiting/separate/divorced/widowed on respect to single), number of household members, economic satisfaction. In Fig. 1, we present the preliminary results of our analysis.

Figure 1 – *Life Satisfaction: OLS Estimation Results.*



*Notes: * p<0.05; ** p<0.01; *** p<0.001*

Analysing the subsample (by Age), marital status, maintains statistical significance, but more relevant for the category of young old compared to older old. Also, number of household members and economic satisfaction maintain the same

sign and statistical significance, but in this case the effect is higher for older old compared to young old. Moreover, as the frequency of socializing with friends increases, life satisfaction also increases, and this effect is most evident and significant for the young old. Feeling stressed reduces life satisfaction for all categories with a higher impact for young old.

Rely on other people, trust in others, log media and non-media leisure and unpaid work are all positively correlated to life satisfaction with a slightly higher effects for older old.

From the results just presented it emerged that having been in couple at least in a part of the individual's life (married/cohabiting/separate/divorced/widowed) increases life satisfaction especially of young old.

The perception of the own economic condition progressively produces a direct effect on the own overall satisfaction. Analogous significant effects are produced by the perceived stress. Years of schooling appears to produce positive effects on life satisfaction only for young old.

Older people who do not live alone show higher levels of satisfaction compared to those who live alone. These findings are supported by the positive and significant coefficient related to the number of household members. On the other hand, living in the southern regions or islands has not a statistically significant effect on life satisfaction neither for the Overall sample, nor for young old and older old.

The time spent to unpaid work and media and non-media leisure activities increases life satisfaction for both subgroups (overall, young old and older old). Trust in others is also a significant factor in the perception of greater life satisfaction.

Note as the influence of control variables on satisfaction are a bit more significant in the subgroup of the older old (Fig. 1).

We found that the estimation results of a life satisfaction function, obtained by introducing control variables as family size, economic satisfaction, media and non-media leisure and unpaid work, reinforce the idea of a U-shaped trend of life satisfaction with respect to ageing, with an increased effect more for older old than for young old.

This aligns with existing literature highlighting the notion of the "paradox of aging," which suggests that individuals may experience improved well-being as they enter later stages of life (Gerstorf *et al*., 2008; Paggi *et al*., 2016).

Furthermore, our study reveals the importance of social connections and relationship status for life satisfaction among both young and older adults. Those who are in committed partnerships or living with other family members tend to report higher levels of life satisfaction, emphasizing the role of social support, companionship, and emotional well-being (Smith and Kawachi, 2014; Dykstra and Fokkema, 2011). These findings are in line with the social integration theory, which posits that social relationships and social support networks play a vital role in

promoting well-being and satisfaction in later life (Berkman *et al*., 2000). The estimations results are also reported in the appendix (Tab. A1).

5. Conclusion

Our study contributes valuable insights into the determinants of life satisfaction among the elderly population in Italy. The findings underscore engagement in leisure activities in shaping subjective well-being in later life. Engagement in unpaid work and non-media leisure activities emerged as positive contributors to life satisfaction among older adults. These activities provide a sense of purpose, fulfillment, and social connectedness, thereby enhancing life satisfaction and, consequently, overall well-being and life satisfaction.

Furthermore, our study confirms the positive impact of social connections and relationship status for life satisfaction, emphasizing the role of emotional support, companionship, and shared experiences in contributing to overall life satisfaction.

These findings have important implications for interventions and policies aimed at promoting life satisfaction and well-being among the elderly population. Facilitating social interactions, supporting stable relationships, and encouraging meaningful leisure activities can foster positive aging experiences and enhance the overall quality of life for both young and older adults.

It is important to acknowledge that this study has certain limitations. The cross-sectional design of the study limits our ability to establish causal relationships, and the use of self-reported measures may introduce biases. Further research is needed to expand upon these findings, employing robust study designs and considering diverse populations, to deepen our understanding of the factors that contribute to well-being among older adults.

Future research employing longitudinal designs and objective assessments could provide further insights into the dynamics of life satisfaction among older adults.

Appendix

In Table A1, we present the estimation results of our OLS models. The first column displays the coefficients for the full model (Overall), while the second column showcases the results for the Young old subsample. Finally, the last column presents the results for the Older old subsample.

Table A1 – *Life Satisfaction: OLS Estimation Results*

	Overall		Young old		Older old	
	Coef.	p.	Coef.	p.	Coef.	p.
Constant	1.3184	***	1.4721	***	1.2128	***
Young old (Yes)	0.0320	***				
Woman (yes)	0.0029		0.0042		-0.0035	
Years of schooling	0.0011		0.0024	*	-0.0002	
South and Islands	0.0045		-0.0006		0.0104	
Marital status (ref. Single)						
Married/coahabiting	0.0443	**	0.069	***	0.0152	
Sepeareted/divorced	-0.0099		-0.0029		-0.0240	
Widowed	-0.0156		-0.0209		-0.0209	
N* household members	0.0151	***	0.0074		0.0249	***
Satisfaction Economic Situation (ref. Not at all)						
A little	0.1586	***	0.1157	***	0.2025	***
Quite a bit	0.2632	***	0.2137	***	0.3118	***
Very much	0.3854	***	0.3198	***	0.4456	***
Frequency seeing friends (ref. No friends						
Never	-0.0197		0.0858	*	-0.0637	*
A few times a year	0.0595	**	0.1378	***	0.0114	
A few times a month (<4)	0.0961	***	0.1487	***	0.0650	*
Once a week	0.1110	***	0.1451	***	0.1042	***
More than once a week	0.1014	***	0.1600	***	0.0622	*
Daily	0.1211	***	0.1792	***	0.0822	**
Feeling stressed (ref. Never)						
Sometimes	-0.0453	***	-0.0383	***	-0.0533	***
Often	-0.1748	***	-0.1534	***	-0.1982	***
Always	-0.4488	***	-0.4720	***	-0.4249	***
Rely on other people (YES)	0.0297	**	0.0382	**	0.0248	
Trust in others: Most people are trustworthy	0.0322	***	0.0258	*	0.0382	*
Satisfaction with the time dedicated to oneself (ref. Yes)						
No, to much	-0.1687	***	-0.1474	***	-0.1822	***
No, too little	-0.0588	***	-0.0546	***	-0.0766	***
Log Media Leisure	0.0054	*	-0.0071	*	0.0171	***
Log Non-Media Leisure	0.0134	***	0.0071	*	0.0185	***
Log Unpaid Work	0.0283	***	0.0147	***	0.0361	***
N° of observations	9581		4828		4753	
R^2	0.2761		0.2651		0.2773	

*Notes: * p<0.05; ** p<0.01; *** p<0.001*

References

BALTES M.M., SMITH J. 2003. New frontiers in the future of ageing: From successful ageing of the young old to the dilemma of the fourth age, *Gerontology*, Vol. 49, pp. 123-135.

BERKMAN L.F., GLASS T., BRISSETTE I., SEEMAN T.E. 2000. From social integration to health: Durkheim in the new millennium, *Social Science & Medicine*, Vol. 51, No. 6, pp. 843-857.

BRAJŠA-ŽGANEC A., MERKAŠ M., ŠVERKO I. 2011. Quality of life and leisure activities: How do leisure activities contribute to subjective well-being?. *Social Indicators Research*, Vol. 102, pp. 81-91.

BUBER I., ENGELHARDT H. 2008. Children's impact on the mental health of their older mothers and fathers: findings from the Survey of Health, Ageing and Retirement in Europe. *European Journal of Ageing*, Vol. 5, pp. 31-45.

CERCI, P.A., DUMLUDAG, D. 2019. Life Satisfaction and Job Satisfaction among University Faculty: The Impact of Working Conditions, Academic Performance and Relative Income, *Social Indicators Research*, 144, pp. 785–806.

CHANG Y.K., LABBAN J.D., GAPIN J.I., ETNIER J.L. 2012. The effects of acute exercise on cognitive performance: a meta-analysis, *Brain Research*, Vol. 9, No. 1453, pp. 87-101.

CHENG S.T., CHAN A.C. 2006. Relationship with others and life satisfaction in later life: do gender and widowhood make a difference?, *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol. 61, No. 1, pp. P46-P53.

CHEUNG D.S.K., LAI C.K.Y., WONG F.K.Y., LEUNG M.C.P. (2018). The effects of the music-with-movement intervention on the cognitive functions of people with moderate dementia: a randomized controlled trial, *Aging Mental Health*, Vol. 22, No. 3, pp. 306-315.

DIENER E. (1984). Subjective well-being, *Psychological Bulletin*, Vol. 95, No. 3, pp. 542-575.

DIENER E., EMMONS R.A., LARSEN R.J., GRIFFIN S. (1985). The Satisfaction with Life Scale, *Journal of Personality Assessment*, Vol. 49, No. 1, pp. 71-75.

DYKSTRA P.A., FOKKEMA T. 2011. Relationships between parents and their adult children: A West European typology of late-life families, *Ageing & Society* Vol. 31, No. 4, pp. 545-569.

FERRER-I-CARBONELL, A., FRIJTERS, P. 2004. How important is methodology for the estimates of the determinants of happiness?. *The economic journal*, Vol. 114, No. 497, pp. 641-659.

FITZROY, F.R., NOLAN, M.A., STEINHARDT, M.F., ULPH, D. 2014. Testing the tunnel effect: comparison, age and happiness in UK and German panels. *IZA Journal of European Labor Studies*, Vol. 3, pp. 1-30.

FREY B.S., STUTZER A. 2010. *Happiness and economics: How the economy and institutions affect human well-being*. Princeton: Princeton University Press.

GERSTORF D., RAM N., ESTABROOK R., SCHUPP J., WAGNER G.G., LINDENBERGER U. 2008. Life satisfaction shows terminal decline in old age:

Longitudinal evidence from the German Socio-Economic Panel Study (SOEP), *Developmental Psychology*, Vol. 46, No. 4, pp. 1140-1149.
GRAHAM, C. (2012). *Happiness around the world: The paradox of happy peasants and miserable millionaires*. Oxford: Oxford University Press.
HOLT-LUNSTAD J., SMITH T.B., LAYTON J.B. 2010. Social relationships and mortality risk: A meta-analytic review. *PLOS Medicine*, Vol. 7, No. 7, e1000316.
MEGGIOLARO S., ONGARO F. 2015. Life satisfaction among older people in Italy in a gender approach, *Ageing & Society*, Vol. 35, No. 7, pp. 1481-1504.
PAGGI M.E., JOPP D.S., HERTZOG C. 2016. The importance of leisure activities in the relationship between physical health and well-being in a life span sample, *Gerontology*, Vol. 65, No. 3, pp. 299-309.
RALEY S. 2014. Time use, inequality, and child well-being. In *Handbook of child well-being*, Netherlands: Springer, pp. 999–1031.
ROBINSON J.P., GODBEY G. 2010. *Time for life: The surprising ways Americans use their time*. University Park: Pennsylvania State University Press.
SMITH J. (2001) Well-being and health from age 70 to 100: Findings from the Berlin aging study, *European Review*, Vol. 9, No. 4, pp. 461-477.
SMITH N.D.L., KAWACHI I. 2014. State-level social capital and suicide mortality in the 50 US state, *Social Science & Medicine*, Vol. 120, pp. 269-277.
OSHIO T. 2011. Gender Differences Among Elderly Japanese: Importance of family and social relations for life satisfaction. *RIETI Discussion Paper Series 11-E-051*. https://doi.org/10.1007/s10823-012-9169-y.
VILHELMSON B., THULIN E. ELLDÉR E. 2022. Is ageing becoming more active? Exploring cohort-wise changes in everyday time use among the older population in Sweden, *European Journal of Ageing*, Vol. 19, pp. 447–461.
United Nations 2017. *World population prospects: The 2017 revision*. New York: Department of Economic and Social Affairs, Population Division. http://esa.un.org/unpd/wpp/.
WALLACE T. 2008. *Life satisfaction in individuals age sixty-five years of age and older*. PhD thesis, Oklahoma State University. UMI Dissertations Publishing.

Giovanni BUSETTA, University of Messina, gbusetta@unime.it
Maria Gabriella CAMPOLO, University of Messina, mgcampolo@unime.it
Antonino DI PINO INCOGNITO, University of Messina, dipino@unime.it

EVOLUTION OF HOMICIDAL VIOLENCE IN ITALY: THE END OF MEZZOGIORNO'S EXCEPTIONALISM?

Claudio Caterino, Luigi M. Solivetti, Franco Turetta

Abstract. The homicidal violence of Mafia-type gangs has been the primary cause of the Mafia legend and has prompted the publication of numerous studies, which, however, have provided conflicting hypotheses about the determinants of this criminal phenomenon. The persistence, in Italy's Mezzogiorno, the traditional Mafia's turf, of a high rate of homicides has, in particular, attracted worldwide attention. The present study aims to analyse the evolution of Mezzogiorno's homicides and elucidate its causes. The analysis is based on panel data covering 4-decade cross-province time series. The time series showed a drop in Mezzogiorno's homicide rates. The longitudinal models, in turn, showed that this drop has been associated with structural changes that occurred in Mezzogiorno, particularly with the momentous rise in the youth's education and a surge in State enforcement. At the end of the time frame considered, the gap in terms of homicide rates between Mezzogiorno and the rest of Italy evaporated. Mezzogiorno's rate became lower than the average rate for West European countries, ending the long-lasting Mezzogiorno's exceptionalism in terms of homicidal violence.

1. Introductory remarks: Mafia and violence

Behind the interest aroused worldwide by Mafia-type criminal organisations is their violence. The amount and cruelty of such violence have captured people's imagination and prompted countless stories, novels, and films. All this, in turn, has generated the Mafia legend that has inflated an already gruesome reality. In contrast to the vastness of this folkloristic and fictional discourse, there is limited information about the Mafia's nature and determinants. What we know about Mafia-type organisations is that they are criminal gangs, territorially-rooted, close-knit and bound by the silence code. Making the most of the force of their associative bond, they use threat and violence to conduct otherwise legal business or to trade in illegal markets (e.g. that of drugs), recurring to murder to eliminate their competitors and, when expedient, to counter State enforcement. Numerous scientific publications have dealt with the issue of the roots of these organisations. However, the theories

meant to explain persistent criminal behaviour in groups of people do not seem well-suited to provide a convincing account of Mafia-type gangs.

Social disorganisation theory has long been regarded as a valid tool to elucidate the cross-time stability of criminal behaviour in specific territorial units. The reference to *stability* and *specific territories* seems to fit Mafia-type gangs. However, disorganisation theory states that criminal groups emerge in contexts where socioeconomic challenges – and, first of all, high mobility – break down family and community ties, destroy social control and finally, indirectly encourage the creation of criminal gangs (Shaw and McKay 1942; Earls and Visher 1997; Sampson et al. 1997). These characteristics seem lacking in Italy's Mezzogiorno,[1] the turf of Mafia-type gangs. In Mezzogiorno, there has been high mobility, especially in the decades immediately after WWII; but it has been mobility in terms of emigration, not immigration, as suggested by the disorganisation theory. And it would be hard to regard Mezzogiorno as a territory characterised by a breaking down of family ties and parental control. Those territories have always shown much lower separation and divorce rates than the rest of Italy, higher birth rates, a limited number of single-parent families, etc. Only recently emerged a new social disorganisation factor: foreign immigration. However, foreign immigration has been much lower in Mezzogiorno than in the rest of Italy.

Relative deprivation theory seems better suited to explain Mafia-type crime. This theory is indebted to the anomie conceptual framework (Merton 1949), according to which social pressure to succeed materially in the face of scarce legitimate opportunities leads to crime. There was undoubtedly a deprivation scenario in Mezzogiorno in the past. Per capita GDP was 56% of the rest of Italy in the early 1960s. In the early 2020s, Mezzogiorno's share was substantially the same, but from 1980, there had been a substantial increase (45%) in per capita GDP at constant prices all over the country, Mezzogiorno included. In the early 1970s, 32.8% of Mezzogiorno's people were in *absolute poverty* vs 14.4% in the rest of Italy; in 2008, the gap had shrunk to 9.2% vs 2.2% (Vecchi 2011). In 1961, people in the agricultural sector were, on average, 45.4% in Mezzogiorno and 24.1% in the rest of Italy. In 2020, the values had shrunk to 1.9 and 1.2, respectively.

In 1961, in Sicily, there were only 3.2 cars per 100 population; in Calabria, 1.8; in Basilicata, 1.2; while in Lombardy, there were 6.5 cars and, in Latium, 7.6. It was necessary to wait until the 2010s to see the disappearance of the gap, with Latium and Sicily presenting the same diffusion of cars (~65 per 100 population).

As for education, too, the gap was substantial. In the early 1980s, many Mezzogiorno provinces presented a secondary higher-school male enrolment rate

[1] With "Mezzogiorno", we refer to the southern part of Italy, including Sicily, but excluding Sardinia where Mafia-type gangs are not present, corresponding to ITF and ITG in NUTS 1 classification.

between 35% and 45%. In contrast, the rate was at least 10 points higher in most Northern and Central provinces. Only at the end of the 2010s the gap evaporated, with gross enrolment rates of 90% and more in substantially all of Italy's provinces.

Summarising the above, Mezzogiorno's past educational and economic scenario was such to suggest that, owing to the shortage of legitimate opportunities, criminal gang activities would have provided alternative opportunities for economic success. However, the scarcity of legitimate opportunities is not the only requirement of relative deprivation theory. This theory incorporates a precondition: a high social pressure to succeed materially. This precondition does not seem to characterise Mezzogiorno: a macro-region that is more tradition-directed and more inclined to favour ascription against achievement than the Northern and Central regions. Little economic-success-driven mentality can be found, in particular, in the biographies of Mafia bosses, most of them known for their modest, peasant-like lifestyle.

Given all this, other hypotheses have been advanced to explain the diffusion of Mafia-type criminal organisations. A more cultural hypothesis has derived from a peculiar feature repeatedly identified in the Mezzogiorno society: the so-called *amoral familism*, a blend of familism, lack of trust and, consequently, limited social capital (Banfield 1958; Bell 1979; Putnam et al. 1993). Over the last decades, this peculiar feature – initially detected in Italy's Southern regions – has been widely employed to explain the weaker social fabric of some regions and countries worldwide (Coleman 1988; Fukuyama 1995; Putnam 2000). Proxies of social capital, such as turnout at elections and referendums and diffusion of volunteers, persistently present lower values in the Mezzogiorno regions (Solivetti 2020). Unsurprisingly, the dearth of social capital and the lack of trust in the State and people outside the family have been associated with Mafia-type criminal gangs (Hess 1973; Gambetta 1993; Paoli 2004).

On the other hand, the lack of trust in the State is partly explained by the State's weak presence in Mezzogiorno. A weak State, unable to control private violence and conflicts revolving around land possession and natural resources exploitation, would have encouraged the rise of the Mafia as an illegal but effective regulator (Buonanno et al. 2015; Acemoglu et al. 2020). Persistence of the Mafia diffusion would have been assured by its effectiveness as a provider of ''private protection'' (Gambetta 1988, 1993) in a context characterised by ineffective law enforcement.

In any case, Mezzogiorno's criminal exceptionalism seems to have markedly shrunk over the last years. The structural changes that occurred in Mezzogiorno's socioeconomic profile in the previous decades might have been the effective causes of the assumed decline of its traditional violent criminality. This decline might have also resulted from a cultural change generated, at least partly, by the increasing education of the Mezzogiorno's youth. Interestingly, since the 1990s, in the very regions most touched by Mafia-type criminality, people, for the first time, organised

public manifestations against criminal gangs and gave rise to anti-Mafia associations (Cayli 2013). The presence of students and, in general, young people was conspicuous in those manifestations.

A further relevant factor might have been the more robust enforcement action by the State. In 1982, the Italian Parliament passed a new act introducing the offence of *Mafia-type criminal conspiracy*, which entails much longer incarceration terms than the standard *criminal conspiracy*. The same act also introduced the precautionary seizure of all possessions (property, money) of suspicious origin. In a few years, the Italian judiciary performed an effective action against the various Mafia-type criminal organisations ('Ndrangheta, Mafia, Camorra, Sacra Corona, etc.) also by taking advantage of the (contentious) collaboration of Mafia's former affiliates (so-called *pentiti*), to whom reduced sentences were granted. Despite this, in the 1990s, the Mafia killed some high-ranking magistrates overseeing the enforcement actions against these criminal organisations. New acts were passed to make the fight against them more effective. The Parliament extended the hard prison regime to Mafia-type criminals (1992) and passed a new act providing more protection for criminal organisations' affiliates who turned into police co-operators.

The change in Mezzogiorno's crime might have also resulted from a shift in focus of the Mafia's interests. Traditionally, Mafia-type gangs used threat and violence in both the rural and urban contexts of a tradition-directed society, primarily controlling the land and the markets dealing with agricultural produce. However, social scientists noticed changes in the interests of these criminal organisations. This shift concerned the capability of adapting to new political situations and new market conditions (Catanzaro 1985; Shelley 1994) and the move from extortions and violence towards more entrepreneurial – though still illegal or semi-illegal – businesses (Arlacchi and Ryle 1986; Paoli 2008). A further change has regarded these criminal organisations' movement from their habitual Southern turf to the Central and Northern provinces, at least to invest the returns of their illicit activities.

Whatever the reasons for the decline of Mafia violence, investigating the determinants of this decline would lead to a better understanding of the causes of the entire Mafia phenomenon and would indirectly test the theories about its rise.

2. Purpose and hypotheses of the present study

We will begin by measuring the evolution of Mezzogiorno's homicide rates and comparing them to the rest of Italy's and other European countries'. We will check whether the alleged decline of homicides in Mezzogiorno is a mirror of the decrease in homicides registered from the 1990s in Western Europe or a specific phenomenon.

We intend to focus on the Mezzogiorno's exceptionally high rates of homicides because intentional homicides represent not only the ultimate crime but also the most striking fallout of traditional Mafia-type criminal activities.

Next, we intend to measure the impact on homicide rates of changes that intervened over time. In particular, in tune with the abovementioned potential factors of Mafia-type criminal organisations, we advance the following hypotheses:

- H1: because young people with educational aspirations are less likely to engage in crime (Braithwaite 1989: 44 ff.), an increase in their educational level could have had an impact on violent crime by introducing new values and attitudes;
- H2: if the origin and persistence of Mafia-type crime have been favoured by weak State enforcement, then an increase in State enforcement could have been the reason for the decrease in violent crime;
- H3: a general increase in the Mezzogiorno's well-being could have reduced the pressure exerted on the youth to recur to illegitimate routes to success;
- H4: a decrease in birthrate would have allowed families to better provide for their young generations, reducing their drive towards illegitimate means.

3. Data and methods

Most literature on the Mafia is qualitative. The few quantitative studies have usually focused on synchronic/pooled data of territorial units. Such studies are good at identifying the *between* differences, but they miss the *within* differences, i.e. variations in each observational unit. Thus, these studies explain differences across observations but not change over time. Secondly, crime densities could also derive from non-observed, time-invariant factors, which would bias the estimates. Fixed-effects (FE) panel analyses, being based on cross-time variations – i.e. the *within* differences – would bypass the effects of non-observed, persistent crime factors and identify the effects of change. Moreover, FE analyses may reckon the temporal order of events: therefore, they help distinguish between correlation and causality, a problem lurking beneath any synchronic analysis. With panel data, however, FE models would pinpoint the effects of over-time deviations from the mean, but they would say nothing about the pure cross-sectional relationships, those between response and explanatory variables, that transcend time. We expect these cross-sectional relationships to be particularly relevant to Mafia-type crime. To bypass this catch, we recurred to a model combining *within* and *between* effects via a random-effects (RE) estimator based on the average of the estimates produced by between and within estimators. Our regression models use a *within* effects formula (1):

$$(y_{it} - \bar{y}_i) = \beta(x_{it} - \bar{x}_i) + \tau_{t2,\dots,tnth} + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (1)$$

where *y* is the crime rate in the province *i*, during the year *t*, *x* is the set of explanatory variables, τ represents time dummies, by which we controlled for the year-specific variations of *y*, and ε is the error term. This formula is combined with that of a regression based on the mean values for each territorial unit (i.e. province) (2):

$$\bar{y}_i = \propto + \beta\bar{x}_i + v_i + \bar{\varepsilon}_i \quad (2)$$

Little research has been conducted on panels of Mafia-type crimes. Moreover, those panels usually covered relatively short spans of time. In the present study, we made use of the longest possible panel dataset: ~40-year cross-province time series.

The present paper's response variables concern intentional, completed homicide (henceforth IH). We are interested in the total IH rates and the share of IH attributed to Mafia-type gangs (27.4% of Mezzogiorno's IH). Crime data were recorded by the police and operated by Istat. As for the determinants, we considered some socioeconomic measures. We used a lagged birthrate (t–14 years) to measure the juvenile population's impact on crime. As for education, we used higher secondary school male enrolment (gross) rates. Then, we considered infant mortality, which has been increasingly regarded as a barometer for a region's overall welfare. Next, to measure the evolution of State enforcement, we computed a killer's probability of being brought to justice (3):

$$y_{it} = (cm_{it} + \overline{\overline{cm}}) / (m_{it} + \bar{\bar{m}}) \quad (3)$$

where *y* is the probability in province *i*, during the year *t*, *cm* is the number of people charged with IH, and *m* is the number of IHs. A further proxy for State enforcement was calculated as above, using the homicide clearance rate (cleared homicides divided by all homicides). We also considered a standard control variable, namely population density. All the previous variables came from Istat. To measure economic well-being, we used per capita GDP, from Eurostat and Unioncamere. We also used cars per population, from ACI.

We arranged these data as cross-territorial-units panels: they cover the period 1983-2021 and all the Italian provinces, ranging from 95 to 106. All the variables, except car diffusion, were not available as time series and had to be prepared ad hoc. The statistical analysis of Mezzogiorno's homicides was carried out on data concerning all its provinces, ranging from 30 to 33.

4. Results

Figure 1 shows the evolution of IH rates. It is possible to detect two facts. Firstly, the huge gap that existed at the beginning between the Mezzogiorno's rates and the

rest of Italy's: 3 to 6 vs ~1 IHs per 100K pop., when, in the other West European countries, the average rate was ~1.4. Secondly, the fall in the Mezzogiorno's rates after the spike that occurred around 1990. At the end of the four-decade period, the gap in IH rates between Mezzogiorno and the rest of Italy was substantially over. We also notice that there was no shift in IHs from Mezzogiorno to the rest of Italy because as well the latter's rates decreased: from 1.1 to 0.45 per 100K pop. Most recent IH rates for all of Italy (0.5) and Mezzogiorno (0.6) are lower than the average rate in the other Western European countries, which is 0.9 per 100K pop.

Figure 1 – *Evolution of the provinces' mean homicide rates for Mezzogiorno and the rest of Italy; upper and lower bounds (mean +/– 1 s.e.); 1983-2021.*

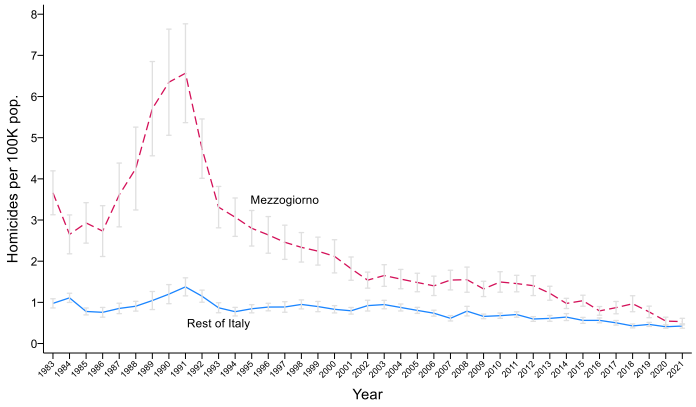


Figure 2 – *Evolution of the provinces' mean Mafia-type homicide rates for Mezzogiorno and the rest of Italy; upper and lower bounds (mean +/– 1 s.e.); 1983-2021.*

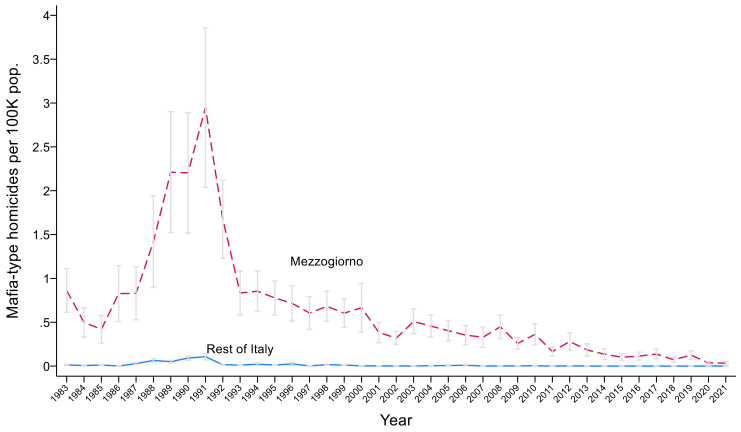


Figure 2 shows the evolution in the rates of only IHs attributed to Mafia-type organisations. In this case as well, by the end of the time frame, Mezzogiorno's rates became very close to those of the rest of Italy, where, too, Mafia-type IH rates declined. The decline in IH and Mafia-type IH rates did not affect only the average values. Some *Spaghetti graphs* (available on demand) of fit lines of evolution of homicide rates for each of Mezzogiorno's provinces or regions show that the decline occurred substantially in all of them: primarily in those regions where homicide had reached its highest incidence: i.e. Calabria (organisations of the *'Ndrangheta* type), Sicily (specific *Mafia* organisations), and Campania (*Camorra* organisations).

Figure 3 – *Pearson pairwise correlations of response and explanatory variables for all provinces of Mezzogiorno, 1983-2021 (N=1157-1233).*

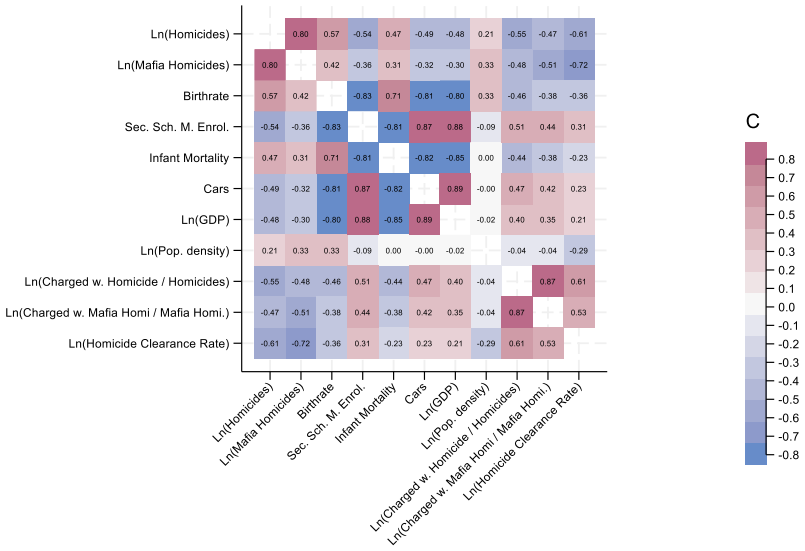


Figure 3 shows the correlations between all response and explanatory variables. All correlations between IHs rates and explanatory variables are significant at <0.000; the same occurs with Mafia IHs. IH and Mafia-type IH rates are closely correlated. IH correlations with birthrates, infant mortality and population density are positive; those with car diffusion, GDP, school enrolment, people charged with IH, and clearance rate are negative. Figure 3 also shows some close correlations between explanatory variables. School enrolment is correlated negatively with infant mortality and birthrate. GDP is correlated with car diffusion; both are correlated with school enrolment and negatively with birthrate and infant mortality. People charged with IH and clearance rates are correlated with each other. All this throws light on

the inwoven set of forces associated with IH. It also suggests the existence of other, unknown factors impacting IH rates. However, pooled-data correlations discard the over-time information in the data, and they do not control for the mentioned factors.

Multiple regression RE models are expected to provide a full test of this paper's hypotheses. Before applying the RE models, we used a Fisher-type test (Phillips–Perron option) to check the existence of a unit root. The results (available on demand) reject the null hypothesis that the panels contain unit roots. The RE regression models (Table 1) show that the variations in all IHs and Mafia-type IHs are negatively associated with school enrolment, car diffusion, the ratio people-charged-with-IH by number-of-IHs, and the IH clearance rate. Mafia-type IHs, in addition, are positively associated with population density. School enrolment and State enforcement measures emerge as the most impactful explanatory variables.

Table 1 – *Random effects regression models: Homicides and Mafia-type homicides on various explanatory variables, Mezzogiorno's provinces, 1983-2021: coefficients, prob. and (s.e.).*

Explanatory variables	Ln(homicides)		Ln(Mafia homicides)	
Birthrate (t–14)	0.011	0.202	–0.050	0.001
(s.e.)	(0.009)		(0.016)	
Higher secondary school male enrolment	–1.100	0.000	–1.120	0.009
(s.e.)	(0.246)		(0.428)	
Infant mortality	0.011	0.052	0.007	0.485
(s.e.)	(0.006)		(0.010)	
Cars per population	–0.013	0.001	–0.016	0.017
(s.e.)	(0.004)		(0.007)	
Ln(Population density)	0.082	0.075	0.352	0.000
(s.e.)	(0.046)		(0.073)	
Ln(Charged with IH / IHs) (t–1)	–0.288	0.000		
(s.e.)	(0.028)			
Ln(Charged with Mafia IH / Mafia IHs)			–0.272	0.000
(s.e.)			(0.024)	
Ln(IH clearance rate)	–0.239	0.000	–1.112	0.000
(s.e.)	(0.045)		(0.073)	
Within – between effects	yes		yes	
Fixed years	yes		yes	
Intercept	1.391	0.000	–1.301	0.022
(s.e.)	(0.340)		(0.568)	
R-squared within	0.638		0.569	
R-squared between	0.331		0.543	
R-squared overall	0.557		0.573	
Number of observations	1157		1157	

Education had a larger impact on the whole of IHs, whereas enforcement on Mafia-type IHs. The *between* effects are much more relevant in Mafia-type IHs,

confirming that this specific category of IHs is particularly affected – as expected – by persistent structural differences between the Mezzogiorno's provinces.

5. Discussion

Apart from a spike around the 1990s, IH rates in Italy's Mezzogiorno exhibited a drastic decrease since the early 1980s. The fraction of IHs attributed to Mafia-type criminal organisations also exhibited a parallel drop. At the end of the lengthy period considered in the present research, both the IH measures concerning Mezzogiorno reached levels only fractionally different from the rest of Italy's. There was no substantial transferral of IHs from Mezzogiorno to the other regions, and indeed, the IH rates in the rest of Italy exhibited a concurrent decrease. The longitudinal analysis conducted on panel data showed that the sharp rise in secondary school male enrolment was a good predictor of the decrease in homicides, also at parity of economic well-being measures. This supports the hypothesis (H1) that an increase in the educational level would have introduced values and attitudes actively opposing those characterising a low-social-capital society.

As for the economic well-being measures, car diffusion emerged as a resilient predictor of the decrease in IH rates. Although GDP was significant in single regressions, it resulted non-significant in multiple regressions when the car diffusion variable was introduced. This is probably due to the fact that cars per population is a well-being indicator that – in contrast to GDP – is expected to take in also the shadow economy profits. These results are in tune with the hypothesis (H2) of an association between traditional homicidal violence and Mezzogiorno's relative deprivation. The same results are instead in conflict with the social disorganisation theory. The vanishing of the traditional socioeconomic scenario and the new foreign immigration would have suggested an increase in violence as a consequence of an increase in social disorganisation: the opposite of what happened.

Birthrate impact on homicides did not meet the expectations (H4 hypothesis) when this variable was used in the multiple regressions together with other predictors closely (and negatively) correlated with birthrate, such as school enrolment.

Population density impacted Mafia-type IHs, suggesting that an increase in population has been relevant to Mafia-type criminal organisations.

What is certain is that an increase in State enforcement (H2) was matched by a decrease in IH rates. The ratio people-charged-with-IH by number-of-IHs and the homicide clearance rate exhibited a large impact on IH and an even larger one on Mafia-type IHs. The impact of State enforcement suggests that increased enforcement would deter criminals and make a criminal career less attractive for aspiring gangsters. This also indicates that being charged by the police probably

tarnishes the prestige not only of the charged individual but of the entire criminal underworld. Therefore, criminal gangs would be perceived as less trustworthy when claiming to be able to take care of their protegees as illegal but effective regulators.

The rise in education had a larger impact on the whole of IHs, whereas variations in enforcement a larger impact on Mafia-type IHs. This can be ascribed to the fact that the whole of Mezzogiorno's IHs contains, together with homicides officially attributed to Mafia-type gangs, some homicides lacking evidence enough to be classified as such, plus other IHs due to passion, domestic disputes, interpersonal violence, etc. These other IHs are more affected by education, values, and attitudes. In contrast, Mafia-type IHs, which are premeditated, are expected to be more affected by a rational calculation of the probability of being brought to justice.

Ultimately, educational progress and enforcement resurgence emerged as the most meaningful determinants of the decline in Mezzogiorno's IH rates. The final outcome was such a fall in IHs that not only all Italy's but also Mezzogiorno's rates became neatly lower than the average IH rate for the West European countries. This drop in homicides, therefore, was not so much the mirror of a universal decreasing trend in homicides as something associated with specific changes in the region.

All this does not mean that homicides disappeared from Mezzogiorno and the rest of Italy, nor that the Mafia-type gangs vanished. However, because in history, and not only in folklore and fiction, homicides have always been the Mafia's trademark, the above means that both surviving and new criminal organisations in Italy are other than traditional Mafia-type gangs. The primary consequence has been the end of the long-lasting Mezzogiorno's exceptionalism in terms of homicidal violence.

References

ACEMOGLU D., DE FEO G., DE LUCA G.D. 2020. Weak States: Causes and Consequences of the Sicilian Mafia. *Review of Economic Studies*, Vol. 87, No. 2, pp. 537–81.

ARLACCHI P., RYLE M.H. 1986. *Mafia Business: The Mafia Ethic and the Spirit of Capitalism*. London: Verso.

BANFIELD E.C. with FASANO BANFIELD L. 1958. *The Moral Basis of a Backward Society*. Glencoe, Ill.: The Free Press.

BELL R.M. 1979. *Fate, Honor, Family and Village: Demographic and Cultural Change in Rural Italy Since 1800*. Chicago: University of Chicago Press.

BRAITHWAITE J. 1989. *Crime, Shame and Reintegration*. Cambridge: Cambridge University Press.

BUONANNO P., DURANTE R., PRAROLO G., VANIN P. 2015. Poor Institutions, Rich Mines. *The Economic Journal*, Vol. 125, No. 586, pp. F175–202.

CATANZARO R. 1985. Enforcers, Entrepreneurs, and Survivors. *The British Journal of Sociology*, Vol. 36, No. 1, pp. 34–57.
CAYLI B. 2013. Italian Civil Society against the Mafia. *International Journal of Law, Crime and Justice*, Vol. 41, No. 1, pp. 81–99.
COLEMAN J.S. 1988. Social Capital in the Creation of Human Capital. *American Journal of Sociology*, Vol. 94, pp. S95–120.
EARLS F., VISHER C.A. 1997. *Project on Human Development in Chicago Neighborhoods: A Research Update*. Washington, D.C: US Department of Justice.
FUKUYAMA F. 1995. *Trust*. New York: The Free Press.
GAMBETTA D. 1988. Fragments of an Economic Theory of the Mafia. *European Journal of Sociology*, Vol. 29, No. 1, pp. 127–45.
GAMBETTA D. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge, Mass.: Harvard University Press.
HESS H. 1973. *Mafia and Mafiosi: The Structure of Power*. Westmead, Farnborough: Saxon House.
MERTON R.K. 1949. *Social Theory and Social Structure*. Glencoe, Ill.: The Free Press.
PAOLI L. 2004. Italian Organised Crime. *Global Crime*, Vol. 6, No. 1, pp. 19–31.
PAOLI L. 2008. The Decline of the Italian Mafia. Pp. 15–28 in *Organized Crime*, edited by D. Siegel and H. Nelen. New York: Springer.
PUTNAM R.D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.
PUTNAM R.D., LEONARDI R., NANETTI R.Y. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, N.J: Princeton University Press.
SAMPSON R.J., RAUDENBUSH S.W., EARLS F. 1997. Neighborhoods and Violent Crime. *Science*, Vol. 277, No. 5328, pp. 918–24.
SHAW C.R., MCKAY H.D. 1942. *Juvenile Delinquency and Urban Areas*. Chicago: University of Chicago Press.
SHELLEY L.I. 1994. Review: Mafia and the Italian State: The Historical Roots of the Current Crisis. *Sociological Forum*, Vol. 9, No. 4, pp. 661–72.
SOLIVETTI L.M. 2020. Political Partisanship versus Turnout in Italy's 2016 Referendum. *Quality & Quantity*, Vol. 54, No. 3, pp. 709–34.
VECCHI G. 2011. *In ricchezza e in povertà: Il benessere degli italiani dall'Unità a oggi*. Bologna: Il mulino.

Claudio CATERINO, Istat, caterino@istat.it
Luigi M. SOLIVETTI, Sapienza University of Rome, luigi.solivetti@uniroma1.it
Franco TURETTA, Istat, turetta@istat.it

A MIXTURE MODEL FOR SELF-ASSESSED STRESS AT WORK ACROSS EU[1]

Stefania Capecchi, Francesca Di Iorio, Nunzia Nappo

Abstract. Occupational stress is currently considered a primary social issue as well as an extensive problem of public health. Latest research on risk factors at work displays that about 25% of workers state to experience work-related stress at least "most of the time". A similar proportion also claim that their job may negatively affect their physical and psychological health and general well-being. Those evidence are usually collected by means of self-reported answers to several questions on working conditions which are often gathered on Likert scales. Implementing the CUB models, a class of mixture models suitable for ordinal data, this study aims to analyse the perception of occupational stress across European Union countries employing information from the European Working Conditions Survey dataset. Modelling results distinguishing by gender and job sustainability are presented. Results show that overall, there is a quite high perception of work-related stress among European workers, who reported their answers on work-related stress with a high degree of certainty. Results by gender show that on average, feeling estimations for women, is slightly higher than those for men. Looking at perceived job sustainability, some countries show higher levels of feeling for respondents who consider their jobs non-sustainable.

1. Introduction and theoretical framework

People spend a prominent share of their adult lives at work; therefore, work environment and other job-related features impact workers' perceived well-being. Working conditions influence employees' physical and psychological health and inevitably affect their everyday life. As established by a wide literature, employees' well-being is fundamental for their quality of life (among many others EU-OSHA, 2013; Helliwell et al., 2021).

[1] The paper is a joint work of the Authors. §1 has been the effort of collective writing; §2 may be attributed to Stefania Capecchi; §3 to Francesca Di Iorio; §4 to Nunzia Nappo. Supported by grant SI-WCWB from University of Naples Federico II (FRA 2022), DR n 3429, 07/09/2023 (CUP: E65F22000050001).

Work factors associated with psychosocial risks include excessive workload and pace, job insecurity, lack of flexibility in working hours, which may be unpredictable or incompatible with a normal social life, poor interpersonal relations, lack of participation, an unclear role in the organisation, limited prospects for professional development, and potential conflicts between family and work demands (Cox and Griffiths, 2005). Exposure to psychosocial risks can be a source of stress among workers, resulting in a decline in their performances and, in the case of chronic exposure, the occurrence of serious health problems. According to the European Agency for Safety and Health at Work (EU-OSHA), work-related stress emerges when the demands of the working place exceed the ability of workers to cope with them (or to control them) (EU-OSHA, 2009 p. 14).

Such issue is now studied as a main societal phenomenon and stress is considered as a question of public health (among others, see Vercamer, 2018) even with respect to its implied costs. Stress affects workers' performance and causes absence from work. If prolonged, it can lead to cardiovascular or musculoskeletal disorders. All this comes with a price: the main individual costs are associated with damage to health, lower incomes and a worsening quality of life. Organisations, on the other hand, may suffer from the associated expenses of absenteeism, presenteeism, reduced productivity or high staff turnover. More specifically, the losses per year in the European Union (EU), with respect to EU15 member states, were estimated at about 26.47 billion of Euros in 2014 (Hassard et al., 2018).

Information on self-assessed stress is examined by means of a statistical modelling framework, accounting for both the perception and the uncertainty components in the response pattern. In this study, we aim to address the issue of work-related stress perception across countries at EU 28 level. The topic is also approached by analysing perceived stress distinguishing by gender and in the light of how "sustainable", in the medium-long term, respondents consider their job to be.

2. Sources and Methods

Data from the Sixth European Working Condition Survey are employed, focusing on (self-assessed) occupational stress. The European Foundation for the Improving of Living and Working Condition (Eurofound, an Agency of the EU) carries out the EWCS every five years since 1991, providing a wide range of evidence on job and workers' characteristics across Europe.

The whole dataset[2] comprises 44,000 respondents, in 35 countries, interviewed between February and September 2015, therefore, the available information offer a thorough picture of "pre-Covid 19 Europe at work", thus denoting more "ordinary" job settings. The survey questionnaire covers several topics, including individual characteristics and conditions, working time, exposure to physical and psychosocial risks, work organization, work-life balance, self-assessed health and well-being. Our response variable stems from question Q61m: "Do you experience stress in your work?", as measured on a 5-point scale ranging from "always" to "never". In the following, the wording scale has been reversed to improve readability of the results.

Several subjective and occupational determinants can be associated with stress perception, apart from the basic socio-demographic covariates: job characteristics and related risks, work-life balance and unpaid duties and other tasks, if any, performed by the respondents, should be considered. However, building on a previous research on self-assessed health (Capecchi et al., 2021), we chose to analyse how stress perception could differ across countries, differentiating by several determinants starting from the previously studied subset. In the present study, for space constraints, we present modelling results distinguishing only by gender and job sustainability. The gender covariate (question Q2a) is expressed by the usual dummy variable (where female = 1), and *sustainable job* stems from the dummy variable obtained from question (Q93): "Do you think you will be able to do your current job or a similar one until you are 60 years old?" (if respondent is older than 55, the item wording is: "Do you think you will be able to do your current job or a similar one in five years time?"). This latter item is meant to grasp the manifold features influencing job sustainability in a medium to long-term perspective and while ageing.

A preliminary screening for missing values of the selected variables, lowers the original sample to 21,118 respondents where 47.4% are men, 64.3% of respondents state to work in the private sector and about 70.2% affirm that they will be able to perform their current job (or a similar one) in the future. As it can be observed from Table 1, in general, almost 11% of the sample claim to be always stressed at work; a similar proportion affirm to be never stressed. Women report a level of perceived stress (Sometimes, Most of the Time, and Always) slightly higher than that of men.

[2] Data have been downloaded from the UK Data Archive, http://discover.ukdataservice.ac.uk. The survey design and a comprehensive report can be found in Eurofound (2017). Apart from the 28 EU Member States interviewees, the questionnaire was indeed administered within the candidate countries for EU membership (Albania, the Former Yugoslav Republic of Macedonia, Montenegro, Serbia, and Turkey), and in Norway and Switzerland. At the country level, the sample size was generally 1,000, with some exceptions: Poland (1,200), Spain (1,300), Italy (1,400), France (1,500), the UK (1,600), Germany, and Turkey (2,000). Some countries, of their own initiative, financed an over-sampling: Belgium, Slovenia, and Spain (with sample sizes of 2,500, 1,600, and 3,300, respectively).

Overall, from those descriptives, the difference in distributions by gender does not seem remarkable.

Table 1 – *Distribution of level of perceived occupational stress by Gender*.

Stress	Male	Female	Total
Never	13.21	11.99	12.57
Rarely	21.57	18.38	19.89
Sometimes	38.87	41.09	40.04
Most of the time	15.75	17.74	16.80
Always	10.60	10.79	10.70
Total	100.00	100.00	100.00

On the other hand, as it can be seen in Table 2, a more noticeable difference can be appreciated with respect to job sustainability assessment. In fact, workers claiming to perform a non-sustainable job and to be always stressed are almost the double of the corresponding workers holding a sustainable job. The reverse, with a lower gap, can be observed among workers "rarely" stressed.

Clearly, a simple descriptive overview could not allow for an in-depth analysis, and given the nature of the responses, it seems appropriate to also implement a modelling approach for ordinal data. The most implemented models for ordinal data usually focus on the cumulative distribution of the responses (Agresti, 2010). In such modelling structures, it is essential to assume that the answer probability of an individual with specific attributes (covariates) lies into a specific category; then, the transformed cumulative probabilities are generally assumed as a linear function of the individual drivers.

Table 2 – *Distribution of level of perceived occupational stress by Job Sustainability*.

Stress	Non-Sustainable Job	Sustainable Job	Total
Never	11.29	13.12	12.57
Rarely	15.74	21.66	19.89
Sometimes	36.57	41.51	40.04
Most of the time	20.60	15.18	16.80
Always	15.80	8.54	10.70
Total	100.00	100.00	100.00

An alternative framework allows to consider jointly the evaluation of a specific rating and the intrinsic uncertainty, even without inserting the individual characteristics. To this aim, we can suitably implement the CUB models, the acronym standing for Combination of Uniform and Binomial. One of the main features of the CUB models, especially when estimated without covariates (denoted

as CUB(0,0)), consists in the effective graphical representation of the results. A CUB(0,0) is specified by two parameters, both lying in the interval (0;1); therefore, each model can be represented as a point in the unit square, the estimated coefficients being the point coordinates. The rationale behind such framework is that respondents, when selecting a category out of an ordinal scale, are driven both by their latent perception of the item and by the uncertainty, intrinsic to any decision-making process. Briefly, CUB models assume that the data generating process of a discrete choice is given by the combination of two features: the *feeling*, expressing substantial agreement/disagreement (satisfaction/dissatisfaction; and so on) and the *uncertainty* components (for a wide review: Piccolo and Simone, 2019).

Formally, let (R_1,… ,R_n) be a sample of observations collected on an ordinal scale of m≥ 3 categories, e.g. answers to a given item of a questionnaire or a ranking, so that R_i denotes the score assigned by the *i*-th respondent. We say that R_i is a CUB distributed random variable with uncertainty parameter $\pi_i \in (0,1]$ and feeling parameter $\xi_i \in [0,1]$ if:

$$\Pr(R_i = r|\pi_i, \xi_i) = \pi_i \binom{m-1}{r-1} \xi_i^{m-r}(1-\xi_i)^{r-1} + (1-\pi_i)\frac{1}{m} \qquad r = 1 \dots m$$

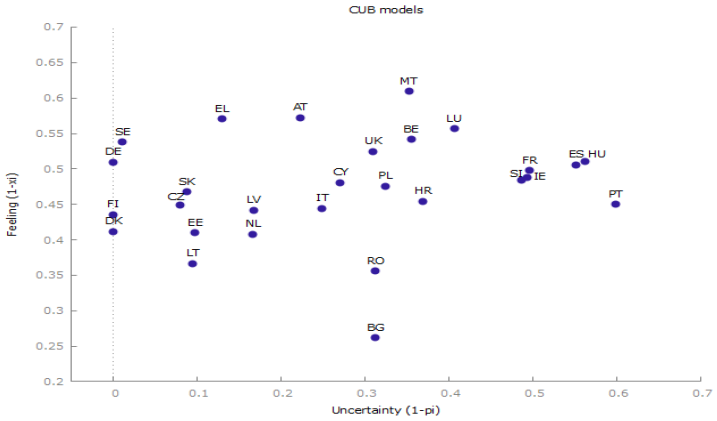
Thus, the mixing proportion 1-π_i is referred to as the *uncertainty* parameter, although 1-π_i measures the weight of the uncertainty of the responses since it associated with the importance of the discrete Uniform in the mixture. In the simple version of CUB model, $\pi_i = \pi$ and $\xi_i = \xi$ are constant among subjects/units. Thus, CUB models allow to characterize different rating/ranking responses in terms of two parameters (1-π; 1-ξ) only, ranging in (0;1]×[0;1], yielding a scatter plot, which easily allows comparisons among selected groups/classes of interviewees. With respect to global measures of fitting, the estimation procedure relies on likelihood methods and dedicated packages are available in R, Stata and Gretl (see Iannario et al., 2018, Cerulli et al., 2022, Simone at al., 2019).

3. Results of the estimated models

From the subset of data as described above, the estimated CUB models for perceived occupational stress for each of the 28 countries of the European Union are depicted in Figure 1. As far as *feeling* parameter is concerned, with very few exceptions, European countries range from 0.40 to 0.62. This circumstance indicates a substantial homogeneity in the perception of work-related stress. A different response behaviour is that of Bulgaria, with a value of feeling much lower, equal to

0.30, and of Romania and Latvia. The highest level of *feeling* is recorded for Malta, immediately followed by Greece and Austria.

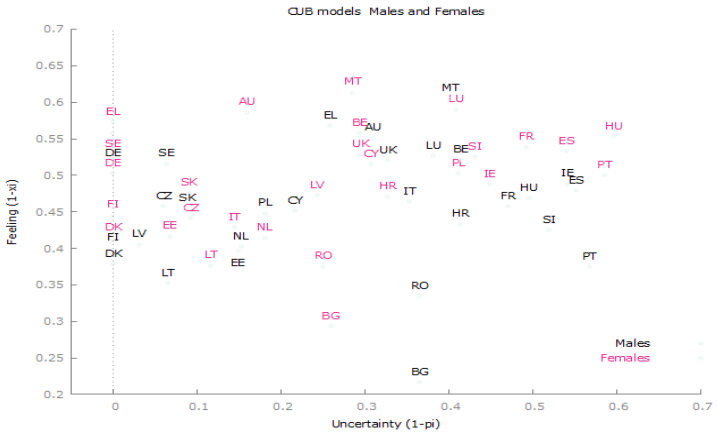
Figure 1 – *Estimated CUB models for stress perception across EU 28 countries.*



The response behaviour is quite varied with regard to uncertainty. First of all, the variability of the estimates ranges from 0.00 to 0.60, the models showing a distinct distribution by countries. First, it can be observed that the Northern European countries (Denmark, Finland, Sweden) and also Germany are characterised by no uncertainty. While, at the opposite extreme, Portugal, Hungary and Spain display the highest level of uncertainty, closely followed by France, Ireland and Slovenia. In general, it can be said that most countries exhibit low uncertainty, which is below 0.35 for more than 2/3 of the countries. Thus, we may affirm that the inherent heterogeneity on stress perception is limited for most countries, implying that the interviewed workers have expressed their responses rather decisively.

The estimation of the CUB models for each country separately by gender (Figure 2) seems to reinforce what was said above: the feeling parameters are between 0.40 and 0.65. An exception, again, is represented by Bulgaria, with a level of feeling below 0.30 for both males and females (0.22 for males). The same can be said for Romanian male respondents, whereas Romanian women are in line with most countries. The Portuguese respondents express a noticeable gap by gender: males are at 0.37; females reach 0.51, denoting a clearly higher perception of work-related stress. Maltese men and women share the highest value in terms of feeling, followed by Austrian and Luxembourg ones. In general, it can be argued that the gap in feeling between male and female workers is limited, for most countries, with feeling estimations for women slightly higher, on average, than those for men.

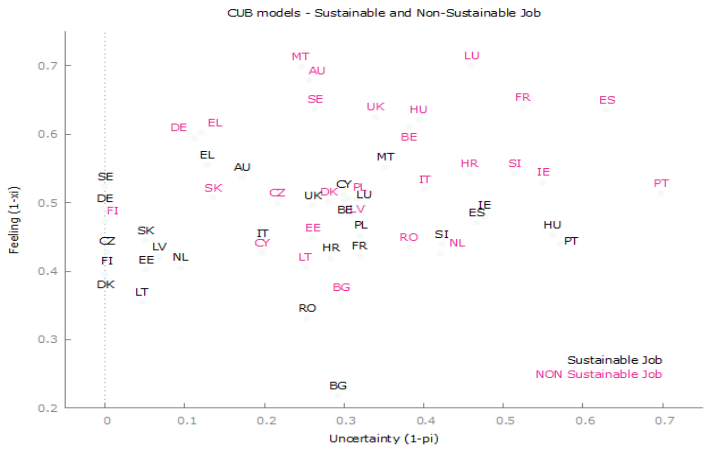
Figure 2 – *Estimated CUB models for stress perception across EU 28 countries by Gender.*



The difference by gender is significantly more pronounced with reference to uncertainty. First, it may be observed that the range for uncertainty is greater than that for feeling, and, in many countries, its level for females is distinctly different from that of males. This is the case in Italy, where women are less uncertain than men, with parameters estimated at 0.12 and 0.38, respectively. Similarly, for Greeks, an uncertainty level of 0.00 is estimated for women, while the parameter is 0.29 for men. As for Poland, the level of uncertainty for males is 0.19, while for females it reaches 0.41. At the same time, for Denmark, Finland, and Germany, both males and females display the same level of uncertainty of 0.00; while, on the opposite of the chart, Portuguese and Spanish males and females, show a level of uncertainty that rises to around 0.59. It would not seem feasible to somehow classify the countries according to the usual welfare patterns or into geographic clusters.

The picture arising from the estimations for stress with respect to the perceived job sustainability (Figure 3) is radically different. First, the ranges of the estimates, both for feeling and uncertainty, are wider: they indeed vary between 0.30 0.70 and between 0.00 and 0.70, respectively. A clear discrepancy emerges, in terms of feeling, between those who believe their jobs sustainable in the medium-long term and those who do not. In fact, a large group of countries, including Malta, Luxembourg, France, Spain and Austria, show higher levels of feeling (of about 0.15-0.20) for respondents who consider their jobs non-sustainable.

Figure 3 – *Estimated CUB models for stress perception across EU 28 countries by job sustainability assessment*



As for uncertainty, Northern European countries maintain a parameter value at 0.00, therefore showing a decisive response pattern. For Germany and the Czech Republic, the uncertainty results to be 0.00 for those who consider their employment sustainable, while it rises to 0.10 for Germans and 0.20 for Czechs who assess their jobs as non-sustainable. In Italy, this value for the uncertainty parameter increases from 0.20 to 0.40. The largest gap in uncertainty belongs to the Netherlands: the uncertainty parameter is at 0.10 for workers for whom their job is sustainable, whereas such parameter rises to 0.45 for those who consider it non-sustainable. In any case, respondents who claim their job to be non-sustainable present greater uncertainty in their response pattern.

4. Discussion and concluding remarks

Work-related stress is correlated with unsafe behavioral outcomes, among which alcohol consumption (Brown and Richman, 2012) and problems maintaining healthy relationships with one's partners and children (Repetti et al., 2009). Furthermore, stress is one of the primary causes of numerous diseases such as coronary heart disease (Sara et al., 2018). In addition, stress negatively affects workers' performance with effects not only at the individual but at organizational level too. Drawbacks of occupational stress are in contrast with goals in the workplace to support workers and organizations succeed (Meyers et al., 2013) and should worry

European countries since results of this research show that workers have a quite severe perception of occupational stress.

The main aim of this paper was to examine differences in the perception of work-related stress in EU 28 estimating CUB models based on data taken from the last European Working Condition Survey. Results show that overall, there is a quite high perception of work-related stress among European workers. Malta shows the highest level of *feeling*, while Bulgaria, Romania and Latvia the lowest. Differences in the perception of occupational stress may depend on working conditions which people face, and on the characteristics of labour markets, which are partly shaped by welfare regimes.

Before the collapse of communist regimes, labour market in Central and eastern European countries was pretty different from the rest of Europe. Before the transition to market economy, gender parity was high, wages were generally low, however workers could benefit from several social benefits. After the transition, rapid structural changes and the increasing unemployment rate were among the causes making uncertain the labour environment with labour markets characterized among other by long working hours and low wages and high job insecurity (Yarmolyuk-Kröck, 2022). Data on working conditions across central and eastern Europe show a sizable presence of psychological risks at work in different industries. However, according to Yarmolyuk-Kröck (2022), compared to other European countries, central and eastern Europe countries' workers have a different perception of psychological risks and a lack of awareness. Therefore, this different perception could be one of the reasons why our results for Bulgaria, Romania and Latvia show a low value of *feeling* (0.30). On the other hand, Malta presents the highest level of *feeling*. According to the European Commission (2018), Malta is one of the fastest growing economies in the EU. This has implied a difficult lack of labour supply in several sectors, which could be one of the causes why workers in Malta experience very long working hours, above the European average (Eurofound, 2017). However, according to Eurofound (2017), working conditions in the country are generally good compared with the European standards. Therefore, both long working hours and the perception workers have that their health or safety is at risk because of their work (Eurofound, 2017) are likely to help to explain the highest level of *feeling* within the country.

Results by gender show that overall, on average, the *feeling* estimations for women are slightly higher than those for men. Results on gender are in line with the literature (Forastieri, 2016; Nappo, 2020). Some psychosocial factors related with work-related stress may be more recurrent for women. Women play a double role at home and at work, generally they are less paid than men and are employed in more insecure job positions, they are more likely than men to deal with sexual harassment at work and domestic violence, they have to demonstrate that they are as good as

men at their jobs, those are some of the reasons why women are likely to have a higher perception of occupational related stress (Forastieri, 2016; Nappo, 2020). The most remarkable gender gap appears to that of Portuguese respondents. Such result can be explained considering that, in Portugal, working rights for women and actions against gender discrimination were formally recognized only in 1986, when Portugal adopted and ratified international organization's legal frameworks such as the Commission on the Status of Women of the United Nations, the Council of Europe and the European Economic Community. However, the effects of such ratifications still need time to produce results in terms of gender equality (Amâncio and Santos, 2021). In Portugal, women face a heavy workload within family. In the country there is still a traditional division of household, this is important for women's careers and job opportunities (Amâncio and Santos, 2021), and it affects their perception of work-related stress.

Results on work-related stress and perceived job sustainability are quite similar for countries with very different labour markets. The very content of sustainable job may be helpful to explain such results. By sustainable job is meant a job which aims to the individual's long-term employability in a position that promotes the worker's personal development over time (Lawrence et al., 2017). Therefore, it likely that non-sustainable jobs, since perceived probably as insecure jobs make workers feel stressed. Indeed, job insecurity is a main stressor for workers (Nappo et al., 2022).

Results show that, across the EU, work-related stress is a problematic issue, however, it seems still considered a subjective challenge rather than a collective/organizational one, connected to poor work organization. The EU requires further policy actions to prevent and to deal with work-related stress. In addition, more research is needed to detect workplace stressors and to encourage healthy work environments.

References

AGRESTI A. 2010. *Analysis of ordinal categorical data*, Hoboken, N.J: Wiley.

BROWN R.L., RICHMAN J.A. 2012. Sex differences in mediating and moderating processes linking economic stressors, psychological distress, and drinking, *Journal of Studies on Alcohol and Drugs*, Vol. 73, pp. 811-819.

AMÂNCIO L., SANTOS M. H. (2021). Gender equality and modernity in Portugal. An analysis on the obstacles to gender equality in highly qualified professions, *Social Sciences*, Vol. 10.

CAPECCHI S., DI IORIO F., NAPPO N. 2021. Volunteering and Self-Assessed Health within EU28 Countries: Evidence From the EWCS, *Nonprofit and Voluntary Sector Quarterly*, Vol. 50, No. 1, pp. 93-117.

COX T., GRIFFITHS A. 2005. The nature and measurement of work-related stress: theory and practice. In WILSON J.R., CORLETT N. (Eds.) *Evaluation of human work* (3rd ed.), London, CRS Press, pp. 553-571.

CERULLI G., SIMONE R., DI IORIO F., PICCOLO D., BAUM C.F. 2022. Fitting mixture models for feeling and uncertainty for rating data analysis, *The Stata Journal*, Vol. 22, No. 1, pp. 195-223.

EU-OSHA 2009. *OSH in figures: stress at work - facts and figures*. Publication Office of the European Union, Luxembourg. Available at: https://osha.europa.eu/en/publications/reports/TE-81-08-478-EN-C_OSH_in_figures_stress_a%20t_work.

EU-OSHA 2013. Psychosocial risks and workers' health, OSHWiki. Available at: https://oshwiki.eu/wiki/Psychosocial_risks_and_workers_health.

EUROFOUND 2017. European Working Conditions Survey, 2015. Data collection, 4th Edition. UK Data Service. SN: 8098.

EUROPEAN COMMISSION 2018. Spring 2018 economic forecast for Malta. Available at: https://ec.europa.eu/info/sites/info/files/economyfinance/ecfin_forecast_spring_030518_mt_en.pdf .

FORASTIERI, V. 2016. Psychosocial risks, stress and violence in the world of work. *International journal of labour research*, Vol. 8.

HASSARD, J., TEOH, K.R., VISOCKAITE, G., DEWE, P., COX, T. 2018. The cost of work-related stress to society: A systematic review, *Journal of occupational health psychology*, Vol. 23, No. 1, pp. 1-17.

HELLIWELL, J.F., HUANG, H., WANG, S., NORTON, M. 2021. World happiness, trust and deaths under COVID-19. *World Happiness Report*, pp. 13-56.

IANNARIO M., PICCOLO D., SIMONE R. 2018. The R package CUB: a class of mixture models for ordinal rating data R package version 1.1.3. Available at: http://CRAN.R-project.org/ package=CUB.

YARMOLYUK-KRÖCK, K. 2022. A case for an EU directive addressing work-related psychological risks: An eastern European perspective. *ETUI Research Paper-Policy Brief*.

LAWRENCE B.S., HALL D.T., ARTHUR M.B. 2017. Sustainable Careers Then and Now. In DEVOS A., VAN DER HEIJDEN B. I. J. M. (Eds.) *Handbook of Research on Sustainable Careers*, Edward Elgar, Cheltenham, UK, pp. 432- 449.

MEYERS M.C., VAN WOERKOM M., BAKKER A.B. 2013. The added value of the positive: a literature review of positive psychology interventions in organizations. European Journal of Work Organizational Psychology, Vol. 22, pp. 618-32.

NAPPO N. 2020. Job stress and interpersonal relationships cross country evidence from the EU15: A correlation analysis.*BMC Public Health*, Vol. 20, pp. 1-11.

NAPPO N. FIORILLO D., LUBRANO LAVADERA G. 2022. Subjective Job Insecurity During the COVID-19 Pandemic in Italy. *Italian Economic Journal*, pp. 1-27.

PICCOLO D., SIMONE R. 2019. The class of CUB models: statistical foundations, inferential issues and empirical evidence, *Statistical Method and Applications*, Vol. 28, pp. 389-475 (with discussions).

REPETTI R., WANG S.W., SAXBE D. 2009. Bringing it all back home: how outside stressors shape families' everyday lives. *Current Directions in Psychological Science*, Vol. 18, pp. 106-11.

SARA J.D., PRASAD M., ELEID M.F., ZHANG M., WIDMER R.J., LERMAN A. 2018. Association between work-related stress and coronary heart disease: a review of prospective studies through the job strain, effort-reward balance, and organizational justice models. *Journal of the American Heart Association*, Vol. 7, e008073.

SIMONE R., DI IORIO F., LUCCHETTI R. 2019. CUB for Gretl. In DI IORIO F., LUCCHETTI R. (Eds.) *Proceedings of the International Conference on the GNU Regression, Econometrics and Time Series Library*, Quaderni della Scuola delle Scienze Umane e Sociali, Vol 12, Federico II University Press, Napoli.

VERCAMER S. 2018. Stress at work. Report to committee on social affairs, health and sustainable development. Council of Europe. Available at: http://website-pace.net/ documents/19855/4491159/20181204-WorkStress-EN.pdf/6c87997b-366a-4 a92-8e80-b540d4cc06d8.

Stefania CAPECCHI, Dipartimento di Scienze Politiche, Università di Napoli Federico II stefania.capecchi@unina.it

Francesca DI IORIO, Dipartimento di Scienze Politiche, Università di Napoli Federico II francesca.diiorio@unina.it

Nunzia NAPPO, Dipartimento di Scienze Politiche, Università di Napoli Federico II nunzia.nappo@unina.it

DISABILITY FREE LIFE EXPECTANCY DECOMPOSITION: A COMPARATIVE ANALYSIS ACROSS EUROPEAN COUNTRIES

Mariarosaria Coppola, Rosaria Simone

Abstract. Life expectancies (LE) at birth and at adult ages have progressively increased in recent decades in Europe. LE is a quantitative indicator and therefore it provides a limited view of the status of the population, ignoring the weight and the effects that the health condition has on the years that remain to live. To address this issue, we refer to Disability Free Life Expectancy (DFLE) indicating the average number of years a person of a given age *x* expects to live in good health. In this paper we analyse the change in LE and DFLE at age 65 from 2004 to 2016 for a set of the most long-lived European countries, as well as for male and female populations in order to highlight gender differences. Then we quantify the extent by which age-specific mortality and morbidity contribute to changes in DFLE for understanding whether the trends observed are due to compression/expansion of morbidity or to variation in mortality patterns. This goal is important both for health policies planning and for sustainability of socio-economic and social security systems. DFLE is calculated using the Sullivan method, measuring disability in terms of the Global Activity Limitation Indicator (GALI). We refer both to step-wise and continuous change techniques for DFLE decomposition. Data sources are Human Mortality Database (HMD) for mortality rates and the European Health Life Expectancy Information System (EHLEIS) for prevalence rates. Results show that all the countries under examination present a stronger increase in LE at age 65 for men than for women. A positive variation in DFLE occurs for Belgium, Sweden, Norway, and Finland. A negative gap in DFLE is identified for Italy, Switzerland, and Spain, meaning that people lived longer but in poorer health conditions. Finally, age-specific mortality and morbidity contributions to the DFLE gaps are highlighted with an innovative visualization tool that allows to pursue a comparative analysis across countries.

1. Introduction

The aging of the population can be considered a success of the globalization, yet it represents an important challenge for governments especially with respect to retirement and welfare policies. For instance, the impact that the lengthening of life could have on the job market emerges when we observe the trend of the elderly dependency rate defined as the ratio between the population aged 65 and over, and the population in working age (15-64 years), multiplied by 100. Data for Italy (released by the National Institute of Statistics - ISTAT) record an acceleration in the increasing trend: specifically, from a value of 30 in 2004, to a value of 34,5 in 2016, up to a value of 37 in 2021.

Thus, it is equally important to value the elderly people, since they constitute a growing component of the population, and to take the longevity risk into account when designing welfare policies. From the individuals' point of view, aging can be perceived as a process that leads to uselessness. Therefore, a new rationale is advocated, based on the idea that the aging phenomenon can be an opportunity to find new ways to thrive and not only as a catastrophic event. In this framework, the study of the elderly of today (and even more of those of tomorrow) has to go beyond the quantitative perspective of age, using dynamic indicators that consider parameters that change over time, such as the state of health and socio-economic conditions. The World Health Organization (WHO) has introduced the term *active aging* understood as the "*process of optimizing opportunities for health, participation and security to gradually improve the quality of life as people get older*". The term "*active*" implies not only the ability to be physically involved but above all a constant participation in the social, economic, and cultural life (WHO, 2002).

Life expectancy (LE) is a widely used tool to assess the well-being of a population. It is a quantitative indicator and represents a measure of how long, on average, a person of a given age x can expect to live, assuming constant age-specific mortality rates; however, it does not give us any information on the quality of these years. The need arises to search for additional measures that could reveal the state of well-being of a population in qualitative terms. To meet this need, in the 70s Sullivan (Sullivan, 1971) proposed a method to calculate the number of remaining years, at a particular age, which an individual can expect to live in a healthy state.

When we talk about Healthy Life Expectancy (HLE) we refer to a family of indicators relating to life expectancy for a given state of health, the most common of which for studying the state of well-being at older ages is the Disability-Free Life Expectancy (DFLE). DFLE (called Healthy Life Years (HLY) according to the Eurostat definition) is a standardized indicator based on the study of limitations in carrying out daily activities. In recent years, many studies have been devoted to the

measurement of the components that determine the life expectancy in health. In particular, researchers have provided tools that allow to establish how much these components influence the differences in LE and DFLE for different populations, or for the same population at different times (Van Raalte and Nepomuceno, 2020). In this regard, the main interest lies in assessing the extent by which these variations are due to changes in morbidity or to changes in mortality (Vaupel, 1986). For this purpose, to the best of our knowledge there are two main methodologies: the step-wise decomposition algorithm (Andreev *et al*., 2002) and the continuous change technique (Horiuchi *et al*., 2008).

In this paper we contribute to the state of the art on the topics by presenting a comparative analysis of the evolution of DFLE at age 65, calculated by the Sullivan's method for a set of European countries selected among the most long-lived ones. Age 65 has been chosen since historically it is an approximate benchmark for retirement in several countries. In details, we first calculate the variation in both LE and DFLE from 2004 to 2016 and assess the evolution of the gender gap for both indicators. Some previous studies in this respect have been conducted for European countries by Robine *et al*., 2001, and for the United States in Crimmins *et al*., 1997. Then we quantify the age-specific contributions of morbidity and mortality to the gap in DFLE between the two time periods, for both male and female populations and for each country. Results are presented with simple yet original graphical tools.

The reminder of the paper is organised as follows: a brief introduction to LE and DFLE is provided in Section 2, where the Sullivan's method is also described; decomposition methods for DFLE gap are shortly recalled in Section 3. Section 4 provides results and comments related to the comparative analysis across the selected countries of the evolution in LE and DFLE gaps for male and female populations. Then a discussion on the decomposition of the DFLE gap follows. Finally, Section 5 summarizes results and provides concluding remarks.

2. Life Expectancy and Healthy Life Expectancy: definition and calculation

As explained in the previous section, LE at age x represents the expected number of years an individual aged x can live. The main differences among the several methods that can be used to calculate this measure concern the calculation of the probability of survival in the first year of life and the determination of the top age ω of the mortality table. Given these premises, the life expectancy at age x, $e_{x,}$ is defined by:

$$e_x = \frac{1}{l_x}\sum_{i=x}^{\omega} L_i \qquad (1)$$

where l_x is the number of survivors at age *x*, and for each age $i = x, \dots, \omega$, L_i denotes the total number of *person-years* lived by the individuals from age i to age $i+1$. For L_i calculation, it is generally assumed that the l_{i+1} survivors at age $i+1$ contributes with a whole year, whereas each of the d_i individuals who die within the age interval *(i,i+1)* contributes with half a year on average, so that $L_i = l_{i+1} + 0.5\ d_i$. For details and examples, see https://www.lifeexpectancy.org/lifetable.shtml.

Starting from the definition of LE as "*the average number of years remaining to live*", it is possible to divide these remaining years of life into years spent in good or ill health. The concept of healthy life expectancies as health indicators was introduced for the first time by Sanders in 1964 (Sanders, 1964). Then Sullivan in 1971 proposed a method for DFLE calculation which is the most popular and widely used still today to study the state of well-being at older ages. Sullivan's method assumes that life follows a one-way process from good health to death, passing through a state of ill-health.

DFLE is a standardized indicator based on the study of limitations in carrying out daily activities. For its determination by Sullivan's method, morbidity data as age-specific and health condition-specific prevalence rates are needed, along with age-specific mortality rates. In the following, we refer to the Human Mortality Database to obtain the mortality rates (HMD, 2023) and to the European Health Life Expectancy Information System (EHLEIS) for prevalence rates. In particular, disability is measured starting from the Global Activity Limitation Indicator (GALI) obtained by surveying the following question: *"Because of health problems, to what extent do you have limitations, which last for at least 6 months, in the activities that you usually carry out?".* Response options are: *"Serious limitations", "Minor limitations", "No limitations"*. These measurements are collected within the European Statistics on Income and Living Conditions survey (EU-SILC). Then, since Sullivan's Index requires a binary variable (presence/absence of disease, presence/absence of disability and so on), the response variable is dichotomized into two categories*: "No limitations"* and *"With some Limitations"*.

As a matter of fact, GALI is a subjective indicator: people affected by the same pathologies can report levels of different functional capabilities. For this reason, among others, its reliability can be argued (Berger *et al.*, 2015). Nevertheless, it allows to grasp how much the restriction in activities affects the physical and mental state of people and, to the best of the Authors' knowledge, there are no alternative indicators with some property. Most importantly, subjective evaluations play an important role in accounting for active ageing in the assessment of health of the elderly, circumstance that must be taken into consideration when investigating our topics.

Hereafter, for each age *x*, we set:

1. ${}_nL_x$ = the number of person-years lived in the age interval *(x; x+n)*, (usually *n=5* is set, apart from the first two and the last age intervals). This measure is obtained summing the total person-years L_y for each single age *y* in the interval *(x; x+n);*
2. ${}_n\pi_x$ = the prevalence rate in ill health for the age interval *(x; x+n).*

Then, life expectancy in bad health at age *x* (*Unhealthy*) can be calculated as:

$$e_x^U = \frac{1}{l_x}\sum_{i=x}^{\omega} {}_nL_i \times {}_n\pi_i \quad (2)$$

whereas Life expectancy in good health (*Healthy*) is obtained as:

$$e_x^H = e_x - e_x^U = \frac{1}{l_x}\sum_{i=x}^{\omega} {}_nL_i\ (1-{}_n\pi_i). \quad (3)$$

Note that other methods can be used for DFLE calculation such as the multistate methods (Saito et al, 2014), according to the data source available (panel or cross sectional). Some criticisms of Sullivan's model are due to the fact that the underlying assumptions limit the representation of people's functional status to prevalence rates influenced by past conditions instead of using measures such as incidence rates allowing to calculate a pure current proportion of unhealthy people. Anyway, it has been shown that the differences in estimates between Sullivan's method and methods using incidence rates (panel data) do not arise from the use of prevalence rates, but from the use of prevalence estimates: indeed, these are influenced by past conditions, in an open population. However, Sullivan's method provides good estimates if the changes in either prevalence and incidence rates are regular over the long term (Mathers and Robine, 1997). In addition, the simplicity as well as the ability to use easily accessible data such as prevalence data make it the most popular and widely used method.

3. Decomposing the gap in DFLE

Decomposition is a widely exploited methodology in Demography to study the gap in an index by analysing the contribution of the variation of each of its components to the overall change. The gap can refer either to the same population in different time periods or to different populations in the same time period.

In other words, the problem of decomposition results in the attribution of the observed difference between two aggregated indexes for two populations (or the same population at different times) to the contribution of the underlying parameters. In the case of the Sullivan's Index, the decomposition allows to know the

contribution given by the variations in mortality and morbidity to the gap observed between two indices. In this way we can understand whether the trends observed for life expectancy in health at a given age are due to a compression/expansion of morbidity or variations in mortality for subsequent age intervals.

Two decomposition methods have been proposed in literature for the Sullivan's Index. The former is the *step-wise decomposition* introduced in Andreev *et al*. (2002). This procedure replaces one at a time each of the parameters (age-specific contribution of mortality and morbidity in the present study) referring to one population (or time period) with the elements of the parameter vector from the second population (or time period). After each step, it recalculates the DFLE to estimate the contribution of the substituted parameter. The latter is the *continuous change* or *line integral model* developed by Horiuchi *et al*. (2008). It assumes that age-specific mortality and morbidity contributions change continuously along a hypothetical dimension, along which decomposition is performed based on multiple smaller time intervals. Then, the procedure estimates the variations needed by parameters to change the final aggregate value of DFLE from one population to another, or if we refer to the same population, from a time period to another. For more deepening on the topics, see also Van Raalte and Nepomuceno (2020).

To the best of the Authors' knowledge, the implementation of both methods has produced similar results in all demographic applications: therefore we can consider that there is no valid reason to prefer one methodology over the other.

For each population, the decomposition of the DFLE gap between the two time periods has been performed by means of the 'DemoDecomp' R package (Riffe, 2018).

4. A comparative analysis across EU countries.

In this section we first calculate the LE and the DFLE at age of 65 for male and female populations by the Sullivan's method for years 2004 and 2016 and for each of the selected countries. Time periods have been chosen on the basis of the available data from the EU-SILC survey. Then, we decompose the DFLE variation from 2004 to 2016 to determine the extent by which the change in each parameter (morbidity/mortality) contributes to the overall gap in the index.

We considered some of the most long-lived EU countries: Norway, Sweden and Finland for the Northern Europe, Belgium and Switzerland for the Central Europe, and Spain and Italy for the Southern part. Ages between 65 and 110 have been grouped into 5-year age intervals, with the addition of an open range one (85+): note that both e_x and $DFLE_x$ refers to the beginning of each age interval.

Table 1 shows that all the countries under examination present an increase in LE at age 65, stronger for men than for women. Belgium, Sweden, Norway, and Finland show a positive variation of DFLE, which is very marked for Norway and Sweden. Conversely, Italy, Switzerland and Spain present a negative gap in the DFLE, meaning that people live longer but the years lived without limitation have decreased.

Table 1 – *Gap in LE and DFLE at age 65 from 2004 to 2016, by country.*

	Female		Male	
Country	ΔLE	ΔDFLE	ΔLE	ΔDFLE
ITA	1	-2,25715	1,67	-1,22397
BEL	1,38	1,45967	1,82	1,66155
CHE	0,83	-3,14947	1,39	-2,22867
ESP	1,67	-0,7016	1,85	-0,07262
FIN	1,11	0,68061	1,58	0,95315
NOR	0,95	2,05546	1,92	2,8602
SWE	0,86	2,38447	1,62	2,87648

Figure 1 displays country-specific gender gap in LE and DFLE in 2004 and in 2016 as points in a scatterplot: by assessing the distance to the bisector line, results indicate that gender gap slightly reduced from 2004 and 2016 for all countries in a similar way, for DFLE to a greater extent than for LE.

In order to grasp the relative importance of DFLE compared to LE, Figure 2 displays the ratio between these indicators in 2004 and in 2016. It follows that overall, there is no relevant variations between men and female populations. Countries can be grouped in two sets: the 'continental' ones, for which DFLE increased relative to LE from 2004 to 2016 (standing over the bisector line), and the Scandinavian ones (except for Finland), for which DFLE decreased relative to LE between the two time periods (standing below the bisector line).

Figure 1 – *Gender gap in $DFLE_{65}$ and LE_{65} in 2004 and 2016, by country.*

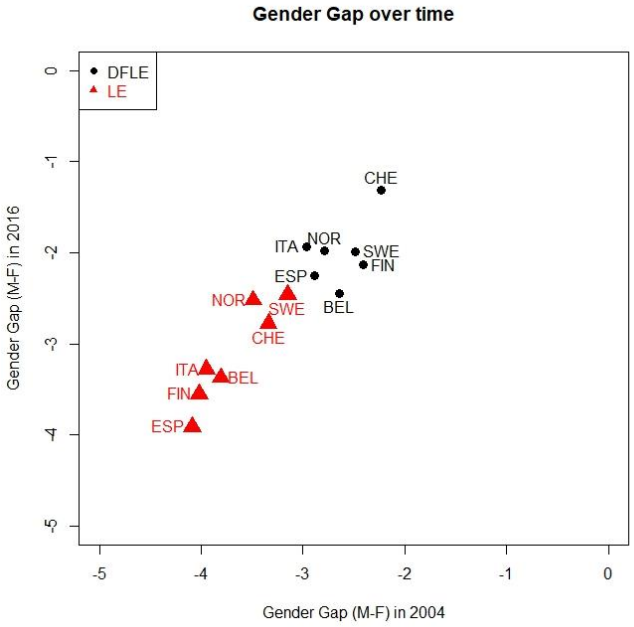
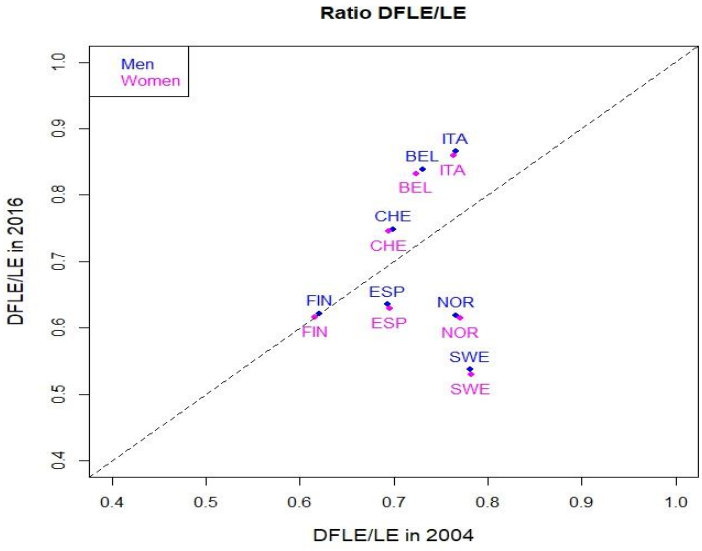
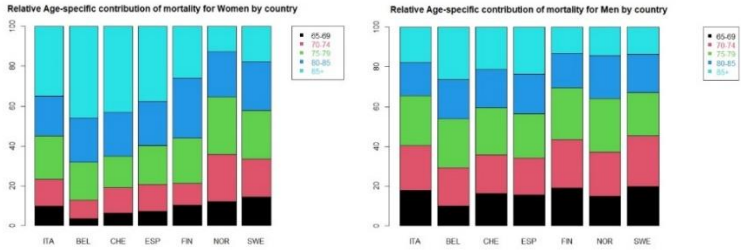


Figure 2 – *Ratio $DFLE_{65}/LE_{65}$ in 2004 and 2016, by country.*



For investigating to what extent mortality and morbidity have determined the overall variation in DFLE between 2004 and 2016, we performed the DFLE decomposition by the stepwise and the continuous change methods. As expected, we obtained similar results, so only those related to the stepwise method are reported. Figures 3 and 4 illustrate age-specific contributions of mortality and morbidity components to gap in DFLE at age 65 for women and men by country, relative to the total contribution of mortality and morbidity to DFLE gap, respectively. For mortality, the contribution of every age interval is positive, whereas for morbidity some negative values are observed. For this reason, Figure 4 considers the absolute value of age-specific contribution to morbidity, relative to their total. In order to preserve information about the sign of the contribution, dashed areas indicate negative variations.

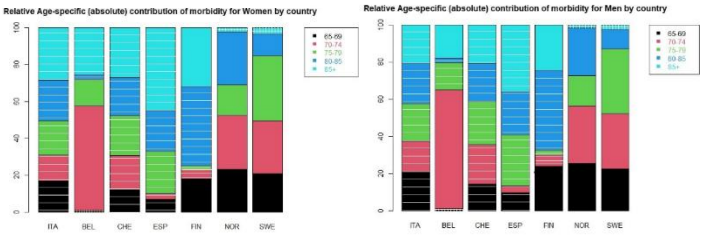
Figure 3 – *Age-specific contribution of mortality to gap in $DFLE_{65}$ by country.*



We can observe that the negative gap estimated for DFLE from 2004 to 2016 for Italy and Spain is mainly due to a significant worsening of morbidity (respect to the improvement in mortality) in all age groups for both men and women (especially in the 85+ range for women). The overall negative gap in the DFLE for Switzerland is due to a strong negative variation of the morbidity component compared to a lower improvement in mortality for all age groups, regardless of gender.

As far as Belgium, on one hand Figures 3 and 4 show that the positive DFLE gap for women is mainly determined by a strong improvement in morbidity in the 70-74 age group while for the 85+ age group mortality plays the main role. For the male population, instead, Figures 3 and 4 reveal that both morbidity and mortality improved, but mortality did so to a greater extent.

Figure 4 – *Age specific (absolute) contribution of morbidity to gap in $DFLE_{65}$ by country (dashed areas indicate negative variations).*



When we consider Finland, we observe that the positive gap in the DFLE is due to the improvement of mortality in all age groups and of morbidity in the 65-69 and 85+ ranges, despite the sharp worsening of morbidity in the 80-84 class for both males and females.

For Norway male population, we observe that the positive contribution of mortality is greater than that of morbidity in the 80-84 range; morbidity contributes negatively to the overall variation of the index only in the last age group. For females a similar conclusion can be drawn, with the difference that also in the 80-84 age group the morbidity contribution in the overall variation of the index is greater than that of mortality.

Finally, for Sweden we observe that, for both male and female populations, morbidity has contributed more strongly to the positive variation in DFLE with respect to mortality, which has improved in all age group. Only in the 85+ range a worsening of morbidity has been observed compared to an improvement in mortality, which is more significant for males than it is for females.

5. Concluding remarks

Given the longevity risk characterizing nowadays most of the developed countries, monitoring the evolution of life expectancy within the population is of foremost importance for public health, as well as for social security system and welfare, in order to promote also the so-called active ageing (Arriaga, 1984).

In this setting, the paper is framed within research on healthy life expectancy at older ages and its relative difference with life expectancy. A comparative analysis among a selection of long-lived EU countries and across gender allowed to reveal the extent by which mortality and morbidity improvements for age groups contribute to overall improvement in disability free life expectancy at age 65. Innovative

visualization of results of decomposition methods are proposed to foster monitoring of the evolution of these important demographic indicators over time and countries.

References

ANDREEV E.M., SHKOLNIKOV V.M., BEGUN A.Z. 2002. Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates, *Demographic Research*, Vol. 7, pp. 499–522.

ARRIAGA E.E. 1984. Measuring and explain the change in life expectancies, *Demography*, Vol. 21, No. 1, pp. 83-96.

BERGER N., VAN OYEN H., CAMBOIS E., FOUWEATHER T., JAGGER C., NUSSELDER W., ROBINE J.M. 2015. Assessing the validity of the Global Activity Limitation Indicator in fourteen European countries, *BMC medical research methodology*, Vol. 15, No. 1. https://doi.org/10.1186/1471-2288-15-1.

CRIMMINS E.M., SAITO Y., INGEGNERI D. 1997. Trends in disability-free life expectancy in the United States, 1970-90, *Population and Development Review*, Vol. 23, No. 3, pp. 555-5872.

HMD. 2023. Human mortality database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at: www.mortality.org. Data downloaded on May 2023.

HORIUCHI S., WILMOTH J.R., PLETCHER. S.D. 2008. A decomposition method based on a model of continuous change, *Demography,* Vol. 45, No. 4, pp. 785–801.

ISTAT. 2023. http://dati.istat.it/.

MATHERS C.D., ROBINE J.M. 1997. How good is Sullivan's method for monitoring changes in population health expectancies, *Journal of Epidemiology and Community Health* (1979-), 51.1, pp. 80–86. ISSN: 0143005X, 14702738. Available at: http://www.jstor.org/stable/25568408.

RIFFE T. 2018. DemoDecomp: Decompose demographic functions. R Package version 101. Available at: https://rdrr.io/cran/DemoDecomp/.

ROBINE J.M., JAGGER C., ROMEU I. 2001. Disability-free life expectancies in the European Union countries: calculation and comparisons, *GENUS*, Vol. 57, No. 2, pp. 89-101.

SAITO Y., ROBINE J.M., CRIMMINS E.M. 2014. The methods and materials of health expectancy, *Statistical Journal of the IAOS, Vol.* 30, No. 3, pp.209-223. DOI: 10.3233/SJI-140840.

SANDERS B.S. 1964. Measuring community health levels, *American Journal of Public Health and Nations Health*, 54.7, pp. 1063–1070.

SULLIVAN D.F. 1971. A single index of mortality and morbidity. *HSMHA health reports*, 86.4, p. 347.

VAN RAALTE A., NEPOMUCENO M.R. 2020. Decomposing gaps in healthy life expectancy. In JAGGER C., CRIMMINS E.M., SAITO Y., DE CARVALHO Y.R.T., VAN OYEN H., ROBINE J.M. (Eds.) *International handbook of health expectancies.* Springer, pp. 107-122.

VAUPEL J.W. 1986. How change in age-specific mortality affects life expectancy, *Population Studies*, Vol. 40, No. 1, pp. 147-157.

WORLD HEALTH ORGANIZATION (WHO). 2002. Active Age a policy framework. *A contribution of the World Health Organization to the Second United Nations World Assembly on Ageing.* Available at: https://extranet.who.int/agefriendlyworld/wp-content/uploads/2014/06/WHO-Active-Ageing-Framework.pdf.

Mariarosaria COPPOLA, Dipartimento di Scienze Politiche, Università degli Studi di Napoli Federico II, m.coppola@unina.it

Rosaria SIMONE, Dipartimento di Scienze Politiche, Università degli Studi di Napoli Federico II, rosaria.simone@unina.it

FRAILTY IN EUROPE: THE ROLE OF SOCIAL NETWORKS AND ACTIVITY PARTICIPATION

Rossella Miglio, Chiara Puglisi, Rosella Rettaroli, Giulia Roli, Francesco Scalone

Abstract. Frailty represents a condition of increased vulnerability among elderly people due to losses in one or more domains of the human functioning. Population aging and the consequential increase in the proportion of elderly, are making frailty a public health priority, especially in the European Union, where the proportion of elderly is projected to increase significantly in the next 30 years. Despite the numerous studies on frailty and its determinants, scholarly interest towards the role of the social determinants, specifically, is recent, with most studies focusing on the Asian context. Thus, evidence on the social determinants of frailty in the European context is scarce. In light of this, the study we propose aims at evaluating the role of social connectedness and activity participation on frailty among the elderly in Europe, using suitable regression techniques. We employ data from the Survey on Health, Aging and Retirement in Europe (SHARE), providing information on the health, socioeconomic conditions and social life of community-dwelling individuals aged 50 or older in European countries. Specifically, we study how social networks (SN) and activity participation (AP) in wave 4 affect frailty of people aged 60 and over in wave 5 by gender, separately. We measure SN and AP via two indexes proposed in the literature. To measure frailty, we rely on Romero-Ortuno et al.'s SHARE Frailty Instrument, providing pre-calculated, population-representative and gender-specific frailty indexes. Our findings hint at the existence of a substantial association between both social connectedness and activity participation and frailty. However, this association appears to change across genders. When comparing the role of social connectedness and activity participation, a more prominent association with activity participation emerges.

1. Introduction

The concept of frailty refers to a condition of increased vulnerability among older people, entailing a high risk of adverse health outcomes (such as falls, hospitalization and death) due to losses in one or more domains of the human functioning (Bessa et al., 2018). Population aging and the consequential increase in the proportion of elderly are making frailty a public health priority, especially in the European Union,

where the proportion of elderly is projected to increase significantly in the next 30 years (Gobbens et al., 2010).

Elderly's physical and mental health, mortality and their risk of entry into institutionalized care have all been shown to be strongly associated with aspects like social support, engagement, and network structure (Wenger, 1997). Numerous studies, in particular, have found a link between social networks, engagement and support and, respectively, health (Berkman et al., 2000; Seeman et al., 1996) and survival (Berkman, 1995; Eng et al., 2002). Possible mechanisms behind this association include physiological, behavioral, and mental/cognitive ones. At the physiological level, social isolation and disconnectedness are considered chronically stressful conditions to which the organism reacts via a higher rate of ageing and functional decline, thus influencing life expectancy and mortality (Berkman, 1988; Berkman et al., 2000). At the behavioral level, social engagement induces health-promoting behaviors, with social networks representing a relevant source of health-related information and advice, with consequential effects on individuals' decisions concerning their own health and wellbeing (Eriksson et al., 1999). For instance, there is evidence of positive associations between social support on the one hand and physical exercise (Treiber et al., 1991) and smoking cessation (Murray et al., 1995) on the other. Lastly, social networks and ties have been argued to affect individuals' mental health via the provision of self-esteem, self-efficacy, and sense of security (Thoits, 2011). In addition, they tend to enhance the perception of surrounding support and, where needed, company and assistance (Berkman et al., 2000; Garcìa et al., 2005), consequentially buffering the negative effects of social isolation on depressive symptoms (Liu et al., 2016; Vance et al., 2005) and cognitive functioning (Miceli et al., 2019)

A strand of literature has further investigated the role played by engagement in social, physical, and cognitive activities and mental and physical health. For instance, beyond the documented positive effects of physical activity on health and cognitive functioning (Blake et al., 2009), participating to sport clubs entails a "social doing" that is argued to be especially beneficial thanks to aspects like reduction in social isolation and loneliness and the increase in the possibility for socialization (Jenkin et al., 2018). Similarly, various studies have highlighted a positive effect of volunteer work and charity activities on various health outcomes (Haski-Leventhal, 2009; Morrow-Howell et al., 2003). Mechanisms behind this effect include the provision of purpose, sense of fulfilment, direct social interaction, and reciprocal support to those engaged in these activities. Indeed, consistently with the "role theory" (Anderson et al., 2014; Choi & Kim, 2011), the active engagement entailed in volunteer and charity activities is especially important for elderly individuals, experiencing the cessation of life-long roles of paid workers, family caretaker or spouses due to events such as retirement, widowhood, and a reduction

in the (grand)parental responsibilities following the aging of both children and grandchildren. Related mechanisms are argued to be behind the association between "formal leisure activities" (Misener et al., 2010, p. 271), including participation in various kinds of clubs and organizations, and physical and mental health (Lei et al., 2022; Munford et al., 2017).

Despite the ample evidence concerning the role played by social support and engagement in various kinds of activities for health, mortality, and the risk of entry into institutionalized care, scholarly interest towards the role that these aspects play specifically for frailty is recent, with most studies exploring the association in the Asian context (Chen et al., 2014; Makizako et al., 2018; Sun et al., 2022; Takatori & Matsumoto, 2023; Xie & Ma, 2021) and very few works looking at the West (Etman et al., 2015; Watts et al., 2017). Thus, evidence on the social determinants of frailty in the European context is scarce. In light of this, relying on recent evidence of a substantial but different role of social connectedness and activity participation on cognitive functioning (Litwin & Stoeckel, 2016), this study aims at exploring the relative influence of social connectedness and activity participation on frailty among elderly in Europe, by relying on two distinct and comprehensive measures, to our knowledge never employed before in the study of frailty.

2. Data and methods

We employ data from the 4th (2011) and 5th (2013) wave of the Survey on Health, Aging and Retirement in Europe (SHARE) (Börsch-Supan 2022a 2022b), a longitudinal survey providing information on health, socioeconomic conditions and social life of community-dwelling individuals aged 50 or older in European countries. Our sample considers a total of 8,677 men and 8,616 women aged 60 and older, across 13 European countries.

We measure social connectedness and activity participation via two indexes proposed by Litwin & Stoeckel (2016), namely the Social Network (SN) scale and the Activity Participation (AP) scale. The SN scale represents a composite measure including aspects regarding social network size, physical proximity, frequency of contacts, support, and number of types of relationships. The AP scale, rather, concerns the number of social, physical, intellectual, and recreational activities individuals participated to in the year before the survey. Activities included volunteer work, educational courses, sport, social, or any other kind of club, activities organized by religious organizations, political or community-related events, games that involve other people, reading, word or number games, physically vigorous sports, and moderate energy activities like gardening or going for walks. Individuals were identified as participating to the activity if they took part in it at

least once a month. The AP score ranged between 0 and 9, with higher scores representing a higher number of activities that participants engaged with.

We employ multinomial logistic regression analysis with country-fixed effects on a pooled sample of all the analyzed countries, separately for men and women, to assess and compare how social network and activity participation measured in wave 4 affect frailty status in wave 5. To measure frailty, we rely on Romero-Ortuno & Kenny's (2012) SHARE Frailty Instrument (SHARE-FI), providing pre-calculated, population-representative and gender-specific frailty index based on Fried et al.'s 2001 Phenotype Model. The index assesses frailty based on the presence of five main symptoms, namely unintended weight loss, weakness, slowness, self-reported exhaustion, and low physical activity. SHARE-FI has been validated as a reliable frailty index, classifying individuals into non-frail, pre-frail, and frail. We further included "deceased" as a fourth level of our outcome variable. Our models control for age, marital status, working status, country-specific wealth quintiles[1], number of chronic conditions, verbal fluency, and EURO-D scale of depression[2]. We estimate three models: the first looks at the "crude" effect of social networks on frailty. The second includes controls for the above-mentioned factors, to assess the "net" effect of social networks. The third further adds the activity participation variable, to assess whether it plays any role and, if so, how it modifies the effect of social networks.

3. Results

The results from multinomial regressions are reported in Table 1, separately by gender and for the three possible transitions of frailty[3], from non-frail to the other three considered levels of the outcome variable.

Among men, social connectedness is negatively associated with the risk of moving from a state of non-frail to one of pre-frail, while it does not seem to affect the risk of becoming frail, nor the risk of death. Rather, activity participation

[1] In line with previous literature (Bono & Matranga, 2019; Miceli et al., 2019), we calculated wealth as the sum of household income and total assets. Total assets were computed as the sum of real assets (including the primary residence net of mortgage, the value of other real estate, owned share of own business and owned cars) and net financial assets (such as bank accounts, stocks, mutual funds, and so forth) (Miceli et al., 2019).

[2] As a robustness check, we further estimated models where we also controlled for education, whether the person lives alone and whether the person has living children. Since results of these alternative specification remained substantially unchanged, we decided to include in our main analysis the most parsimonious models.

[3] Due to space constraints, we include here only regression results. Descriptive analyses are available upon request.

influences all outcomes, so that higher scores on the AP scale are associated with a lower risk of pre-frailty, frailty, and death.

As for women, social network appears to be negatively associated with the risk of becoming pre-frail, while no association is found with the other outcomes; in addition, the statistical significance with the risk of pre-frailty disappears after the inclusion of the AP scale variable in the model. Activity participation, in turn, seems to be an important source of protection against the risk of becoming pre-frail, as well as against the risk of becoming frail and the risk of death.

Considering Litwin & Stoeckel, 2016's finding of a reciprocal compensatory role of social connectedness and activity participation on cognitive functioning, we further estimated an additional regression model where an interaction term between SN scale and AP scale was included, to test whether the association between activity participation and frailty changed at different levels of social connectedness. However, the interaction term was not statistically significant, signaling the absence of any reciprocal compensatory role of the two factors on frailty.

Table 1 – *Effect of SN and AP on the risk of pre-frailty, frailty, and death. Man.*

Men									
	Pre-frail			***Frail***			***Deceased***		
SN Scale	0.82***	0.83***	0.85***	0.88*	0.94	0.98	0.94	1.02	1.05
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.07)	(0.07)
AP Scale			0.84***			0.73***			0.81***
			(0.03)			(0.04)			(0.03)
Age (Cont.)		1.06***	1.06***		1.11***	1.11***		1.10***	1.10***
		(0.01)	(0.01)		(0.02)	(0.02)		(0.01)	(0.01)
Wealth quintile (Ref. 1st)									
2nd		1.20	1.26		1.00	1.09		0.85	0.90
		(0.23)	(0.24)		(0.25)	(0.26)		(0.21)	(0.23)
3rd		1.09	1.15		0.95	1.05		0.64**	0.69*
		(0.11)	(0.12)		(0.13)	(0.16)		(0.10)	(0.11)
4th		0.91	0.98		0.49**	0.56*		0.69	0.75
		(0.09)	(0.11)		(0.12)	(0.13)		(0.18)	(0.20)
5th		0.84	0.92		0.52***	0.60**		0.69	0.76
		(0.11)	(0.12)		(0.08)	(0.09)		(0.21)	(0.24)
Married/in partnership (Ref. Single)		0.89	0.91		1.12	1.16		0.64	0.65
		(0.10)	(0.11)		(0.22)	(0.22)		(0.15)	(0.15)
Currently working (Ref. Not working)		0.75	0.76		0.66	0.67		0.71	0.72
		(0.12)	(0.12)		(0.18)	(0.19)		(0.15)	(0.15)
N. of chronic conditions		1.24***	1.25***		1.24***	1.24***		1.08*	1.08*
		(0.03)	(0.03)		(0.01)	(0.02)		(0.03)	(0.03)
Verbal Fluency score		0.99*	0.99		0.96**	0.97*		0.95***	0.96**
		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)
EURO-D scale of depression		1.16***	1.15***		1.19**	1.18**		1.10	1.09
		(0.04)	(0.04)		(0.07)	(0.07)		(0.06)	(0.06)

Table 1 – *Effect of SN and AP on the risk of pre-frailty, frailty, and death. Woman (continued).*

	Women								
	Pre-frail			***Frail***			***Deceased***		
SN Scale	0.88***	0.92*	0.94	0.85*	0.94	0.99	0.82	0.96	1.02
	(0.03)	(0.03)	(0.03)	(0.06)	(0.08)	(0.07)	(0.10)	(0.12)	(0.13)
AP Scale			0.87***			0.74***			0.67***
			(0.03)			(0.04)			(0.06)
Age (Cont.)		1.07***	1.07***		1.09***	1.09***		1.10***	1.10***
		(0.01)	(0.01)		(0.01)	(0.01)		(0.01)	(0.01)
Wealth quintile (Ref. 1st)									
2nd		0.90	0.92		0.74	0.77		0.62	0.66
		(0.06)	(0.06)		(0.12)	(0.12)		(0.23)	(0.24)
3rd		0.82**	0.85*		0.63*	0.67		0.83	0.89
		(0.06)	(0.06)		(0.13)	(0.14)		(0.33)	(0.35)
4th		0.76***	0.80***		0.40***	0.45***		0.89	1.01
		(0.06)	(0.05)		(0.06)	(0.08)		(0.32)	(0.36)
5th		0.67***	0.71***		0.40**	0.45**		0.83	0.96
		(0.06)	(0.06)		(0.12)	(0.14)		(0.31)	(0.36)
Married/in partnership (Ref. Single)		1.04	1.02		1.18**	1.13*		0.74	0.69
		(0.07)	(0.06)		(0.07)	(0.06)		(0.16)	(0.15)
Currently working (Ref. Not working)		0.92	0.92		0.45***	0.47**		0.80	0.81
		(0.10)	(0.10)		(0.11)	(0.11)		(0.19)	(0.19)
N. of chronic conditions		1.20***	1.21***		1.40***	1.41***		0.92	0.93
		(0.03)	(0.03)		(0.09)	(0.09)		(0.09)	(0.09)
Verbal Fluency score		0.98***	0.99**		0.93***	0.95***		0.96	0.97
		(0.00)	(0.00)		(0.01)	(0.01)		(0.03)	(0.03)
EURO-D scale of depression		1.17***	1.16***		1.21***	1.20***		1.12	1.11
		(0.03)	(0.03)		(0.05)	(0.05)		(0.08)	(0.08)

Source: Authors' elaborations on SHARE data. *Notes:* Results are reported in the form of Relative Risk Ratios (RRR). 95% C.I. *** $p<0.001$, ** $p<0.01$, * $p<0.05$

When it comes to the role played by health-related characteristics, our findings suggest that aspects like the number of chronic conditions of the individual and depression were significantly associated with higher-risks of pre-frailty and frailty, both among men and women. Similarly, verbal fluency was slightly negatively associated with both of such outcomes.

As for the role of socioeconomic characteristics, wealth appeared to be negatively associated with the risk of becoming frail, particularly for the 4th and 5th quintile. Importantly, while we found no difference between men who are working and those who are not, women who are still in the labor market were found to be substantially less likely to become frail. This aspect might depend on the fact that while working status possibly reflects available income (thus related to a monetary flow), our wealth variable is, rather, a "stock" indicator. Thus, our result might be driven by the fact that while among men the stock aspect might be a relevant determinant of frailty,

among women, rather, the presence or absence of an available income can operate as a determinant of frailty even at initial stages. Surprisingly, moreover, women in partnership were found to be more exposed to the risk of becoming frail, while this association was not significant among men.

Taken together, these results suggest that, among men, having strong social network resources is important to prevent the development of pre-frailty, less so among women. Rather, consistently with previous literature on the matter, participation to various kinds of activities appears to be a notably powerful protective aspect, negatively affecting the risk of becoming pre-frail, frail, as well as the risk of death, both among men and women.

4. Conclusions

The goal of this study was to investigate and compare the role of social connectedness and activity participation on frailty among community-dwelling individuals from 13 European countries, aged 60 and older. To measure social connectedness and activity participation we relied on two indexes employed in Litwin & Stoeckel (2016). Frailty was measured through Romero-Ortuno & Kenny's (2012) SHARE-FI. In our models, we controlled for several relevant demographic and health-related confounders.

Our findings reveal for both social connectedness and activity participation to be associated with frailty. However, comparing the role played by the two aspects, activity participation emerged as a more relevant determinant, strongly associated with the risk of moving from a state of non-frail to one of pre-frail and frail. Engaging in various kinds of activities, compared with maintaining relatively frequent interpersonal relationship, might benefit individuals to a greater extent, thanks to the "doing" component that it involves: the sense of purpose and self-efficacy provided by the involvement in community/political organizations, charity events or educational courses might be a more relevant protective factor for physical and mental health, thus contributing to buffer the negative effects of the cessation of key life roles of paid workers and family caretaker of elderly individuals. This is consistent with research comparing the effect of social connectedness and activity participation on frailty (Chen et al., 2014). Similar findings, with regards to cognitive functioning, are reported by Litwin & Stoeckel (2016) and Miceli et al., (2019).

Nevertheless, differences between man and women should be underlined. Indeed, in this study social connectedness matters mainly among men, with regards to the risk of moving from a state of non-frail to one of pre-frail. Moreover, activity participation was found to be especially relevant in protecting both genders from the risk of becoming pre-frail and frail, as well as from the risk of death.

Naturally, this study is not without limitations. First, social network is a dynamic concept that is subject to change as people age (Schwartz et Litwin 2017). Thus, a more in-depth analysis of these changes would be appropriate. Secondly, due to the lack of time-dependent variables in our data, our analytical strategy allows us to control for the substantial changes occurring in elderly individual's social network's composition and structure only partially. In addition, our measure of activity participation includes various kinds of activities, *i.e.*, "collective" ones such as participation to community/political and sport clubs as well as charity organizations, and more "solitary" ones, such as playing words or number games and reading books or newspapers. Hence, our measure incorporates together widely different types of activities, requiring different levels of social engagement and interaction with others. For this reason, future research should investigate whether there exists any difference in the role played by different kinds of activities, to assess whether it is the "social doing" entailed in the more interactive activities that protects against the risk of frailty, or whether the "solitary" activities, too, matter in preventing individuals from becoming frail.

Besides such limitations, our study has also various strengths. Among these, it is the first one, to our knowledge, exploring and comparing the role played by social connectedness and activity participation in protecting elderly Europeans against the risk of becoming pre-frail and frail. Understanding these aspects can inform about potentially important policy measures to ensure a healthy ageing, such as the implementation of programs aiming at promoting the involvement of the elderly in various kinds of activities.

Acknowledgements

This research was funded by the 2020 PRIN-Research Project of National Relevance "SOcial and health Frailty as determinants of Inequality in Aging (SOFIA)".

We acknowledge co-funding from European Union – Next Generation EU, in the context of The National Recovery and Resilience Plan, Investment Partenariato Esteso PE8 "Conseguenze e sfide dell'invecchiamento", Project Age-It (Ageing Well in an Ageing Society).

This paper uses data from SHARE Waves 4 and 5. The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-

COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, and VS 2020/0313. Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

ANDERSON, N. D., DAMIANAKIS, T., KRÖGER, E., WAGNER, L. M., DAWSON, D. R., BINNS, M. A., BERNSTEIN, S., CASPI, E., COOK, S. L., & The Bravo Team. 2014. The benefits associated with volunteering among seniors: A critical review and recommendations for future research. *Psychological Bulletin*, Vol. 140, No.6, pp. 1505-1533.

BERKMAN, L. F. 1988. The changing and heterogeneous nature of aging and longevity: A social and biomedical perspective. *Annual Review of Gerontology & Geriatrics*, Vol. 8, No.1, pp. 37-68.

BERKMAN, L. F. 1995. The Role of Social Relations in Health Promotion. *Psychosomatic Medicine*, Vol. 57, No. 3, pp. 245-254.

BERKMAN, L. F., GLASS, T., BRISSETTE, I., & SEEMAN, T. E. 2000. From social integration to health: Durkheim in the new millennium. *Social Science & Medicine*, Vol.51, No.6, pp. 843-857.

BESSA, B., RIBEIRO, O., & COELHO, T. 2018. Assessing the social dimension of frailty in old age: A systematic review. *Archives of Gerontology and Geriatrics*, Vol. 78, pp. 101-113.

BLAKE, H., MO, P., MALIK, S., & THOMAS, S. 2009. How effective are physical activity interventions for alleviating depressive symptoms in older people? A systematic review. *Clinical Rehabilitation*, Vol.23, No.10, pp. 873-887.

BÖRSCH-SUPAN, A. 2022a. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w4.800

BÖRSCH-SUPAN, A. 2022b. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 8.0.0. SHARE-ERIC. Data set. DOI: 10.6103/SHARE.w5.800

CHEN, L.-J., CHEN, C.-Y., LUE, B.-H., TSENG, M.-Y., & WU, S.-C. 2014. Prevalence and Associated Factors of Frailty Among Elderly People in Taiwan. *International Journal of Gerontology*, Vol.8, No.3, pp. 114-119.

CHOI, N. G., & KIM, J. 2011. The effect of time volunteering and charitable donations in later life on psychological wellbeing. *Ageing and Society*, Vol.31, No.4, pp. 590-610.

ENG, P. M., RIMM, E. B., FITZMAURICE, G., & KAWACHI, I. 2002. Social ties and change in social ties in relation to subsequent total and cause-specific mortality and coronary heart disease incidence in men. *American Journal of Epidemiology,* Vol.155, No.8, pp. 700-709.

ERIKSSON, B. G., HESSLER, R. M., SUNDH, V., & STEEN, B. 1999. Cross-cultural analysis of longevity among Swedish and American elders: The role of social networks in the Gothenburg and Missouri longitudinal studies compared. *Archives of Gerontology and Geriatrics*, Vol. 28, No.2, pp. 131-148.

ETMAN, A., KAMPHUIS, C. B. M., VAN DER CAMMEN, T. J. M., BURDORF, A., & VAN LENTHE, F. J. 2015. Do lifestyle, health and social participation mediate educational inequalities in frailty worsening? *The European Journal of Public Health,* Vol. 25, No.2, pp. 345-350.

FRIED, L. P., TANGEN, C. M., WALSTON, J., NEWMAN, A. B., HIRSCH, C., GOTTDIENER, J., SEEMAN, T., TRACY, R., KOP, W. J., BURKE, G., MCBURNIE, M. A., & Cardiovascular Health Study Collaborative Research Group. 2001. Frailty in older adults: Evidence for a phenotype. *The Journals of Gerontology*. Series A, Biological Sciences and Medical Sciences, Vol.56, No.3, pp.146-156.

GARCÌA, E. LÓPEZ., BANEGAS, J. R., PÉREZ-REGADERA, A. GRACIANI., CABRERA, R. HERRUZO., & RODRÌGUEZ-ARTALEJO, F. 2005. Social network and health-related quality of life in older adults: A population-based study in Spain. *Quality of Life Research*, Vol.14, No.2, pp. 511-520.

GOBBENS, R. J. J., LUIJKX, K. G., & WIJNEN-SPONSELEE, M. T. 2010. Towards an integral conceptual model of frailty. *The Journal of Nutrition*, Vol.14, No.3., pp. 175-181

HASKI-LEVENTHAL, D. 2009. Elderly Volunteering and Well-Being: A Cross-European Comparison Based on SHARE Data. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 204, 388-404.

JENKIN, C. R., EIME, R. M., WESTERBEEK, H., & VAN UFFELEN, J. G. Z. 2018. Sport for Adults Aged 50+ Years: Participation Benefits and Barriers. *Journal of Aging and Physical Activity*, Vol.26, No.3, pp. 363-371.

LEI, Y., LAO, J., & LIU, J. 2022. Participation in community seniors' organizations and mental health among retired adults in urban China: The mediating role of interpersonal needs. *Frontiers in Public Health*, Vol.10, pp. 1-13.

LITWIN, H., & STOECKEL, K. J. 2016. Social Network, Activity Participation, and Cognition: A Complex Relationship. *Research on Aging*, Vol.38, No.1, pp. 76-97.

LIU, L., GOU, Z., & ZUO, J. 2016. Social support mediates loneliness and depression in elderly people. *Journal of Health Psychology*, Vol.21, No.5, pp. 750-758.

MAKIZAKO, H., SHIMADA, H., DOI, T., TSUTSUMIMOTO, K., HOTTA, R., NAKAKUBO, S., MAKINO, K., & Lee, S. 2018. Social Frailty Leads to the Development of Physical Frailty among Physically Non-Frail Adults: A Four-Year Follow-Up Longitudinal Cohort Study. *International Journal of Environmental Research and Public Health*, Vol.15, No.3, pp. 490.

MICELI, S., MANISCALCO, L., & MATRANGA, D. 2019. Social networks and social activities promote cognitive functioning in both concurrent and prospective time: Evidence from the SHARE survey. *European Journal of Ageing*, Vol.16, No.2, pp.145-154.

MISENER, K., DOHERTY, A., & HAMM-KERWIN, S. 2010. Learning From the Experiences of Older Adult Volunteers in Sport: A Serious Leisure Perspective. *Journal of Leisure Research*, Vol.42, No.2, pp. 267-289.

MORROW-HOWELL, N., HINTERLONG, J., ROZARIO, P. A., & TANG, F. 2003. Effects of Volunteering on the Well-Being of Older Adults. *The Journals of Gerontology*: Series B, Vol.58B, No. 3, pp. 137-145.

MUNFORD, L. A., SIDAWAY, M., BLAKEMORE, A., SUTTON, M., & BOWER, P. 2017. Associations of participation in community assets with health-related quality of life and healthcare usage: A cross-sectional study of older people in the community. *BMJ Open*, Vol.7, No. 2, pp. 1-9

MURRAY, R. P., JOHNSTON, J. J., DOLCE, J. J., LEE, W. W., & O'HARA, P. 1995. Social support for smoking cessation and abstinence: The lung health study. *Addictive Behaviors*, Vol.20, No.2, pp. 159-170.

ROMERO-ORTUNO, R., & KENNY, R. A. 2012. The frailty index in Europeans: Association with age and mortality. *Age and Ageing*, Vol.41, No. 5, pp. 684-689.

SCHWARTZ, E., & LITWIN, H. 2019. The Reciprocal Relationship Between Social Connectedness and Mental Health Among Older European Adults: A SHARE-Based Analysis. *The Journals of Gerontology: Series B*, *74*(4), 694–702.

SEEMAN, T. E., BRUCE, M. L., & MCAVAY, G. J. 1996. Social network characteristics and onset of ADL disability: MacArthur studies of successful aging. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, Vol.51, No.4, pp.191-200.

SUN, J., KONG, X., LI, H., CHEN, J., YAO, Q., LI, H., ZHOU, F., & HU, H. 2022. Does social participation decrease the risk of frailty? Impacts of diversity in frequency and types of social participation on frailty in middle-aged and older populations. *BMC Geriatrics*, Vol.22, No.1, pp. 553-564.

TAKATORI, K., & MATSUMOTO, D. 2023. Effects of social activity participation and trust in the community on the transition of frailty classification in late-stage older adults: A 4-year prospective cohort study. *BMJ Open*, Vol.13, No.5, pp.1-10.

THOITS, P. A. 2011. Mechanisms Linking Social Ties and Support to Physical and Mental Health. *Journal of Health and Social Behavior*, Vol.52, No.2, pp. 145-161.

TREIBER, F. A., BARANOWSKI, T., BRADEN, D. S., STRONG, W. B., LEVY, M., & KNOX, W. 1991. Social support for exercise: Relationship to physical activity in young adults. *Preventive Medicine*, Vol. 20, No.6, pp. 737-750.

VANCE, D. E., WADLEY, V. G., BALL, K. K., ROENKER, D. L., & RIZZO, M. (2005). The Effects of Physical Activity and Sedentary Behavior on Cognitive Health in Older Adults. *Journal of Aging and Physical Activity*, *13*(3), 294–313.

WATTS, P., WEBB, E., & NETUVELI, G. 2017. The role of sports clubs in helping older people to stay active and prevent frailty: A longitudinal mediation analysis. *International Journal of Behavioral Nutrition and Physical Activity*, Vol.14, No.1, pp. 95.

WENGER, G. C. 1997. Social networks and the prediction of elderly people at risk. *Aging & Mental Health*, Vol.1, No.4, pp. 311-320.

XIE, B., & MA, C. 2021. Effect of social participation on the development of physical frailty: Do type, frequency and diversity matter? *Maturitas*, Vol.151, pp. 48-54

Rossella MIGLIO, University of Bologna, rossella.miglio@unibo.it
Chiara PUGLISI, University of Bologna, chiara.puglisi5@unibo.it
Rosella RETTAROLI, University of Bologna, rosella.rettaroli@unibo.it
Giulia ROLI, University of Bologna, g.roli@unibo.it
Francesco SCALONE, University of Bologna, francesco.scalone@unibo.it

COMPOSITE INDICATOR FOR MEASURING THE DEPOPULATION OF ITALIAN MUNICIPALITIES BY THE PERMANENT POPULATION CENSUS RESULTS[1]

Valeria Quondamstefano, Mariangela Verrascina

Abstract. The study aims to measure, through synthetic indices, the depopulation of Italian municipalities from demo-social indicators calculated on the basis of the results of the first 4 editions of the Permanent Population Census.

Since 2018, the Census has become Permanent, meaning it is conducted every year on a sample of municipalities and households and whose output, referring to the entire population, is disseminated annually. The new census strategy is based on the integration of data available in administrative sources and collected in the field. Outputs covering population structure by gender, age and citizenship (Italian-Foreigner/Stateless) and educational attainment are considered for the analysis, for the 2018-2021 period. The analysis focuses on a small group of municipalities, those belonging to the peripheral and ultra-peripheral Inner Areas, to highlight critical issues in these areas. Some indicators are produced on 1,756 Italian municipalities. The composite indicator Adjusted Mazziotta Pareto Index (AMPI) is constructed to summarize multidimensionality in space and time and allow comparisons on depopulation levels.

In addition, groups of municipalities with similar characteristics emerged through the calculation of the CHAID (Chisquared Automatic Interaction Detector) classification tree. The AMPI is the dependent variable, while the independent variables are administrative subdivisions, some geographic characteristics, and the municipality demographic size.

Finally, by representing on a map the municipalities of the best and worst nodes over the 4 years, it is observed that among the municipalities in the most peripheral Areas, the phenomenon evolves negatively quickly. Critical situations are also noted in the North, particularly in the Northeast: municipalities, even in the wealthy areas of the country, are at a risk of depopulation as much as those in the South and Islands.

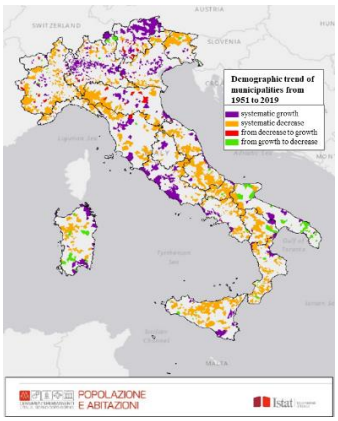
[1] The article is exclusively expressing the authors' opinions. Sections are attributed as follows: paragraphs 4 to Valeria Quondamstefano, paragraphs 1, 2 and 3 to Mariangela Verrascina. Paragraphs 5 and 6 are written jointly by the authors and can not be divisible.

1. Introduction

Depopulation is defined as a gradual decrease in the population of a place.

In Italy, many territories have experienced population loss in recent decades. A cartographic representation of the demographic trend between the 1951 Population Census and the 2019 Permanent Census (Figure 1) shows how many municipalities/territories have suffered a systematic decrease, from census to census, since the first post-war one (in orange in the map).

Figure 1 – *Demographic trend of municipalities from 1951 to 2019.*



Source: https://gis.censimentopopolazione.istat.it/apps/dashboard

There are many causes that determined and still determine the reduction in the size of municipalities, linked mainly to demographic components, but not only. The phenomenon mainly involves small territories. The most appropriate way to study the phenomenon is to use the information at the finest administrative territorial level; so, we consider outputs available per municipality.

The Permanent Population Census comes to our aid in this; the Census had traditionally been carried out on the entire population until 2011, but since 2018 a paradigm shift has been applied. Since then, the Census is conducted annually on a sample of municipalities and households, with results referring to the entire population. So, while until 2011 very fine spatial level data (down to the sub-municipal level) was available every 10 years, with the new strategy a reduced set of data is available every year for all municipalities.

The objective of this work is to measure the depopulation of Italian municipalities starting from socio-demographic indicators computable with the results of the first four editions (2018, 2019, 2020 and 2021) of the Permanent Census of Population and Housing. The focus is on a small group of municipalities, the most marginalised ones, those included in some Inner Areas (Peripheral and Ultra-peripheral Areas).

Starting from the disseminated data, a number of indicators are calculated and then synthesised into a composite Indicator. Through a further Classification Tree analysis we highlight differences and similarities between municipalities in Peripheral and Ultra-peripheral Areas in the census years considered, also to identify where the municipalities most at risk of depopulation are and what structural/demographic characteristics they have.

2. Context

Depopulation derives from a combination of factors, from the reduction of births to emigration, mainly of young people who leave their place of origin for centres that may offer more opportunities for education and work, and who are unlikely to return (return migration). Studies have shown that there is a link between ageing and depopulation: the ageing process of the population has contributed to the subsequent depopulation, and depopulation has emphasised the ageing of the population, creating a vicious circle. (Reynaud and Miccoli, 2018)

Depopulation, population ageing, emigration and low birth rate lead to a different population distribution in the Italian territories and a modified age structure of the population in the different Italian areas. Populations that remain in the places of origin have an altered age structure. These phenomena affect more rural, already penalized and marginalized areas.

The Inner Areas[2] are the country's most peripheral territories in terms of access to basic public services. They are more than 20 minutes' travel time away from a pole that plays the role of a centre of supply of fundamental services relating to education, mobility and health care. (Dipartimento per lo sviluppo e la coesione economica, 2014).

The municipalities of Areas 4 (Intermediate), 5 (Peripheral), 6 (Ultra-peripheral) are 4,055 out of the total of 7,903[3] municipalities existing in Italy in 2021: they comprise more than half of the municipalities, cover about 60.0% of the entire surface area of the national territory and have more than a fifth of the total population (12,690,730 on 59,030,133 i.e. 21.5%).

For each area, the number of municipalities that lost population is calculated; depopulation is considered here as negative numerical change in the population of municipalities, continuous since the beginning of the Permanent Census (Table 1).

[2] The definition of Inner Areas dates back to 2014, when the Department for Development and Economic Cohesion (DPS), now the Agency for Territorial Cohesion, launched a "National Strategy for Inner Areas" (SNAI) as part of the strategic options for the programming of EU funds 2014-2020.

[3] The municipality of Misiliscemi born from Trapani in 2021 is not considered, the population was counted together with Trapani.

Table 1 – *Number of municipalities and municipalities with a negative trend for Inner Area.*

Area	Municipalities		Municipalities with Negative trend	
	AV	%	AV	%
1-Poles	217	2.7	100	46.1
2-Intermunicipal poles	122	1.5	49	40.2
3-Belt areas	3,509	44.4	981	28.0
4-Intermediate areas	2,288	29.0	952	41.6
5-Peripheral areas	1,475	18.7	734	49.8
6-Ultra-peripheral areas	292	3.7	171	58.6
Total	**7,903**	**100.0**	**2,987**	**37.8**
Areas 5 and 6	***1,767***	***22.4***	***905***	***51.2***

Analysing the phenomenon area by area, it can be seen that the municipalities in Area 3 have the smallest negative trend, fewer municipalities have steadily lost population since 2018 (28.0%, 981 municipalities), while the municipalities that have experienced population loss are in Area 5 half (49.8%, 734 municipalities), the share reaches 58.6% in Area 6 (171 out of 292 municipalities). Taking Peripheral and Ultra-peripheral areas together, more than half of the municipalities lose population from year to year. Only 51 municipalities in Areas 5 and 6 show a positive trend (overall in Italy 453).

To better understand the evolution of depopulation and aging in Italian territories (in absolute and relative terms) we consider the percentage of population over 65 years old and the aging index (Table 2).

Table 2 – *Percentage of population over 65 years old and the aging index for Italy, Areas 5 and 6 and municipalities with negative trend – Years 2018, 2019, 2020, 2021.*

	Nr Municipalities	% Population over 65 years old				Aging index			
		2018	2019	2020	2021	2018	2019	2020	2021
Italy	7,903	22.9	23.2	23.5	23.8	174.0	179.3	182.6	187.6
Areas 5 and 6	1,767	24.4	24.9	25.4	25.7	204.3	211.5	216.4	222.4
Areas 5 and 6 - Negative trend	905	25.6	26.1	26.7	26.9	223.9	232.2	238.2	245.4

Thus, if we consider the share of the population over 65 years old we can see how between 2018 and 2021 the value increases slowly but inexorably from 22.9% to 23.8%, an increase of one point over the four-year period. If we go on to consider only the 1,767 municipalities belonging to the peripheral and ultra-peripheral areas, we go from 24.4% to 25.7%. Reducing the number of municipalities to those that have recorded a negative trend among the most peripheral municipalities, there is an increase of 1.3% but the share of over-65s increases further, 25.6% in 2018 and will be close to 27.0% in 2021. It should be noted, confirming the increase in the share of the elderly in recent years, that at the 2011 Census the over-65s were one fifth of the population (20.8%) while in 2001 they were 18.7%.

While the percentage of the population over 65 expresses the weight of the elderly in absolute terms in the total population, an index that represents the weight of the elderly in relative terms is the Ageing index. It is a relative index because it relates the elderly population to the population of the younger age groups. It is an indicator that synthetically represents the degree of by taking into account a part of the age structure of the population. It is a synthetic index of ageing of the population but also a dynamic one, which exalts the effect of ageing because there is usually a concomitant increase in the number of elderly people and a decrease in the number of young people, so the numerator and denominator change in opposite directions.

At the 2001 Census, there were 131.4 over-65s for every 100 young people, in 2011 this figure rose to 148.7. With the start of the Permanent Census, there is a further increase in the value from 174.0 in 2018 to 187.6 in 2021. Analysing the value for the more peripheral municipalities, the 200 threshold is exceeded, if in 2018 it was 204.3 in 2021 it reaches 222.4. An extremely critical situation is observed for the municipalities in Areas 5 and 6 with the negative trend (905 municipalities that have lost population steadily over the last four years): the values are 224 elderly people for every 100 young people and almost 2.5 over 65 years for every young person under 15 years.

These numbers confirm, if any were needed, the very close link between the ageing of the population and the depopulation of the territories. And they also confirm how in the most peripheral municipalities, those that are experiencing depopulation and hence ageing, the age structure of the population is altered, with a significant weight of the elderly. This is why we decided to focus the analysis on a small group of municipalities, those belonging to the Peripheral and Ultra-peripheral Areas, in order to highlight the critical aspects of these areas.

3. Data and variables

An interpretation of depopulation at the local level is carried out using a few simple and non-redundant indicators characterising the population remaining in the municipality. We also consider the resident foreigners because the phenomenon of immigration has characterised the demographic change from the mid-1990s to the present and in some territories of the country has reduced the intensity of population ageing. We have calculated some indicators representing the demographic and socio-economic characteristics of the population residing in the municipalities:

A. *Proportion of foreign population (per thousand persons)* calculated as the ratio of foreign population to total resident population per thousand inhabitants;
B. *Proportion of population aged 0-17 years (percentage)* calculated as the ratio of the population aged under 18 to the total resident population per hundred inhabitants;

C. *Proportion of foreign population aged 0-17 years (percentage)* calculated as the ratio of the foreign population aged less than 18 years out of the total resident foreign population per hundred inhabitants (how many young foreigners out of the total foreign population);
D. *Young age dependency ratio (percentage)* calculated as the ratio of population under 15 to the working age population (15-64) per hundred inhabitants (how many young people out of the working population);
E. *Old age dependency ratio (percentage)* calculated as the ratio of the population over 65 years of age to the population of working age (15-64) per hundred inhabitants (the burden of the elderly on the working population);
F. *Labour force turnover ratio (percentage)* (revised) calculated as the ratio of the 65-69 year-old population over the 20-24 year-old population per hundred inhabitants (the population about to leave the labour market over the population about to enter it[4]);
G. *Percentage of population with diploma of upper secondary education* calculated as the ratio of population with diploma of upper secondary education or vocational qualification over the population aged 9 years and over per hundred inhabitants;
H. *Percentage of population with master's degree or second level academic diploma and Research Doctorate (PhD)* calculated as the ratio of population with bachelor's degree, master's degree and PhD over the population aged 9 years and more per hundred inhabitants.

The indicators described are calculated for the 4 editions of the Permanent Census of Population and Housing (Istat, 2022; Istat, 2021; Istat, 2020).

The territory taken into account refers to the municipalities existing in 2021: 1,756[5] municipalities belonging to Inner Areas Peripheral and Ultra-peripheral.

4. Methodology

4.1. Composite Indicator

In recent years, the measurement of multidimensional phenomena has become increasingly relevant in the scientific community. Socio-economic phenomena cannot be measured by a single descriptive indicator but should be represented with multiple dimensions.

[4] Compared to the standard indicator that considers the population 60-64 years old and 15-19 years old, the indicator has been revised in light of the changes and lengthening of the age of exit from the labour market and the lengthening of the age of entry.
[5] 11 municipalities were excluded because they did not have values at some denominators, therefore some indicators could not be calculated.

"A composite index is a mathematical combination (or aggregation as it is termed) of a set of single indicators (or variables) that represent the different components of a multidimensional phenomenon to be measured (e.g., development, well-being or quality of life). Therefore, the composite indices are used for measuring concepts that cannot be captured by a single indicator" (Mazziotta and Pareto, 2018).

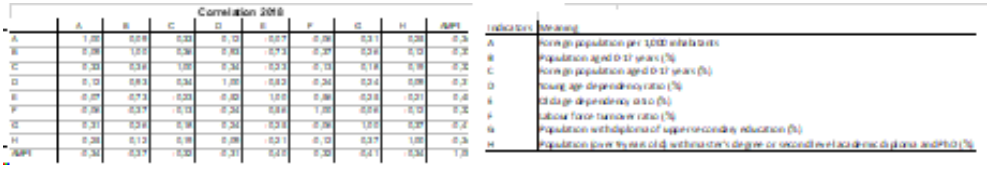
The individual indicators used are summarised through the Adjusted Mazziotta Pareto Index (AMPI). It is a partially non-compensatory composite indicator based on a standardization of the individual indicators, at the reference time, that makes the indicators independent from the unit of measure (De Muro et al., 2011). By choosing this composite indicator, comparisons in both space and time can be ensured. The individual indicators are converted into a scale with a range in the open interval (70;130), where the base value is that of Italy in the first year of analysis (2018).

4.2. Descriptive data analysis AMPI

For the analysis, it is very important to identify the individual indicators that most influence the AMPI and the resulting ranking of the municipalities. In our case, the most influential indicators over the years in the Inner Areas (i.e. the one that is added or removed shifts on average most positions) are for the first 3 years the 'Population aged 0-17 years' (134, 130, 130, 125 in 2018, 2019, 2020 and 2021 respectively) and for the last year 2021 the 'Population (over 9 years old) with master's degree or second level academic diploma and PhD' (119, 118, 129, 135). In contrast, the least influential indicator is the 'labour turnover ratio (revised)' (33, 34, 34, 34).

Having calculated the AMPI, we proceeded with the correlation analysis. It can certainly be said that there is no correlation between the composite indicator and the individual indicators. In fact, the highest value of the Pearson index is found for the indicator 'Population with diploma of upper secondary education' in all years of analysis (-0.407; -0.424; -0.429; -0.415). Figure 2 shows correlation matrix for 2018.

Figure 2 – *Correlation matrix (Areas 5 and 6) – Year 2018*

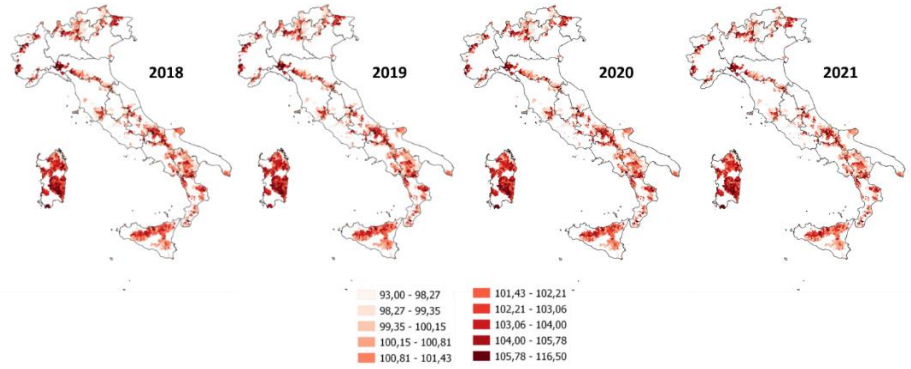


5. Results

5.1. AMPI

Figure 3 shows the mapping of Italian municipalities belonging to Areas 5 and 6 for the 4 years considered according to the level of depopulation.

Figure 3 – *AMPI ranking: Maps of Italian municipalities (Areas 5 and 6).*



Municipalities with the highest values of AMPI are shown in dark red, while those with the lowest values are shown in light red. Peripheral municipalities are located along the mountain ranges: Alps and Apennines. The municipalities at high risk of depopulation are concentrated: in northern Italy, especially on the borders of Piedmont, in some municipalities in Lombardy, in the internal areas of Sardinia, between Basilicata and Calabria, between Molise and Abruzzo, and in a small area called Four Provinces, on the border between four regions[6] that represent the first northern stretch of the Apennines.

5.2. Classification tree

To find out the reasons for this ranking, the analysis continues with the Chisquared Automatic Interaction Detector (CHAID) 'regression tree' classification method (Kass, 1980). We hypothesised that certain contextual factors of administrative, geographical and spatial nature may influence depopulation on a local scale and consequently the demographic structure of the remaining population. Close relationships have been found between some contextual factors of a geographical and geomorphological nature and the depopulation of local realities (Reynaud et al., 2020).

[6] Four Provinces refers to a portion of the Ligurian Apennines included in the Italian provinces of: Alessandria, in Piedmont, Genoa, in Liguria, Pavia, in Lombardy, Piacenza, in Emilia-Romagna.

The AMPI is the dependent variable, while the independent variables are the administrative subdivisions (Geographical area, Region, Province/Metropolitan city), certain geographical characteristics (Altitude zone[7], Population density[8]) and the Demographic size of municipalities class. Since the target variable (the AMPI) is continuous, a F test is used as the criterion of splitting nodes, instead of a chi-squared test (Ritschard, 2013).

The CHAID produces a hierarchical segmentation: disjointed groups (the nodes of the tree) are created, which are internally homogeneous by level of AMPI. In this approach through the dependent variable - the AMPI index - homogeneous groups are constructed within them according to the AMPI values; with the geographical-territorial indicators the municipalities are progressively segmented and 'characterised'. This permits the groups to be read through a distinctive combination of categories of these discriminating variables. All independent variables are influential in the classification, although differently from year to year: in 2018 altitude zone and population density are not influential, in 2019 geographical area, in 2020 and 2021 population density. Variables that are not influential for each year are indicated in red.

For each Census year, we analyse the characteristics of the municipalities belonging to the first two and last two nodes, named respectively 'best nodes' and 'worst nodes'. (Figure 4). The municipalities in the best nodes are medium-large municipalities (between 2,001 and 100,000 inhabitants), mainly in the North-West, North-East and Central Areas. A few municipalities in the South (Campania, Abruzzo, Molise, Basilicata and Calabria) also appear, but there are no municipalities from the Islands. For the years 2018, 2019, 2020 149 municipalities fall into the best nodes, they are 248 for the 2021 Census edition.

In the worst nodes we find small and very small municipalities (up to 1,000 inhabitants). Among the geographical areas, the Centre does not appear. The Regions present are, for the North-West: Liguria, Piedmont, Lombardy; for the North-East: Friuli-Venezia Giulia, Veneto, Emilia-Romagna; for the South: Basilicata and Molise; for the Islands: Sardinia. The number of municipalities in the worst nodes are 109 in 2018, 121 in 2019 and 183 in 2020 and 2021.

Finally, depicting on a map the municipalities of the best and worst nodes over the four-year period, we observe that among municipalities in the most peripheral areas, the phenomenon evolves negatively very quickly (Figure 5**)**. Critical situations are also found in the North, whose municipalities are at risk of depopulation as much as those in the South and the Islands.

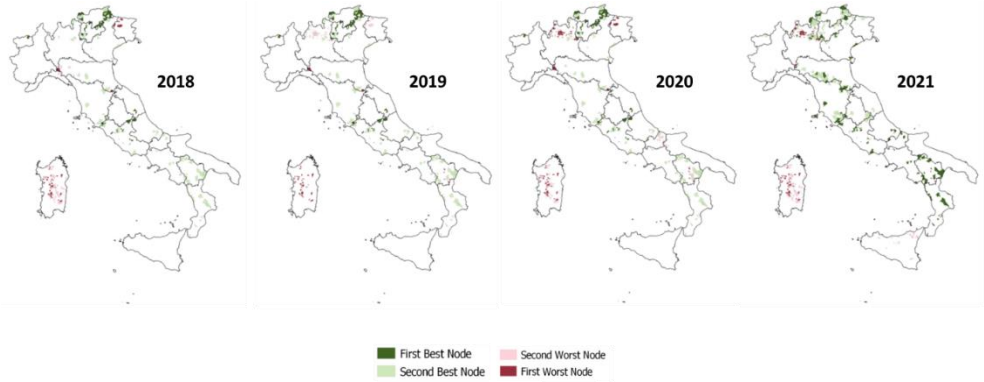
[7] *Altitude zone* derives from the division of the national territory into homogeneous zones resulting from the aggregation of contiguous municipalities on the basis of altimetric threshold values. (Eurostat, 2019)

[8] Relation between the number of inhabitants and the surface of the territory (number of inhabitants per km^2). (Eurostat, 2019)

Figure 4 – *Characteristics of municipalities in the best and worst nodes.*

	2018 Edition	2019 Edition	2020 Edition	2021 Edition
First Best Node	NODE 18	NODE 18	NODE 22	NODE 15
	1-Demographic size of municipalities class: between 2,001 and 5,000 persons	1-Demographic size of municipalities class: between 2,001 and 5,000 persons	1-Demographic size of municipalities class: between 2,001 and 5,000 persons	1-Demographic size of municipalities class: between 5,001 and 100,000 persons
	2-Region: Valle d'Aosta, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Marche, Lazio	2-Region: Valle d'Aosta, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Marche, Lazio	2-Geographical area: North-West, North-East, Centre	2-Region: Lombardy, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Veneto, Emilia-Romagna, Tuscany, Lazio, Campania, Abruzzo, Molise, Basilicata, Calabria
			3-Region: Valle d'Aosta, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Marche, Lazio	
	Number of municipalities: 50	Number of municipalities: 50	Number of municipalities: 50	Number of municipalities: 99
Second Best Node	NODE 16	NODE 16	NODE 15	NODE 16
	1-Demographic size of municipalities class: between 5,001 and 100,000 persons	1-Demographic size of municipalities class: between 5,001 and 100,000 persons	1-Demographic size of municipalities class: between 5,001 and 100,000 persons	1-Demographic size of municipalities class: between 2,001 and 5,000 persons
	2-Region: Lombardy, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Veneto, Emilia-Romagna, Tuscany, Lazio, Campania, Abruzzo, Molise, Basilicata, Calabria	2-Region: Lombardy, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Veneto, Emilia-Romagna, Tuscany, Lazio, Campania, Abruzzo, Molise, Basilicata, Calabria	2-Region: Lombardy, Autonomous Province of Bolzano/Bozen, Autonomous Province of Trento, Veneto, Emilia-Romagna, Tuscany, Lazio, Campania, Abruzzo, Molise, Basilicata, Calabria	2-Geographical area: North-West, North-East, Centre
	Number of municipalities: 99	Number of municipalities: 99	Number of municipalities: 99	Number of municipalities: 149
	Number of municipalities: 149	**Number of municipalities: 149**	**Number of municipalities: 149**	**Number of municipalities: 248**
Second Worst Node	NODE 27	NODE 7	NODE 20	NODE 10
	1-Demographic size of municipalities class: between 501 and 1,000 persons	1-Demographic size of municipalities class: up to 500 persons	1-Demographic size of municipalities class: between 501 and 1,000 persons	1-Demographic size of municipalities class: between 501 and 1,000 persons
	2-Region: Liguria, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Molise, Puglia, Sardinia	2-Region: Lombardy, Friuli-Venezia Giulia	2-Region: Liguria, Lombardy, Veneto, Friuli-Venezia Giulia, Emilia-Romagna, Molise, Puglia, Basilicata, Sardinia	2-Geographical area: Islands
	3-Province: Genova, Piacenza, Parma, Sassari, Nuoro, Oristano, Sud Sardegna		3-Altitude zone: Hill	
	Number of municipalities: 53	Number of municipalities: 69	Number of municipalities: 62	Number of municipalities: 74
FirstWorst Node	NODE 23	NODE 8	NODE 7	NODE 20
	1-Demographic size of municipalities class: up to 500 persons	1-Demographic size of municipalities class: up to 500 persons	1-Demographic size of municipalities class: up to 500 persons	1-Demographic size of municipalities class: up to 500 persons
	2-Region: Piedmont, Liguria, Lombardy, Friuli-Venezia Giulia, Emilia-Romagna, Campania, Basilicata, Sardinia	2-Region: Emilia-Romagna, Basilicata, Sardinia	2-Region: Lombardy, Friuli-Venezia Giulia, Emilia-Romagna, Basilicata, Sardinia	2-Geographical area: North-West, South, Islands
	3-Geographical area: North-East, Islands			3-Region: Lombardy, Basilicata, Sardinia
	Number of municipalities: 56	Number of municipalities: 52	Number of municipalities: 121	Number of municipalities: 109
	Number of municipalities: 109	**Number of municipalities: 121**	**Number of municipalities: 183**	**Number of municipalities: 183**

Figure 5 – *Maps of municipalities (Areas 5 and 6) in the best and worst nodes.*



The focus selected municipalities that are already disadvantaged at the start since they are identified on the basis of their belonging to the most peripheral Inner Areas, i.e. those furthest from the centres offering essential services. If when analysing the Italian territory as a whole the North-South dichotomy usually appears, when attention is focused on a small group of municipalities, peculiarities emerge that deviate from the most widespread stereotypes. But municipalities belonging to the two worst nodes have the same demographic and socio-economic characteristics: the non-arrival of foreigners and the exodus of young people from depopulation areas mean that the (less educated) elderly are left alone, far from the centres offering services (especially hospitals), to deal with age-related problems. In the four years,

several municipalities in Sardinia are constantly present and critical situations of municipalities more at risk of depopulation in several southern Italian regions, but also in wealthy areas in the North.

6. Conclusions

The low birth rate and significant emigration have as a clear consequence a demographic vulnerability due to an ageing population and an abandonment of the territory. These phenomena are even more striking in some local contexts that are already penalised and marginalised. Depopulation of the territory, particularly when combined with an ageing population, represents a major challenge for society, the economy and politics in the future. It will be necessary to identify situations of demographic distress and implement appropriate anti-depopulation measures.

The analysis described with the available data from the four editions of the Permanent Census is carried out to show the potential of using the annual data and how the yearly municipal dissemination allows a temporal and spatial analysis. In fact, thanks to both longitudinal and territorial analyses, it will be possible to highlight trends and/or criticisms in specific territories and identify sub-populations or areas that are particularly vulnerable or in difficulty, i.e. elements that can support public decision-makers at local level in planning action policies.

With this study, it is not possible to identify a trend because 2021 shows slightly different results from the previous three years, but it does make it possible to identify the territories with situations to be monitored. The different trend relative to the last available census year will be further investigated, the worst nodes after the second will be analysed to identify the position in the ranking occupied by the municipalities that are in the last positions in the first three census years. But the result in 2021 could also be an effect of the pandemic, and only the analysis of trends in the coming years will allow this to be investigated.

References

DE MURO P., MAZZIOTTA M., etc. 2011. *Composite Indices of Development and Poverty: An Application to MDGs.* Social Indicators Research 104: pp. 1-18.

DIPARTIMENTO PER LO SVILUPPO E LA COESIONE ECONOMICA, 2014, "Le aree interne: di quale territori parliamo? Nota esplicativa sul metodo di classificazione delle aree". Punto 5 della Nota del Ministro: Ma di quale territorio parliamo (agenziacoesione.gov.it)

EUROSTAT, 2019. *Methodological manual on territorial typologies, 2018 edition*. Luxembourg: Publications Office of the European Union.

ISTAT, 2022. *Popolazione residente e dinamica demografica. Anno 2021*. https://www.istat.it/it/files//2022/12/CENSIMENTO-E-DINAMICA-DEMOGRAFICA-2021.pdf
https://www.istat.it/it/files//2022/12/Nota-metodologica-censipop-_2021.pdf
ISTAT, 2021. *Popolazione residente e dinamica demografica. Anno 2020.* https://www.istat.it/it/files/2021/12/CENSIMENTO-E-DINAMICA-DEMOGRAFICA-2020.pdf https://www.istat.it/it/files/2021/12/NOTA-TECNICA-CENSIMENTO-POPOLAZIONE_2020.pdf
ISTAT, 2020. *Il Censimento permanente della popolazione e delle abitazioni*. *Prima diffusione dei dati definitivi 2018 e 2019.*
https://www.istat.it/it/files/2020/12/REPORT_CENSIPOP_2020.pdf
https://www.istat.it/it/files/2020/12/NOTA-TECNICA-CENSIPOP.pdf
ISTAT, 2017. Rapporto annuale 2017. La situazione del Paese. Roma: Istituto Nazionale di Statistica.
https://www.istat.it/it/files//2017/05/RapportoAnnuale2017.pdf
KASS, G.V. 1980. An Exploratory Technique for Investigating Large Quantities of Categorical Data. Journal of The Royal Statistical Society Series C-applied Statistics, Vol. 29, pp. 119-127.
RITSCHARD G., 2018. CHAID and Earlier Supervised Tree Methods. In J.J. McArdle & G. Ritschard (eds), Contemporary Issues in Exploratory Data Mining in Behavioral Sciences, Routeledge, New York, pp 48-74
MAZZIOTTA M., PARETO A., 2018. *Measuring well-being over time: The adjusted Mazziotta–Pareto index versus other non-compensatory indices.* Social Indicators Research 136, pp. 967-976.
REYNAUD C., MICCOLI S., etc. 2020. *Unravelling a demographic 'Mosaic': Spatial patterns and contextual factors of depopulation in Italian Municipalities, 1981–2011*. Ecological Indicators, 115, 106356.
https://doi.org/10.1016/j.ecolind.2020.106356
REYNAUD C., MICCOLI S., 2018. *Depopulation and the Aging Population: The Relationship in Italian Municipalities.* Sustainability. 10, 1004.
https://doi.org/10.3390/su10041004.

Valeria QUONDAMSTEFANO, Istat, quondamstefano@istat.it
Mariangela VERRASCINA, Istat, verrasci@istat.it

FERTILITY AND UNEMPLOYMENT IN TIMES OF CRISIS IN ITALY

Thaís García-Pereiro, Roberta Pace, Raffaella Patimo, Roberta Misuraca

Abstract. Previous studies found a pro-cyclical relationship between economic downturns and fertility in Western countries, while less is known about the Great Recession's consequences on Italy's fertility levels. Using macro-data from different ISTAT sources, from 2006 to 2018, this study investigates how changes in the economic context, specifically unemployment, have influenced fertility at the regional level. The study also explores whether and how this relationship has changed. Our findings reveal a pro-cyclical relationship between fertility levels and unemployment during the period under observation, more sensitive after 2013 and less homogeneous across macro-areas. A deeper understanding of these dynamics can inform policy interventions aimed at supporting fertility and addressing demographic behaviours in times of uncertainty.

1. Introduction

The Great Recession started in the autumn of 2007 in the USA and rapidly spread to most European Countries, causing worsening economic conditions and increasing unemployment in the period 2008-2013. The effects of the Great Recession were felt across various domains of the economy, including employment, financial markets, and consumer spending. During this period, several countries experienced a decline in economic growth, as well as a rise in unemployment rates and financial instability. This crisis profoundly impacted individuals and households, leading to job losses, reduced income levels, and financial hardships. Its effects were not limited to economic aspects but also had social and demographic consequences. The adverse economic conditions relating to downturns led to increased stress and a general perception of uncertainty about the future that induced individuals to postpone or reduce their plans for having children (Comolli, 2021). The strongest fertility declines were registered in Southern Europe, especially in countries such as Italy and Spain (Matysiak et al., 2020), given the existent and persistent precarious labor market conditions brought about the recent reform of labor market flexibilization (Cirillo et al. 2017).

The Italian Great Recession had the unique characteristic of occurring at a time when fertility rates were in the process of recovering from a period of significant decline (Goldstein et al., 2013; Comolli, 2017; De Rose and Strozza, 2015). Meanwhile, the mean age of women giving birth to their first child increased, further constraining reproductive choices. Moreover, despite the prevalence of strong family ties and traditional values, Italy has long shown very-low fertility levels.

The central aim of this article is to investigate the association procyclical or countercyclical between changes in economic conditions, specifically the deterioration of employment rates, and fertility rates before, during and after the Great Recession in Italy. A macro-panel dataset was constructed, encompassing annual data for all 20 Italian regions, spanning from 2006 to 2018, to explore the impact of Total Unemployment Rates on the Total Fertility Rates. Subsequently, we delve into variations in the fertility response to unemployment over time (before and after 2013) and between the macro areas in Italy (North, Center, South).

2. Theoretical background, literature and research hypotheses

The question of whether and how business cycles affect fertility has been widely debated over the years (Sobotka et al., 2011; Goldstein et al., 2013). While there is an abundance of studies on the effects of economic downturns on births in the United States and some European countries, less is known about Great Recession's consequences on Italy's fertility trends.

All things considered, fertility depends on a wide range of individual and institutional circumstances, including sex, age, and social status as well as welfare and family policies (Vignoli et al., 2012), but it also differs across countries and contexts (Alderotti et al., 2022).

Childbearing is the result of choices made by individuals and strictly depends on personal preferences, family situation, and income constraints evaluation, in the present and for the future. The general perception of uncertainty about economic conditions plays a crucial role in childbearing decision-making, especially during periods of economic decline (Ayllón, 2019). If the individual's future economic expectations are positive, the number of children may increase; conversely, in the case of economic downturns (job loss or job instability), uncertainty about the future may induce individuals to avoid long-term commitments and postpone childbearing decisions. Income constraints and employment insecurity, at the individual and family level, act as signals of possible future uncertainty, making the decision to have children less attractive, leading young adults to prolong their education in order to improve their labour market position (Kohler et al., 2002; Matysiak et al., 2021). This may be the result of an income effect, related to the greater difficulty of childbearing during an economic downturn (Bellido and Marcèn, 2019). On the contrary, some couples could take advantage of unemployment spells or instability

employment periods to decide to have children, exploiting this low wage window characterized by lower opportunity costs of parental leave and childcare, especially for mothers, who generally represent the main childcare providers. In this case, a substitution effect arises as childcare costs decrease when women are out of the labour market (Butz and Ward 1979; Friedman et al., 1994; Alderotti et al. 2019). Childbearing decision-making is generally made by the couple, therefore, the labour market status of both is potentially able to affect future fertility intentions (Alderotti et al., 2022; Comolli et al., 2019).

As mentioned, the effects of the recent economic decline on fertility have been widely studied. Schneider and Hasting (2015) investigated how the worsening of economic conditions that occurred during the Great Recession impacted state-level fertility in the US. The authors found a negative relationship between the detrimental economic conditions and fertility decisions of low-educated women. Along this line, Seltzer (2019), combining statistical and survey US data, examines the long-term effect on fertility in post-recession periods. In particular, the author considers the structural changes in the industry composition (job displacement and deindustrialization) as playing a primary role in the decrease of TFR, with a particular focus on the Great Recession period. Finally, Comolli (2017) further develops the approach in Goldstein et al. (2013) and finds that fertility seems to negatively respond to the economic uncertainty during the Great Recession. In particular, the author points out the central role played by the deterioration of the labour market structure in determining a decline in fertility rates.

Goldstein et al. (2013), using European data from Eurostat and the OECD database, investigated how changes in labour market conditions, through an increase in the unemployment rates, affected fertility rates during the Great Recession. They found a strong negative correlation between the decline in fertility rates and the economic recession, especially for those European states hardest hit by the crisis, albeit in the presence of strong variation by region, age, and parity. Following the Goldstein et al. approach, Comolli (2017) confirms the negative association between a rise in unemployment, as a consequence of the Great Recession in European countries, and the decline in fertility rates, in particular among young women. An interesting work in this direction is Bellido and Marcén (2019). The authors investigated the business cycle impact on fertility for 30 European countries over three decades and found that, although the business cycle effect on fertility seems to be negative but moderate, different scenarios are identified, strictly related to socioeconomic and institutional factors at the country level, such as the generosity of welfare or high levels of gender equality. Similar results were found by Ayllón (2019).

More recent European studies confirmed a pro-cyclical relationship between employment instability and fertility (Matysiak et al., 2020). Specifically, relying on

different measures, such as the unemployment rates (total unemployment and youth unemployment) and the proportion of long-term unemployment (unemployment lasting 12 months or more), they found a strong relationship between increased unemployment rates and fertility decline in Italy. The findings indicated that deteriorating economic conditions during the recession were associated with a more pronounced decline in fertility compared to the pre-recession period. Furthermore, well-established literature confirms the presence of gender-specific responses: the effect of the economic crisis on fertility strongly depends on men's economic and labour conditions (Vitali and Billari, 2017).

The strength and the nature of the crisis effects tend to be different across countries (Alderotti et al., 2022). The high degree of heterogeneity of social and family welfare coverage, as well as regional labour market conditions, strongly moderate the overall effect of economic downturns on fertility decisions (Hiilamo, 2017). In Southern Europe, especially in countries such as Italy and Spain, structural weaknesses and labour market instability worsened the effects of the economic downturn on fertility decisions. Focusing on Italy, very few studies investigated the link between the Great Recession and fertility decisions from a macro perspective. An exception is Comolli and Vignoli (2021) who aimed to investigate the causal effect of the economic crisis on birth rates, because of decreasing levels of confidence about the future and increasing perceived uncertainty. They documented a procyclicality of the relationship between fertility and economic downturns in Italy.

We draw on the literature strand that shows that there is a negative relationship between unemployment and fertility (Sobotka et al., 2011; Goldstein et al., 2013; Cherlin et al., 2013; Seltzer, 2019). Considering changes over time, some studies found that the Great Recession has had a negative impact on fertility, particularly in terms of first births (Adserà and Menendez, 2011; Neels et al., 2013). Especially in Italy and Spain, the effects of the crisis on fertility levels were intensified by labor market instability, after having already experienced a substantial decline in the past. We, therefore, hypothesize a pro-cyclical relationship between fertility and unemployment (H1).

The studies conducted by Goldstein et al. (2013) and Zambon et al. (2020) provide evidence of a decrease in fertility rates during periods (and especially in Southern European countries) characterized by pronounced labour market instability. We, therefore, expect to find a more sensitive fertility response to unemployment after 2013 (H2).

Finally, we test the previous hypotheses by grouping regions into macro-areas. A well-differentiated geographical fertility trend has been identified in Italy. The upturn in the TFR occurred mainly in Northern regions, where most of the increase has been attributed to the contribution to the fertility of foreign women (Caltabiano

et al., 2009). Along this line, Zambon et al. (2020) investigated the TFR at the regional level in Italy, between 1952 and 1998. They found that fertility levels in Italy are strongly associated with economic changes, also confirming different patterns across the Italian territory. Fertility rates in the economically advanced regions of Northern Italy appeared less sensitive to the worsening of economic conditions but more responsive to economic growth, if compared to more disadvantaged regions (Southern Italy). We, therefore, expect to find important differences in fertility responses to unemployment between Italian macro-areas (H3).

3. Data and Methods

We aim to examine the impact of Total Unemployment Rates on the Total Fertility Rates over a period that covers before, during, and after the occurrence of the Great Recession using macro data at the regional level for Italy. To do this, from 2006 to 2018 a macro-panel dataset was constructed, encompassing annual data for all 20 regions. Our dependent variables are TFRs, which were drawn from the Population Roster; while our main variables of interest are TURs, coming from the Labour Force Survey. The mean values of these indicators for the period under examination are displayed in Table 1.

We can observe variations in both TUR and TFR across different regions in Italy. The variations in both the Rates across different regions in Italy suggest an inverse relationship between unemployment and fertility rates. Regions with higher TUR values, such as Puglia, Calabria, and Sardegna tend to have lower TFR values, indicating a negative correlation between unemployment and fertility. This pattern implies that higher levels of unemployment may be associated with lower fertility rates in the Italian regions. Conversely, regions with lower TUR values, such as Trentino-Alto Adige, and Veneto, tend to have higher TFR values, suggesting a positive correlation between lower unemployment rates and higher fertility rates.

Additionally, we explore whether and how the relevance of unemployment to fertility has changed over the observed period. The period trend is represented as a categorical variable, divided into three categories: 2006-2008, 2009-2013, and 2014-2018. Furthermore, we also investigate possible variations in the observed relationship across macro-areas by grouping regions (Northern, Central, and Southern regions).

Following the mainstream literature (Goldstein et al., 2013; Hiilamo, 2017; Seltzer, 2019; Matysiak et al., 2020) a one-year lag has been incorporated into empirical analyses. All our model specifications also include the following control variables: the percentage of the population with a university education (men and women), mean age at marriage (men and women), marriage and divorce rates, the percentage of foreign women and childcare coverage (Bellido and Marcén, 2019; Comolli, 2017).

Table 1 – *Mean values of Total Unemployment Rate (TUR) and Total Fertility Rate (TFR) for Italian regions, 2006-2018.*

Region	TUR	TFR
Sicilia	16,52	1,39
Calabria	15,81	1,29
Campania	15,03	1,40
Sardegna	13,98	1,11
Puglia	13,94	1,29
Basilicata	11,03	1,18
Molise	10,19	1,17
Lazio	8,67	1,38
Abruzzo	7,92	1,29
Valle d'Aosta	6,18	1,50
Piemonte	5,97	1,39
Marche	5,87	1,35
Liguria	5,84	1,35
Umbria	5,41	1,34
Toscana	5,30	1,35
Lombardia	4,50	1,32
Friuli-Venezia Giulia	4,50	1,43
Emilia-Romagna	4,14	1,45
Veneto	4,04	1,61
Trentino-Alto Adige	2,95	1,47

Source: Own elaboration, ISTAT data.

To test our hypotheses, we run two groups of Ordinary Least Squares (OLS) regression models. To account for the influence of unobserved variables that change over time, calendar-year fixed effects are utilized. This approach helps to control for time-varying and region-varying factors that might affect the relationship between unemployment rates and fertility rates (Goldstein et al., 2013; Seltzer, 2019). To address potential bias resulting from correlation within and between regions, we applied robust standard errors clustered by regions (Kohler and Kreuter, 2005).

The first group of models (Table 1) is aimed at identifying the pro-cyclical or counter-cyclical nature of the relationship between TFRs and TURs and changes occurred over time during the period. The first three model specifications are aimed at testing our first research hypothesis (H1) while comparing results across a null model (M1), a model with fixed calendar year and region effects (M2) and a fixed effects model that considers the influence of control variables (M3). To test for our second research hypothesis (H2) models also include interaction terms between period trends and unemployment measures to search for differences in fertility responses over time (M4).

The second group of models (Table 2) follows the same structure of the former, but divided into macro-areas (North, Center and South), to approximate geographical differences in the effects of unemployment on fertility and test for our third and last research hypothesis (H3).

4. Results

To test for the presence of a negative relationship between fertility and unemployment (H1), we estimated OLS models without and with fixed effects (FE). Table 2 (M1-M3) reports a pro-cyclical relationship between fertility and unemployment levels during the examined period, which aligns with findings from both Italian and international literature. In M1, as the TUR increases by one-percentage-point, the respective TFR decreases by 0.0128. This negative association is also found when including region and calendar year fixed effects and robust standard errors clustered by regions (M2), but the strength of the relationship decreases. As stated before, this finding is consistent with most studies conducted in US, Europe, and Italy (Sobotka et al., 2011; Goldstein et al.,2013; Cherlin et al., 2013; Cazzola et al.,2016; Seltzer, 2019; Zambon et al.,2020). It is interesting to note that the negative effect of TURs on TFRs diminishes when adding control variables to model estimations (from M2 b = -0118 to M3: b = -0.0077).

Table 2 – *Estimates of the OLS regression analysis of the Total Unemployment Rate (TUR) on the Total Fertility Rate (TFR), 2006-2018, without and with region and calendar fixed effects.*

	M1	**M2**	**M3**	**M4**
TUR	-0.0128***	-0.0118***	-0.00768***	-0.00394
	(0.00138)	(0.00213)	(0.00254)	(0.00282)
*Period trend*TUR*				
2009-2013*TUR				-5.15e-05
				(0.00130)
2014-2018*TUR				-0.00181*
				(0.00140)
Constant	1.477***	1.466***	2.719***	2.535***
	(0.0152)	(0.0210)	(0.221)	(0.218)
Observations	256	256	256	256
R-squared	0.253	0.316	0.655	0.664
Number of id		20	20	20

*Notes: Robust Standard Errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Models 3 and 4 (M3 and M4) control for the percentage of the population with a university education (men and women), mean age at marriage (men and women), marriage and divorce rates, the percentage of foreign women, and childcare coverage*

Table 3 – *Estimates of the OLS regression analysis of the Total Unemployment Rate (TUR) on the Total Fertility Rate (TFR), for Italian macro areas, 2006-2018.*

Macro-areas	**M1**	**M2**	**M3**	**M4**
North				
TUR	-0.0174***	-0.0123**	-0.0186***	-0.0239**
	(0.00447)	(0.00478)	(0.00378)	(0.00864)
*Period_trend*TUR*				
2009-2013*TUR				0.00790
				(0.00594)
2014-2018*TUR				0.00728
				(0.00708)
Constant	1.543***	1.511***	2.811***	2.758***
	(0.0294)	(0.0297)	(0.447)	(0.396)
Observations	100	100	100	100
R-squared	0.134	0.176	0.784	0.792
Number of id		8	8	8
Center				
TUR	-0.0152***	-0.0207***	-0.0132	-0.00428
	(0.00320)	(0.00204)	(0.00662)	(0.00516)
*Period_trend*TUR*				
2009-2013*TUR				0.00570
				(0.00297)
2014-2018*TUR				-0.00108
				(0.00510)
Constant	1.480***	1.525***	3.808**	2.862**
	(0.0273)	(0.0166)	(0.711)	(0.527)
Observations	52	52	52	52
R-squared	0.311	0.501	0.825	0.871
Number of id		4	4	4
South and Island				
TUR	0.00245	-0.00920***	-0.00450*	-0.00444*
	(0.00249)	(0.00237)	(0.00199)	(0.00217)
*Period_trend*TUR*				
2009-2013*TUR				0.000408
				(0.00109)
2014-2018*TUR				0.000287
				(0.00121)
Constant	1.231***	1.397***	2.784***	2.745***
	(0.0369)	(0.0336)	(0.323)	(0.306)
Observations	104	104	104	104
R-squared	0.009	0.472	0.723	0.723
Number of id		8	8	8

*Notes: Robust Standard Errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Models 3 and 4 (M3 and M4) control for percentage of the population with a university education (men and women), mean age at marriage (men and women), marriage and divorce rates, the percentage of foreign women and childcare coverage.*

As for our second hypothesis (H2), we suggested that TFR at the regional level will experience a more sensitive response after 2013**,** due to the increase in unemployment rates. As shown in Table 2, M4 tests for changes on the unemployment-fertility relationship during different phases of the Great Recession after controlling for possible confounders. The interaction term between the last period and unemployment is negative and highly significant, suggesting that between 2014-2018 the negative unemployment-fertility association was stronger that it was during the 2006-2008 pre-recession period (M4: 2014-2018*TUR b = -0.0018).

Table 3 presents the results of OLS regressions for Italian macro areas. In the Northern area, there is a pro-cyclical relationship between unemployment and fertility, although no clear period effects are observed.

5. Conclusions

In this study we examined the association between changes in economic conditions, specifically the deterioration of total employment rates, and its impact on fertility rates in Italy before, during, and after the Great Recession. The study aligned with previous literature that suggests a negative relationship between unemployment and fertility confirming a pro-cyclical relationship between fertility and unemployment (H1).

Additionally, building on the literature that indicates a decrease in fertility rates during periods of labor market instability due to economic downturns (Goldstein, 2013; Zambon et al., 2020), the results supported our second hypothesis (H2), showing a more sensitive response in fertility rates after 2013. Finally, the study recognised differentiated geographical fertility trends among Italian regions because of varying sensitivity to the economic crisis impact, as found in previous literature, therefore confirming important differences in fertility responses to unemployment between Italian macro-areas (H3).

Understanding the relationship between economic conditions and fertility behaviours in Italy can inform policy discussions and interventions to address low fertility rates and promote more women employment's friendly measures. The political debates on population ageing/shrinking and gender equality, both high in the (inter)national agenda, could benefit from these insights. Specifically, initiatives such as flexible working arrangements, affordable and accessible childcare facilities, and parental (not only maternal, so as to share family duties in the couple) leave provisions can help maintain a work-family life balance, encouraging women to start and boost their careers while considering family planning. Regional policies would benefit at a deeper level: regions showing greater sensitivity to economic crisis should receive specific support, such as job creation programs, skill development,

and incentives for family planning helping the reduce the inherent territorial cultural gaps.

Acknowledgments

We acknowledge co-funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE8 – Project Age-It: "Ageing Well in an Ageing Society". This resource was co-financed by the Next Generation EU [DM 1557 11.10.2022]. The views and opinions expressed are only those of the authors and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. Additional co-funding was secured by PON Ricerca e Innovazione 2014-2020 DM 1062 by the project "La promozione della gender parity tra dinamiche sociodemografiche e politiche occupazionali innovative" CUP H95F21001450006.

References

ADSERÀ, A., & MENENDEZ, A. 2011. Fertility changes in Latin America in periods of economic uncertainty. *Population studies*, 65(1), 37-56.

ALDEROTTI, G. 2022. Female employment and first childbirth in Italy: what news?. *Genus*, 78(1), 1-19.

ALDEROTTI, G., MUSSINO, E., & COMOLLI, C. L. 2022. Natives' and migrants' employment uncertainty and childbearing during the great recession: a comparison between Italy and Sweden. *European Societies*, 1-35.

ALDEROTTI, G., VIGNOLI, D., BACCINI, M., & MATYSIAK, A. 2019. Employment instability and fertility in Europe: A meta-analysis. *Demography*, 58(3), 871-900.

AYLLÓN, S. 2019. Job insecurity and fertility in Europe. *Review of Economics of the Household*, 17(4), 1321-1347.

BELLIDO, H., & MARCÉN, M. 2019. Fertility and the business cycle: the European case. *Review of Economics of the Household*, 17(4), 1289-1319.

BUTZ, W. P., & WARD, M. P. 1979. The emergence of countercyclical US fertility. The *American economic review*, 69(3), 318-328.

CALTABIANO, M., CASTIGLIONI, M., & ROSINA, A. 2009. Lowest-low fertility: Signs of a recovery in Italy? *Demographic research*, *21*, 681-718.

CAZZOLA, A., PASQUINI, L., & ANGELI, A. 2016. The relationship between unemployment and fertility in Italy: A time-series analysis. *Demographic Research*, 34, 1-38.

CHERLIN, A., CUMBERWORTH, E., MORGAN, S. P., & WIMER, C. 2013. The effects of the great recession on family structure and fertility. *The Annals of the American Academy of Political and Social Science*, 650(1), 214–231

CIRILLO, V., FANA, M. & GUARASCIO, D. 2017. Labour market reforms in Italy: evaluating the effects of the Jobs Act. *Economia Politica* 34, 211–232. https://doi.org/10.1007/s40888-017-0058-2.

COMOLLI, C. L. 2017. The fertility response to the Great Recession in Europe and the United States: Structural economic conditions and perceived economic uncertainty. *Demographic research*, 36, 1549-1600.

COMOLLI, C. L., NEYER, G., ANDERSSON, G., DOMMERMUTH, L., FALLESEN, P., JALOVAARA, M., LAPPEGÅRD, T. 2019. Beyond the economic gaze: Childbearing during and after recessions in the Nordic countries. *European Journal of Population*, 37(2), 473-520.

COMOLLI, C. L., & VIGNOLI, D. 2021. Spreading uncertainty, shrinking birth rates: A natural experiment for Italy. *European Sociological Review*, 37(4), 555-570.

DE ROSE, A., & STROZZA, S. (Eds.). 2015. *Rapporto sulla popolazione: l'Italia nella crisi economica*. Società editrice il Mulino, Spa.

GOLDSTEIN, J. R., KREYENFELD, M., JASILIONIENE, A., & ÖRSAL, D. K. 2013. Fertility reactions to the "Great Recession" in Europe: Recent evidence from order-specific data. *Demographic research*, 29, 85-104.

FRIEDMAN, D., HECHTER, M., & KANAZAWA, S. 1994. A theory of the value of children. *Demography*, 31, 375–401.

HIILAMO, H. 2017. Fertility response to economic recessions in Finland 1991–2015. *Finnish Yearbook of Population Research*, 52, 15-28.

KOHLER, H. P., BILLARI, F. C., & ORTEGA, J. A. 2002. The emergence of lowest-low fertility in Europe during the 1990s. *Population and development review*, 28(4), 641-680.

KOHLER, U., & KREUTER, F. 2005 *Data analysis using Stata*. Stata press.

MATYSIAK, A., SOBOTKA, T., & VIGNOLI, D. 2020. The Great Recession and fertility in Europe: A sub-national analysis. *European Journal of Population*, 37(1), 29-64.

NEELS, K., THEUNYNCK, Z., & WOOD, J. 2013. Economic recession and first births in Europe: recession-induced postponement and recuperation of fertility in 14 European countries between 1970 and 2005. *International Journal of Public Health*, 58(1), 43–55.

SCHNEIDER, D., & HASTINGS, O. P. 2015. Socioeconomic variation in the effect of economic conditions on marriage and nonmarital fertility in the United States: Evidence from the Great Recession. *Demography*, 52(6), 1893–1915.

SELTZER, N. 2019. Beyond the Great Recession: labor market polarization and ongoing fertility decline in the United States. *Demography, 56,* 463–1493.

SOBOTKA, T., SKIRBEKK, V., & PHILIPOV, D. 2011. Economic recession and fertility in the developed world. *Population and development review*, 37(2), 267-306.

VIGNOLI, D., DREFAHL, S., & DE SANTIS, G. 2012. Whose job instability affects the likelihood of becoming a parent in Italy? A tale of two partners. *Demographic Research*, 26, 41-62.

VITALI, A., & BILLARI, F. C. 2017. Changing determinants of low fertility and diffusion: A spatial analysis for Italy. *Population, Space and Place*, 23(2), 1998.

ZAMBON, I., RONTOS, K., REYNAUD, C., & SALVATI, L. 2020. Toward an unwanted dividend? Fertility decline and the North–South divide in Italy, 1952–2018. *Quality & Quantity*, 54, 169.

Thaís GARCIA-PEREIRO; Università degli studi di Bari; thais.garciapereiro@uniba.it

Roberta PACE, Università degli studi di Bari; roberta.pace@uniba.it

Raffaella PATIMO, Università degli studi di Bari; raffaella.patimo@uniba.it

Roberta MISURACA, Università degli studi di Bari; roberta.misuraca@uniba.it

DISTANCE LEARNING DURING THE PANDEMIC: OPINIONS AND ATTITUDES OF YOUNG STUDENTS[1]

Alessio Buonomo, Cinzia Conti, Francesca Di Patrizio, Salvatore Strozza, Marco Dionisio Terribili

Abstract. During the pandemic, school closures and distance learning were one of the main instruments to contain the spread of the virus. Even if the tool of distance education did not originate with the pandemic, as the virus spread and schools closed, the use of online tools became more and more conscious and widespread. This scenario has amplified the scientific (and non-scientific) debate on the role played by technologies and online education. The aim of the paper is to study the effect of distance learning on the (self-perceived) performances of students in Italy. Studies carried out in other countries about school performances during the pandemic show that the performance seems to have been negatively influenced, and the young were reported to have suffered more. The study exploits the data collected in 2021 by the Italian Institute of Statistics (ISTAT), with the support of the Ministry of Education, through a survey on "Children and Teenagers: behaviours, attitudes and future projects". The use of regression models allows us to understand the connections between personal and familiar characteristics of the students and the perception of the effects of distance learning on their school performance. Our results indicate that foreign students appear to have a greater probability, compared to native, of a negative influence on school results, as do those who feel they belong to a fairly or very poor family.

1. Introduction

Because of the spread of Covid-19, more than a hundred countries around the world have closed schools at all levels of education as a measure to contain contagions and impacted 91.3% of enrolled students worldwide (UNESCO, 2020). This represents the first global school closure since World War II. The first countries that closed schools were those on the Asian continent. Italy was the first country in Europe to close schools (9 March 2020) followed by several other European countries (Sarno, 2020). School closures and distance education were one of the

[1] This work is the result of a close collaboration among the authors.

main instruments to contain the spread of the virus (Schneider and Council, 2021). The tool of distance education, however, did not originate with the pandemic but was already researched, known, and implemented before COVID-19. Of course, as the virus spread and schools closed, the use of online tools became increasingly aware and widespread. This scenario has amplified the scientific (and non-scientific) debate on the role played by technologies and online education (Ranieri *et al.*, 2020).

Distance learning (DL), during the period of the pandemic, has increased social inequalities (UNESCO, 2020). In fact, differing conditions of access to technology also take the form of unequal access to education. In other words, the digital divide is configured as a cultural divide. It is estimated that only 15% of the countries that applied school closures were adequately equipped to allow students to access distance education, therefore, the remaining 85% could not ensure students a properly continued learning during the Covid pandemic (Sarno, 2020). In the Italian case, the National Institute of Statistics (ISTAT) highlighted that 45% of students, aged between 6 and 17, had no access to DL because they lacked devices (e.g., computers) in their homes (Ferraro *et al.*, 2020; Iavarone and Trocchia, 2020). Such students were mostly concentrated in southern Italy where 20% of students had no access to any device and were excluded from DL (Ferraro *et al.*, 2020).

The international literature has also highlighted other characteristics (less direct than those mentioned above) of the impact of DL on students' education. Compared to traditional learning (with the co-presence of teachers and students), DL requires the student to have a high level of motivation and self-discipline to be effective because the student is more exposed to the risk of distraction and dropping out of studies (Clark, 2020). Disengaged or unmotivated students have a higher risk of being penalized in educational performance with distance education (Burke and Dempsey, 2020).

In the case of vulnerable young people, e.g. low-income families, foreigners, single parents, education (and care) in the presence of teachers is particularly important, and being deprived of it puts them at greater risk of low performances and school dropout (Baker, 2020). In addition, belonging to low socio-economic status increases the risk of not having the proper tools (e.g., internet connection, computer, etc.) to effectively take advantage of distance education. During the period of the pandemic, the debate on DL in Italy was very heated and there were several authors who underlined how this increased inequalities (Saraceno, 2021).

2. Objectives and research questions

Contemporary research is increasingly investigating the consequences of distance education and school closures. However, there are still few answers in the literature on the role played by distance education on student learning. There is also a lack of quantitative information about school performance and long-DL. The purpose of this paper is to fill this gap by exploring, from the point of view of children and adolescents, the association between perceived school performance and DL. Two are our main research questions: has distance education played a positive or negative role in (perceived) student performances? What factors most significantly influenced students' perceptions of their own performance during DL? It is important to understand who were the most vulnerable students during the pandemic, also in order to plan specific actions to fill the gaps.

3. Data and methods

In 2021 ISTAT, with the support of the Ministry of Education, carried out a survey on "Children and Teenagers: behaviours, attitudes and future projects". It collected fundamental information for understanding the impact of the spread of the Covid-19 pandemic on the daily lives of young people, to highlight the vulnerabilities and strengths of the new generations. Through a short questionnaire administered exclusively with the CAWI technique (which can also be filled using a smartphone or a tablet), school issues were addressed - with a specific focus on the use and effects of DL-, relationships with schoolmates, friends in general and the family, the use of social media, citizenship, and future projects of the new generations.

The survey involved a random sample of around 41,000 pupils who, in the 2020/2021 school year, attended one of the first and second-grade secondary schools (Lyceum and technical and professional institutes) selected by the Student Registry of the Ministry of Education and distributed throughout the country. Continuing with the previous survey on the integration of second generations, foreign children constituted a specific survey target: 11,000 foreigners were included in the sample. The sampling design is based on two stratified stages, in which the primary sampling units are the schools and the final units are the pupils. The stratification concerns the first stage units (the schools) and therefore indirectly the final units.

To achieve the research objective and to investigate, according to the students' point of view, the association between perceived school performance and DL, purifying the effects of interrelationships with other different dimensions, we resorted to the application of a logistic multinomial model (Agresti, 2012). The

dependent variable - non-binary categorical - is represented by the perception expressed by the student in response to the question "How do you think DL has affected your grades in this school year 2020/2021?", with possible responses "It has affected them positive", "It has affected them negatively", "It hasn't affected them at all". As explanatory variables, we considered socio-demographic characteristics (citizenship, gender, cultural background, perceived economic status and type of family), characteristics related to the student's educational background (type of school and its location geography, opinion on one's school performance before the pandemic, incidence of pupils with foreign citizenship in one's school) and in particular to opinions on the effects of DL (understanding of lessons, degree of difficulty of individual subjects, lack of contact with teachers). Among the regressors, we included also other variables detected during the interview that identify particular aptitudes of the pupil (interest in school, psychological discomfort experienced during the pandemic) and the instrumental and non-instrumental difficulties they encountered (type of internet connection, availability of navigation devices, possibility of following the DL from the beginning of the first lockdown, opinion on the preparation of the school to face a year in DL). We applied the multinomial model by setting the reference value of the dependent variable "The DL had no influence on the grades", giving rise to model 1 "The DL had a positive influence vs she had no influence" and model 2 "DL had a negative influence vs she had no influence". We applied also a stepwise procedure with similar results. In addition, specific models for different types of schools were applied without particular shreds of evidence.

Due to the hierarchical nature of school data (students grouped into sections, in turn, grouped in classes, which belong to a school, and so on) we studied multilevel models. Specifically, being categorical the dependent variable, we proposed the multinomial multilevel logit models (MMLM) in our study (Skrondal and Rabe-Hesketh, 2003). MMLMs are direct extensions of models for binary responses and can incorporate random slopes and contextual effects for nominal responses. By using these models, we tried to investigate how school policies, teachers' practices, or class dimensions affect student outcomes while accounting for the hierarchical structure of the data.

MMLMs have proven to be useful in analysing school data, and the effects of the levels in these models have been found to be significant but exiguous. On the other hand, the interpretation of the results from MMLMs got more difficult, compared to simple multinomial logit models: given these considerations, we decided to continue the analysis without considering random school effects, by using the simple multinomial logit model that will be illustrated in the following sections.

4. Results

4.1 DL in the Italian Schools during the Pandemic: a descriptive analysis

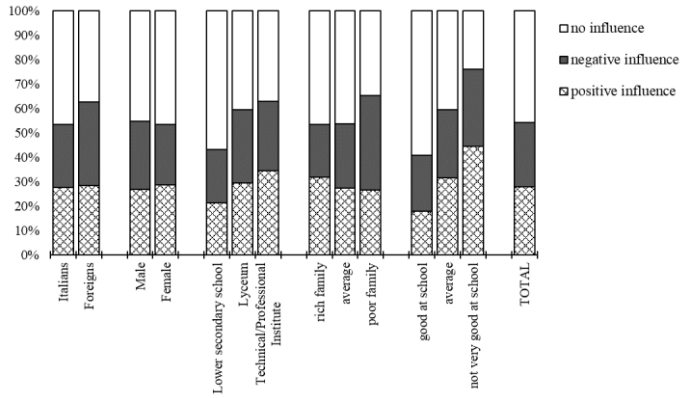
During the pandemic, boys and girls experienced for the first time a totally new way of "going to school" while staying at home. DL was an experience of great impact on the daily life of the very young and for many of them, it lasted over time. The DL has not only imposed a different way of holding and following lessons, but it has also greatly reduced the role of socialization agent that the school has always played, limiting the opportunities for meeting and comparison. This is a generalized in fact 98.7% of students (equal to over 4 million and 220 thousand) have faced periods of DL.

Surely the new generations were not without digital skills, many have defined these young people as "digital natives" because they were habitual users, even before the pandemic, of digital technologies for communication, information, gaming and the use of audio-visuals. However, the "forced" use of DL has introduced not only a change of pace in the use of ICT but also new elements of inequality connected to pre-existing socio-economic and digital gaps.

If it is true that the teenagers were already "very connected", not all of them had the most suitable tools, both in terms of hardware and network connection, to follow many hours of DL. 80% of Italian students were able to follow DL immediately and continuously in the period between March and June 2020; among foreigners, the percentage of those who were able to constantly attend online lessons dropped to 71.4%. DL has not convinced the vast majority of first and second-grade secondary school pupils: 67.7% prefer face-to-face teaching, 20.4% consider the two types of teaching to be the same and only 11.9% prefer DL. A slight difference between the two genders can be noted: it is girls who support face-to-face teaching more (69.5%) than boys (66.1%).

Regarding the influence on the performance of the DL, almost 46% of the interviewees think that it has not influenced the votes, 27.7% believe that it has had a positive influence and 26.4% that it has had a negative influence. Negative influence is reported more often by boys than girls. Middle school students most often said that DL had no influence on grades. Those who believe that the family is quite or very poor believe that in 39% of cases, there has been a negative influence of the DL on grades. Those who were very good at school before the pandemic in almost 60% of cases believe that the DL has not influenced their grades.

Figure 1 – *"How do you think distance learning has influenced your gaps in this 2020/2021 school year?", percentages by citizenship, sex, kind of school, family perceived economic situation, school performances before the pandemic and Total.*



Source: Istat, Survey on "Children and teenagers: behaviours, attitudes and future plans", 2021.

Studies carried out in other countries about school performances during the pandemic show that the performance seems to have been negatively influenced, and young were reported to have suffered more (Panagouli *at al.*, 2021; Engzella *et al.*, 2021; Hammerstein *et al.*, 2021). In Italy, we haven't yet pieces of evidence in this field. So, it could be interesting a deep and multidimensional analysis of the connection between school performances and some individual characteristics. We must underline that in this article the personal perceptions of the students are studied and not the objective achievements or the final marks reported at the end of the school year. The perceived performance aspect seems to be connected to various factors and a multivariate analysis can allow us to better understand the connection between the variables and the perception of the influence on students' performance.

4.2 DL and school performance: results of the regression model

The first result that emerges from the regression models is that, among the factors considered, some appear to have a specific impact in only one of the two models, determining either a negative or a positive influence of the DL on the grades received, others have a general impact on both models. Among the latter, citizenship, the type of school attended, the perception of school performance and the hardships experienced should be considered. Being a foreign citizen, attending an upper secondary school (rather than a middle school), showing interest in school, not having experienced psychological discomfort during the Covid-19 pandemic and

having declared very good school performances before the pandemic determined the same type of impact in both models, albeit with different intensity.

Net of all the other covariates considered, the students' perception of his or her performance at school before the pandemic appears to be, however, the factor with the greatest impact in both models. In the case of the positive influence on grades: having defined oneself as "Not very good" compared to "Very good" determines a more than 4 times greater probability of perceiving a positive influence compared to those who have not detected any influence of the DL on the grades (the probability of having a negative influence on votes, compared to having none, is 84%). Analysing model 1 (positive influence): even the self-perception of an "average" performance before the pandemic determines a double probability of having a positive influence compared to having no influence; the probability is double also in the case of those who attend a upper secondary school (both Lyceum and technical/professional institute) compared to those who attend a lower secondary school. The probability of perceiving a positive influence on grades is higher than 63% among students who declare that they had no difficulty in any subject during remote lessons, about 20% among foreign students and 10% among women. Among those who perceive that their family is neither rich nor poor, compared to those who instead feel they belong to a very rich family, or those have ascertained that the school, on the return of the 2020/2021 school year (after the total block of activities that took place in the spring of 2020), was not ready to face a year of DL, the probability of a positive influence is lower, by 25% and 17% respectively.

Analysing model 2 (negative influence): the perception of a negative impact of DL on grades is particularly accentuated among Lyceum students (the odds ratio is higher than 88%), less strongly it is in the case of students of technical/professional institutes (+48%). Once again, but more intensely, foreigners appear to have a greater probability, compared to native classmates, of a negative influence on school results (+50%), as do those who feel they belong to a fairly or very poor family (+40%). Even having faced the initial emergency with difficulty - not having been able to participate immediately in DL - or among those who complain of connection problems or a school unprepared for the management of DL at the beginning of the school year 2020/2021, the probability of a negative influence is higher (plus 65%, 21% and 21% respectively). On the other hand, the following are protected from a negative impact on grades: women (-24%), students of schools in the Centre and Southern Italy (-23% and 26%), those who show an interest in school (-31%), who did not suffer psychological discomfort (-47%) and who believed that the DL did not make the subjects or the understanding of the lessons more difficult (-42%).

Table 2 – *Logistic multinomial model 1: distance learning has had a positive influence on your grades in school year 2020/2021; reference category "No influence" (1).*

Variables		Od. R.	Confidence interval	
Citizenship (ref. Italian)	Foreign	1.188***	1.113	1.269
Sex (ref. Male)	Female	1.101***	1.044	1.160
Kind of school (ref. Medie)	Lyceum	1.801***	1.685	1.926
	Tech/Professional Inst.	1.991***	1.873	2.117
Geographical area of the school (ref. North)	Center	1.005	0.940	1.075
	South	1.097**	1.032	1.166
Parents' highest educational level (ref. Low (2))	I do not know	0.921	0.794	1.069
	High (3)	0.887*	0.828	0.951
Family perceived economic situation (ref. Rich)	Average	0.750***	0.685	0.820
	Poor	0.822	0.711	0.951
Household type (ref. Both parents)	Single parents or no parents	0.999	0.931	1.071
School performances before the pandemic (ref. Good)	Average	2.301**	2.166	2.446
	Not very good	4.347***	3.836	4.925
Interest in school (ref. Little)	Very much	0.806***	0.757	0.859
Psychological discomfort during the Covid-19 (ref. Yes)	No	0.748***	0.709	0.789
Internet connection at home (ref. Excellent connection)	With problems/No	0.903**	0.857	0.951
PC ownership (ref. Yes)	No	0.907**	0.845	0.973
DL immediately with continuity (ref. Yes)	No/partially	0.947	0.887	1.012
Preparation of the school in the management of DL (ref. Prepared)	Not prepared	0.835***	0.788	0.885
DL lessons more difficult to understand (ref. Yes)	No	1.003	0.952	1.056
DL has made more difficult to study subjects (ref. Yes)	No	1.627***	1.534	1.724
Miss direct contact with teachers (ref. A lot)	Little/No	1.148***	1.084	1.215
Incidence of foreigners at school (ref. No foreigners)	Low	1.023	0.841	1.245
	High	1.063	0.874	1.294

*Note: levels of significance * p<0.1, ** p<0,05, *** p<0.001*
Source: Istat, Survey on "Children and teenagers: behaviours, attitudes and future plans", 2021. (1) Number of observations = 37.523; Pseudo R^2 =0,1815; Likelihood ratio = 6.556 (d.f. 50; p-value <,0001); (2) Includes: untitled, primary, lower secondary; (3) Includes: Upper secondary, Tertiary

Table 3 – *Logistic multinomial model 2: distance learning has had a negative influence on your grades in school year 2020/2021; reference category "No influence" (1).*

Variables		Od. R.	Confidence interval	
Citizenship (ref. Italian)	Foreign	1.504***	1.409	1.606
Sex (ref. Male)	Female	0.761***	0.721	0.803
Kind of school (ref. Medie)	Lyceum	1.876***	1.755	2.007
	Tech/Professional Inst.	1.483**	1.392	1.580
Geographical area of the school (ref. North)	Center	0.767***	0.716	0.822
	South	0.737***	0.692	0.786
Parents' highest educational level (ref. Low (2))	I do not know	0.880	0.758	1.022
	High (3)	0.844**	0.787	0.905
Family perceived economic situation (ref. Rich)	Average	1.082**	0.979	1.197
	Poor	1.400***	1.210	1.619
Household type (ref. Both parents)	Single parents or no parents	1.152***	1.075	1.234
School performances before the pandemic (ref. Good)	Average	1.360	1.282	1.442
	Not very good	1.841***	1.618	2.095
Interest in school (ref. Little)	Very much	0.693***	0.650	0.739
Psychological discomfort during the Covid-19 (ref. Yes)	No	0.527***	0.498	0.558
Internet connection at home (ref. Excellent connection)	With problems/No	1.205***	1.142	1.272
PC ownership (ref. Yes)	No	1.092**	1.019	1.170
DL immediately with continuity (ref. Yes)	No/partially	1.646***	1.549	1.750
Preparation of the school in the management of DL (ref. Prepared)	Not prepared	1.211***	1.144	1.282
DL lessons more difficult to understand (ref. Yes)	No	0.579***	0.549	0.612
DL has made more difficult to study subjects (ref. Yes)	No	0.583***	0.552	0.615
Miss direct contact with teachers (ref. A lot)	Little/No	0.945*	0.891	1.003
Incidence of foreigners at school (ref. No foreigners)	Low	1.098	0.894	1.348
	High	1.028	0.836	1.263

*Note: levels of significance * p<0.1, ** p<0,05, *** p<0.001*
Source: Istat, Survey on "Children and teenagers: behaviours, attitudes and future plans", 2021. (1) Number of observations = 37.523; Pseudo R^2 =0,1815; Likelihood ratio = 6.556 (d.f. 50; p-value <,0001); (2) Includes: untitled, primary, lower secondary; (3) Includes: Upper secondary, Tertiary.

5. Conclusions and next steps

New generations have a greater ability to socialize through digital tools than before. They know how to use social media platforms effectively and create networks useful for learning through the exchange of materials and through mutual help (Hernandez-de-Menendez *et al.*, 2020). However, their increasing ability to use digital tools to interact with peers' results in a decrease in in-person relationships.

If for many students (46%) DL was neutral in terms of performance, for others, it exerted a negative influence and for almost the same percentage it played a positive role. The regression analysis helped to understand better the characteristics of different groups of students. For some categories, the results seem very polarized (for example for the foreigners). In some students' subpopulations DL increases the propensity to have both better and worse scholastic performance at the same time. If these results could seem "counterintuitive", they can be easily interpreted like this: DL is a school modality which leaves nobody indifferent. Someone has loved it, gaining, in terms of school performance. Someone else has not liked it, obtaining during the pandemic period, worse scholastic results.

It can depend on many different aspects that are difficult to include in the regression models. For example, the fact that during the pandemic in Italy exams – and school tests - were simplified and in general the number of failures was reduced. Some students may considered these aspects, while others did not. In general, the perception is not in favour of reduced performance as mostly reported in the literature, but, of course, it can depend on the fact that almost all the considered studies are based on objective information. In addition, not all the studies agree on the effect of DL. Zierer (2021) reported that in the case of secondary education, no significant differences could be found between the learning progress of pupils in face-to-face and those in DL. An aspect that emerges from the regression models and that is in line with reported cases in the international literature is the protection effect played by high-level performances before the pandemic. High-achieving students suffer minor negative effects of DL.

In the next months, it will be important to check the results of the present analysis with the information collected by the Ministry of Education – above all about failures and interruptions of the studies - and the ones collected by INVALSI. In this way, the perception of the students could be compared with objective data about the performances during the pandemic and after.

Acknowledgements

We acknowledge co-funding from Next Generation EU, in the context of the National Recovery and Resilience Plan, Investment PE8 – Project Age-It: "Ageing Well in an Ageing Society". This resource was co-financed by the Next Generation EU [DM 1557 11.10.2022]. Additional co-funding was secured by MIUR in the context of project titled "Immigration, integration, settlement. Italian-Style" [PRIN 2017 - grant No. 2017N9LCSC_004] and by University of Naples Federico II, PON "Research and innovation" 2014-2020 (PON R&I) – ACTION 4. Title "Digital literacy as a determinant of the social inclusion of migrants and their children" [CUP E65F21003040003].

References

AGRESTI A. 2012. *Categorical data analysis*. Vol. 792, John Wiley & Sons.

BAKER J. 2020. The kids who will never return to school after COVID-19. Sydney Morning Herald. Available from https://www.smh.com.au/national/the-kids-who-will-never-return-to-school-after-covid-19-20200411-p54j0e.html.

BURKE J., DEMPSEY, M. 2020. *Covid-19 practice in primary schools in Ireland report*. Maynooth, Ireland.

CHEN E., KACZMAREK K., OHYAMA H. 2021. Student perceptions of distance learning strategies during COVID-19, *Journal of dental education*, 85(Suppl 1), 1190-1191.

CLARK J.T. 2020. Distance Education. In VV.AA. *Clinical Engineering Handbook*, Elsevier: Amsterdam, The Netherlands, pp. 410–415.

ENGZELLA P., FREYA A., VERHAGENA M.D, 2021. Learning loss due to school closures during the COVID-19 pandemic, *PNAS*, Vol. 118, No. 17, Available from https://doi.org/10.1073/pnas.2022376118.

FERRARO F. V., AMBRA F. I., ARUTA L., IAVARONE M. L. 2020. Distance learning in the COVID-19 era: Perceptions in Southern Italy, *Education Sciences*, Vol. 10, No. 12, pp. 355-365.

HAMMERSTEIN S., KÖNIG C., DREISÖRNER T., FREY A. 2021. Effects of COVID-19-Related School Closures on Student Achievement. A Systematic Review, *Frontiers in Psychology*, Vol. 12. Available from https://doi.org/10.3389/fpsyg.2021.746289.

HERNANDEZ-DE-MENENDEZ M., DÍAZ C.A.E., MORALES-MENENDEZ R. 2020. Educational experiences with Generation Z, *I International Journal on Interactive Design and Manufacturing (IJIDeM)*, 14, pp. 847–859. Avaible from https://doi.org/10.1007/s12008-020-00674-9.

IAVARONE M.L., TROCCHIA N. 2020. *Il Coraggio delle cicatrici: storia di mio figlio Arturo e della nostra lotta*, Utet: Milan.

PANAGOULI E., STAVRIDOU A., SAVVIDI C., KOURTI A., PSALTOPOULOU T., SERGENTANIS T.N., TSITSIKA A. 2021. School Performance among Children and Adolescents during COVID-19 Pandemic: A Systematic Review, *Children*, Vol. 8, No. 1134, pp. 1-12. Avaible from https://doi.org/10.3390/children8121134.

PIKULSKI P., PELLA J., CASLINE E., HALE A., DRAKE K., GINSBURG. 2020. School connectedness and child anxiety, *Journal of Psychologists and Counsellors in Schools*, Vol. 30, No. 1, pp. 13–24. Avaible from https://doi:10.1017/jgc.2020.3.

RAJAMOHAN S., BENNETT E.; TEDONE D. 2019. The hazards and benefits of social media use in adolescents, *Nursing*, Vol. 49, No. 11, pp. 52–56. Avaible from https://doi.org/10.1097/01.nurse.0000585908.13109.24.

RANIERI M., GAGGIOLI C., BORGES M. K. 2020. La didattica alla prova del Covid-19 in Italia: uno studio sulla Scuola Primaria, Praxis educativa, Vol. 15, pp.1-20.

SARACENO C. 2021. *La dimensione sociale della crisi covid in Italia, Lavoro ed equità sociale*, Fondazione Friedrich Ebert in Italia.

SARNO E. 2020. Emergenza sanitaria e chiusura di scuole e università. Il divario culturale come ulteriore effetto del Covid-19, *Documenti geografici*, Vol. 1, pp. 219-229.

SCHNEIDER S. L., COUNCIL M. L. 2021. Distance learning in the era of COVID-19, *Archives of Dermatological Research*, Vol. 313, No. 5, pp. 389-390.

SKRONDAL A., RABE-HESKETH S. 2003. Multilevel logistic regression for polytomous data and rankings, *Psychometrika*, Vol. 68, pp. 267-287.

UNESCO 2020. *COVID-19 Educational Disruption and Response*. Available from: https://en.unesco.org/covid19/educationresponse.

ZIERER K. 2021. Effects of Pandemic-Related School Closures on Pupils' Performance and Learning in Selected Countries: A Rapid Review, *Education Sciences*, Vol. 11, 252.

Alessio BUONOMO, University of Naples Federico II, alessio.buonomo@unina.it
Cinzia CONTI, Istat, ciconti@istat.it
Francesca DI PATRIZIO, Istat, dipatriz@istat.it
Salvatore STROZZA, University of Naples Federico II, strozza@unina.it
Marco Dionisio TERRIBILI, Istat, terribili@istat.it

THE NEW MONITORING OF THE DETECTION "ANTHROPIC PRESSURE AND NATURAL RISKS" SUMMARY INDICATORS COMPARED AS TOOLS FOR MEASURING THE DATA COLLECTION OF THE EXTRACTION PHENOMENON[1]

Lucia Mongelli, Sabrina Angiona, Barbara Boninfante, Valentino Parisi, Stefano Moscatelli, Barbara Stobbia

Abstract. The survey "Anthropic pressure and natural risks" on mining activities from quarries and mines aims to update every year the Istat Mining Database on the extraction activities of non-energy mineral resources, natural mineral waters and thermal waters censusing the authorized extraction sites situated in Italy. Istat carries out the survey through two thematic survey models: - Quarries and mines: solid non-energy mineral resources: - Natural mineral waters and thermal waters. For the first time a monitoring system was developed with the processing of the data collected through the GINO++ platform. In the latest edition of the survey, an additional tool was introduced to control the sending of data from all the expected sites, from each institution, through the use of the site identification code. Where the site identification code was not unique, other identifiers based on the combination of other variables collected, were created ad hoc. This innovation had contributed to raise of the survey's response rate of eight percentage points both in relation to the number of respondents and the number of completed questionnaires compared to those expected.

1. Introduction

The Territorial-Environmental Survey "Anthropogenic Pressure and Natural Hazards on Quarrying and Mining Activities", carried out by Istat from 2015, is an annual census about all the activities of exploitation of non-renewable natural resources present in the national territory.

The data are collected directly from the technical offices of the local public Institutions which are responsible for mining (Regions, Provinces, Autonomous Provinces of Trento and Bolzano, Mining Districts of Sicily).

[1]The paper is the result of the common work of the authors. In particular, paragraphs 1 and 5 are attributed to Lucia Mongelli, paragraph 2 is attributed to Barbara Boninfante, paragraph 3 to Sabrina Angiona, paragraph 3.1 and 4 to Valentino Parisi, 4.1. to Stefano Moscatelli, paragraph 4.2 to Barbara Stobbia and for 2.1 Barbara Stobbia e Stefano Moscatelli developed graphs and cartograms.

In recent years, numerous initiatives have been implemented for the consistent maintenance of information flows to ensure the effectiveness and robustness of archives. Particular importance has been given to the improvement of data collection both to increase the amount of spatial data and to improve the flow of information (Vignani, D. 2017).

This work aims to present the important methodological innovations of the monitoring system introduced in the last edition of the survey (Ed. 2022) which allowed the construction of synthesis indicators on the progress of data collection and allowed the integration of a more effective preliminary validation process.

2. A brief overview of data sources

The statistics on the extraction of non-energetic mineral resources from quarries and mines are produced and disseminated with the aim of updating the Istat Mining Database. (Banca dati I.STAT 2013-2021)

The Istat Mining Database contains data and information (historical series years 2013-2018) on the extraction of minerals of category I (mines) and category II (quarries) by mining site and type of mineral resource, according to the classification of mineral substances reported in the current Royal Decree 1443/1927, reference point of the national legislation on mining and quarrying "Legislative rules governing the research and mining of the Kingdom".

Since 2017, the field of observation of the survey has been extended to data on withdrawals of mineral and thermal waters by mining site at a territorial level (ISTAT, ediz.2017-2023). Since 2018, the collection of statistical-geographical data has been strengthened to complete the georeferencing of the Istat geominerary database, with technical offices of local public Institutions (ISPRA, 2020).

Our data sources are the administrative archives held by the Technical Offices which are responsible for mining in the local public Institutions.

The archives contain the following types of data and information:

- ✓ Regional legislation;
- ✓ Regional Plans and Provincial Mining Plans (PRAE, PPAE),
- ✓ Authorizations for the cultivation of mining sites,
- ✓ Regional and provincial databases;
- ✓ Annual declarations submitted by the individual undertakings authorized to cultivate by the granting authorities They contain information about the exercise of their activities for extraction site.

2.1. Survey plan

Local public Institutions represent the survey units, while individual mining sites (quarries and mines) represent the analysis units.

This is a total survey that Istat carries out annually through two thematic survey models for self-completion of the respondent:

1. Quarries and mines: Solid non-energy mineral resources
2. Natural mineral and thermal waters

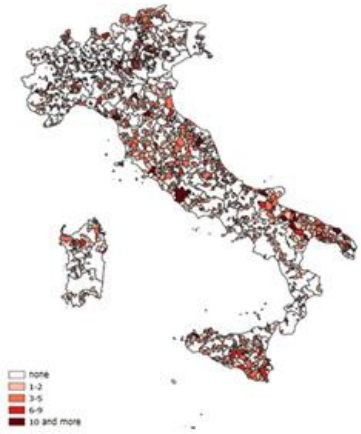
In the following, we can see the geolocation of the sites on the national territory (ISTAT, 2017-2022).

Mining sites (quarries and mines) are mainly concentrated in the North of the country, North-East 943 out of 3,968 total (23.8%) sites and North-West 933 (23.5%), followed by the Centre (21.7% of sites), South Italy (19.2% of sites) and Islands (11.8% of sites). (Vignani, D. 2019)

The natural mineral water and thermal water sites are mainly concentrated in the North-East of the country, 283 sites out of 822 total or 34.4%, followed by Southern Italy, 24.1% of sites, the Centre (21.0% of sites), the North-West (15.2% of sites) and the Islands (5.2%).

Figure 1 – *Quarries and Mines by municipality. Percentage values.*

Figure 2 – *Territorial distribution of sites. 2022 Edition Natural and Thermal Mineral Waters by municipality. Percentage values.*



3. Data collection

To support the data collection process, starting from the year 2020, Istat has introduced important methodological innovations for the detection "Anthropogenic pressure and natural risks" with the aim to improve, standardize the encodings and formats of the variables, and simplify the data collection process.

In particular, the technique CAWI (Computer Assisted Web Interviewing) has been introduced to acquire data in electronic format, through the Gino++ (Gathering information Online) portal of Istat (Torelli, R. 2011).This is according to the provisions of the Code of Digital Administration (d.lgs 82/2005 and subsequent additions and modification) in which data must be transmitted to Istat in computerised mode.

GINO++ is a generalized software that allows not only the collection of data but also the complete management of surveys via web (creation of web questionnaire, online data acquisition and/ or file upload and monitoring) (ISTAT, *Newsstat 4/2012)*. In the Gino++ data acquisition system, the technical offices of the local public Institutions responsible for mining have two alternative ways of sending data: fill in web questionnaires in a guided manner or upload files in format .csv according to Istat records specially prepared.

In the "Documents and Instructions" area of the Gino Portal, the responding public Institutions find the support material to fulfill all the obligations provided by the survey: the facsimile of the web questionnaire of "Quarries and mines: solid non-energy mineral resources" and "Natural and Thermal Mineral Waters", Excel-format detection models, Excel-format track and technical specifications, compilation guide and non-energy mineral resources list by lithotype.

In addition, in Gino Institutions are available two support tools that facilitate and speed up the compilation and provision of data: the possibility to access the data entered in the previous edition and the list of all authorized mining sites that are present in the Istat Database. These two tools make it possible to avoid the provision of redundant data, such as mining sites ceased and communicated in previous years, and to facilitate the identification of missing sites and thus to detect unauthorized, undocumented sites.

In the Gino data acquisition system, rules (soft and hard) have also been implemented to report anomalies such as the use of unforeseen encodings and formats or the inconsistency between variables, which may occur during compilation. These rules contribute to improve the completeness, the quality of the data, the timeliness in providing it and to reduce the statistical annoyance on the respondent.

In Gino++, thanks to the monitoring function, it is possible to constantly monitor the activity of the respondents: the registration, the access by type of user, the data acquisition status. Thanks to the wide range of features present in the software and thanks to the information collected on the identification code of the extractive site, it was possible to implement the new monitoring tool for the investigation "Anthropogenic pressure and natural risks" which will be explained in the following paragraphs

3.1. The pre-validation process

The mining site identification code is one of the main information collected in each questionnaire. This is a mandatory field that may consist of the cadastral code of the site or the code in use by the institutional respondent, if it is present in the administrative archive. The site code is fundamental to identify the extractive site both during the monitoring of data collection and in the subsequent phase of data analysis in historical series.

In particular, during the data collection it allows, on the one hand, to verify the transmission by the Institution of the questionnaire containing the site data related to the year under survey, on the other hand, a precise and faster interlocution with the respondent if there is the presence of inconsistent data than those provided in previous years or missing or incorrect data.

It is therefore important that the site code is present for each survey unit and that: a) does not vary from year to year; b) is unique.

For this reason, at the beginning of the data collection, to Institutions that don't have in their archives the site identification it is recommended to create one for each site and to keep it over time. Despite the provided recommendations, not all Institutions comply with these indications.

In particular, in the 2021-2022 survey edition for about one third of the sites the identification code was not unique or there was a change of the site code from one year to the next which hasn't been accompanied of a code link table. The significant presence of cases of mismatch (ISTAT, 2013) or non-uniqueness of the site codes has not allowed the monitoring of the arrivals of the questionnaires per single site but to compare the data at the aggregate level through the comparison of questionnaires provided and questionnaires expected by the individual institution for province of the site location and type of site (quarry, mine, natural mineral water, thermal water).

The situation has definitely improved in the 2022-2023 survey edition: the percentage of unique or varied site codes compared to the previous year has fallen to 18% (10% for water, 20% for quarries and mines). Part of the lack of correspondence is due to the fact that the data relating to some sites, including the identification code, had not been provided by the institution in the previous year but recovered by the Production Service through other sources. These cases represent 8% of the units to be detected. Excluding such sites, the cases of mismatch of the code are equal to 10% of the sites (9% for water, 11% for quarries and mines).

4. Data collection monitoring tools

The monitoring is carried out weekly in the first months of the survey and twice weekly in the last month by processing the data collected through the GINO platform.

Each institution is required to fill in as many questionnaires as there are the sites of extraction of individual materials (of the territory of competence) in which each authorized extraction company operates. To verify the completeness of the data sent by the institution, the number of sent questionnaires containing the data of the reference year is compared with that of the expected questionnaires that are those communicated by the institution in the previous survey edition plus any other data for which the Production Service has recovered information after the data collection ends.

Considering that each institution must to complete a survey model for each new site, the transmission of a higher number of questionnaires than expected might suggest, in the first place, that the supply is complete and that the more questionnaires in more are referred to new sites. In order to verify that this is the case, the number of questionnaires sent and expected for province of location of the site and type's (natural mineral waters, thermal waters, quarries, mines), in some cases, it can highlight that the sending of data for some territories or for some types of sites is, in reality, incomplete.

To support the monitoring of the data collection trend, in 2022-2023 survey edition, for each expected site, Istat has verified the status of the questionnaire (unanswered, in processing, sent) and the reference year of the reported data through the coupling of the code of the extraction site reported in the questionnaire of the previous survey edition and in the current one, through relational databases (Atzeni P., et al. 2015).

Where there has been a change of site codes compared to the previous year or the institution has reported the same site identifier in several questionnaires, for the purpose of verification, other site identifiers have been defined which have taken into account the combination of the following information collected: site type, matter extracted, province, municipality, locality and geographical coordinates of the site and tax code of the undertaking authorized to extract the matter.

Particular attention was paid to sites for which, in the previous survey edition, the respondent had communicated, the validity of the permit for the extraction of the material by the entrusted company also in the year following the reference data year.

The one-to-one comparison between expected sites and received sites has allowed the Istat monitor to be able to provide the respondent with timely information on the sites for which the requested information has not been provided, with a view to the provision of all data by the end date of collection.

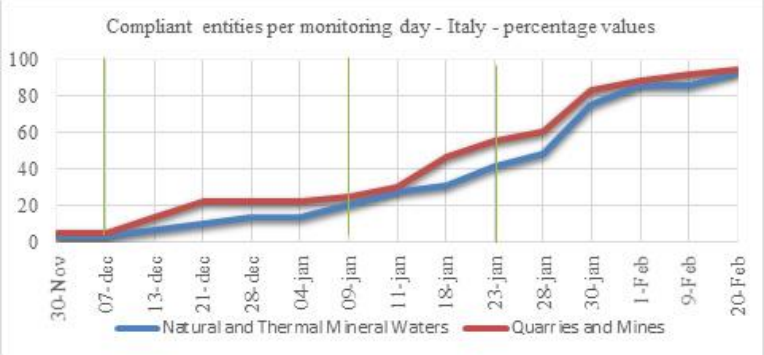
The monitoring also included: a) the expected and received questionnaires for the validity of the extraction authorization (expired in the reference year of the collected data, expired in previous years, active from that year, valid also in the following year); b) questionnaires sent with non-blocking errors, questionnaires being processed with blocking errors and questionnaires being processed with non-blocking errors.

4.1. Progress of the data collection

The survey began on October 20, 2022, expiring on January 31, 2023, Institutions involved were 29 for natural and thermal mineral waters and 36 for quarries and mines.

Through the processing of the data made available by the implemented monitoring system, it can be seen that the response rate has increased as the detection deadline approaches, also following the sending of reminders, in particular for those reminders sent close to the data transmission deadline (Figure 3).

Figure 3 – *Response rate for Natural and Thermal Mineral Waters and Quarries and Mines. Italy. 2022-2023 edition. Percentage values.*



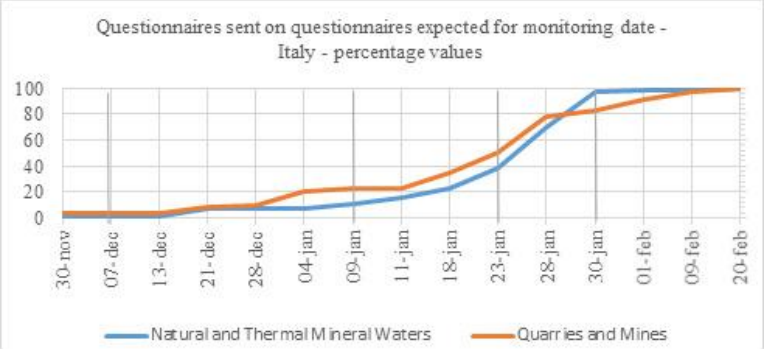
The graphs show a very slow start of the collection, quite common phenomenon, with a very small percentage of questionnaires sent compared to those expected until you get close to the two weeks before the deadline.

During the survey, four reminders were sent via email, in addition to the support provided by the survey contact persons for the data collection.

As can be clearly seen from the graphs, the most noticeable effects are found especially with the latest reminders. A total of 790 natural mineral water and thermal water questionnaires and 3,618 quarry and mine questionnaires were collected.

Figure 4 shows the trend of data collection, in particular the collection of questionnaires sent on the expected questionnaires in percentage terms for the two issues in the survey period.

Figure 4 - *Questionnaires sent on expected questionnaires. Natural and Thermal Mineral Waters and Quarries and Mines. Italy. 2022-2023 edition. Percentage values.*



The Umbria Region should be reported as a virtuous region, which completed the survey in the first week of data collection and two Institutions for each survey (for waters the Liguria Region and the Caltanissetta Mining District, for quarries and mines the Sardinia Region and the Caltanissetta Mining District), which replied after the end of the survey, as a consequence of the memorandum of 30 January 2023, for which it was been necessary to extend the survey until 20 February 2023.

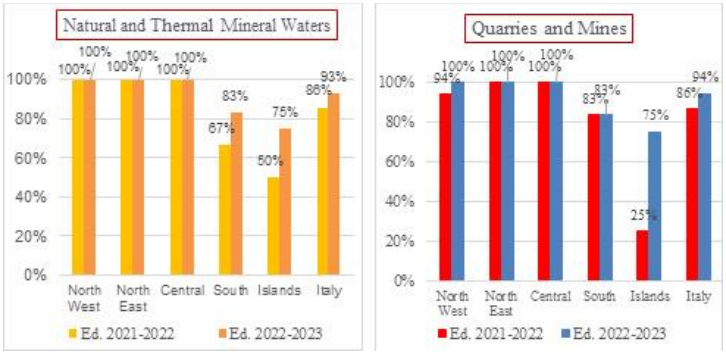
The Region of Calabria and the Mining District of Palermo were totally in default.

4.2. Response rate

In the 2022-2023 edition, the response rate was slightly lower than 95% for both surveys, recording an increase, compared to the previous edition, of about 8 percentage points. The increase is due to the greater participation of the Agencies of the Northwest, of the South and above all of the Islands.

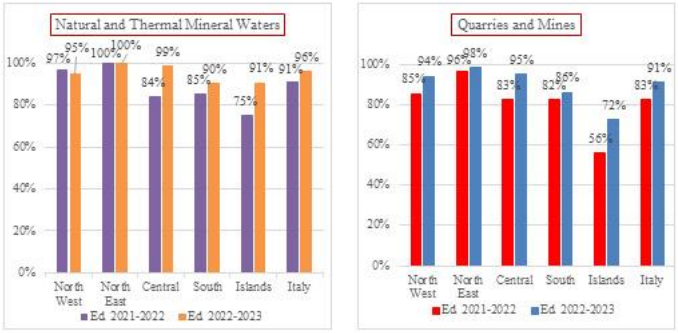
Figures 5 and 6 show the response rate of the survey for the two issues and by territory comparing the data of the 2021-2022 edition and the 2022-2023 edition.

Figures 5 and 6 - *Response rate for the two themes and geographical breakdown. 2021-2022 and 2022-2023 editions. Percentage values.*



Figures 7 and *8* - *Processing status of the questionnaires for the two topics. 2021-2022 and 2022-2023 editions. Percentage values.*

Figures 9 and 10 – *Coverage rate of expected questionnaires for the two themes. 2021-2022 and 2022-2023 editions. Percentage values.*



In the 2021-2022 edition there was also a partial submission of data by two Institutions for the theme "quarries and mines".

In the 2022-2023 edition, unlike what happened in the previous edition, for both surveys, Mineral and thermal waters and Quarries and mines, all the Institutions that participated in the survey, sent the required data by the end of the collection with no Institution left in the state "in progress", Figures 7 and 8.

In the last edition of the survey, more than 90% of the expected questionnaires were received, with an increase, compared to the 2021-2022 edition, of 5 percentage points for the survey on mineral and thermal waters and 8 percentage points for the survey on quarries and mines, Figures 9 and 10.

5. Conclusion

The creation of a comprehensive and up-to-date information framework at territorial level is useful for economic and environmental analysis, allowing the planning of national policy strategies on raw materials, for their sustainable exploitation and for the identification of specific intervention measures at national and territorial level.

Data collection is one of the fundamental activities within the production process of a data; the study of operational strategies to improve the quality, timeliness and response rate, are, therefore, of primary importance.

Since the first edition of the survey, Istat has tried to introduce important changes in the data collection process, implementing methodological innovations that would allow to standardize the collected information and improve the quality of statistical production (ISTAT, 2013-2014, 2015-2016, 2017, 2018, 2020).

The 2022 edition was characterized by the introduction of some important innovations in data collection that have allowed to carry out the survey with greater efficiency and effectiveness. In this contribution we highlighted the role of monitoring to improve data quality.

The use of careful monitoring carried out throughout the survey, not only using the classic reports of the Gino++ Portal, but also through a timely comparison between expected sites and sent sites, identified any missing information to be requested from respondents.

The designed and implemented process was therefore effective, it performed very well the tasks for which it was created, and efficient as specified quantitatively in the analysis of the monitoring results.

References

ATZENI P., CERI S., FRATERNALI P., PARABOSCHI S., TORLONE R., 2013. *Basi di dati. Modelli e linguaggi di interrogazione*. 4D, Ediz. McGrill.

CAMAGNI P., NIKOLASSY R., 2014. *Creare database relazionali con SQL e PHP*. Ediz. HOEPLI.
CHIANESE A., MOSCATO V., PICARIELLO Z., SANSONE L. 2015. *Sistemi di basi di dati e applicazioni*. Editore Apogeo Education - Maggioli. Collana Apogeo Education.
ISPRA, 2020 – *Database GEMMA.* Banca Dati Nazionale Geologico Mineraria Museale Ambientale.
ISTAT 2012, *Gino++, un sistema generalizzato per indagini statistiche via web,* NEWSSTAT 4/2012, Newsletter.
ISTAT 2013. *Metodi statistici per il record linkage*.
ISTAT 2013-2014, Statistica Report *Le attività estrattive da cave e miniere.*
ISTAT 2015-2016, Statistica Report *Le attività estrattive da cave e miniere.*
ISTAT 2017, Statistica Report *Le attività estrattive da cave e miniere.*
ISTAT 2018, Statistica Report *Le attività estrattive da cave e miniere.*
ISTAT 2020. Statistica Report *Le attività estrattive da cave e miniere.*
ISTAT 2022. *Regole e strategie nel trattamento digitale e nella produzione dei dati quantitativi e qualitativi* (pag. 49). Roma: Istat, Metodi - Letture statistiche.
ISTAT 2023. *Le statistiche dell'Istat sull'acqua. Anni 2020-2022.* Statistica Report per la Giornata Mondiale dell'acqua.
ISTAT 2017-2022, Annuario Statistico Italiano Cap. 2 - Ambiente ed energia, Pressione antropica.
ISTAT 2013-2021. Banca Dati I.STAT Tema Ambiente ed energia, Sezione Cave e miniere.
TORELLI R., 2011. *GINO++: A generalized system for web surveys*, Proceedings of Statistics Canada Symposium 2011.
VIGNANI D., 2017. *La rilevazione sulle attività estrattive. Nuove informazioni per lo studio dell'ambiente e del territorio.* Istat Convegno Scientifico "Le attività estrattive e l'ambiente: fabbisogni informativi e nuove disponibilità di dati", Roma 2017.
VIGNANI D., 2019 *Il quadro delle attività estrattive in Italia*. Ecomondo 2019 Rimini Fiere, 5 novembre.

Lucia MONGELLI, mongelli@istat.it
Sabrina ANGIONA, ansabrin@istat.it
Barbara BONINFANTE barbara.boninfante@istat.it
Valentino PARISI parisi@istat.it
Barbara STOBBIA stobbia@istat.it
Stefano MOSCATELLI stmoscat@istat.it

TOURISM AS A MEANS TO COUNTERACT INNER AREAS (IAS) DEPOPULATION: THE CASE STUDY OF CAMPANIA REGION[1]

Antonella Bianchino, Daniela Fusco, Paola Giordano, Maria Antonietta Liguori, Donato Summa

Abstract. In recent years, there has been a growing interest in rural areas, affected by profound changes mainly in the shift of economic activity and population to urban areas. These phenomena have led to a crisis in the traditional structure and organization of rural areas, exposing them to the risk of economic, social and environmental decline. To combat demographic decline and promote sustainable territorial competitiveness, the Italian National Strategy for "Inner Areas" (SNAI) was born in 2013. Rural tourism has a high potential to stimulate local economic growth and social change because of its complementarity with other economic activities. These positive externalities can stop Inner Areas (IAs) population loss. This study proposes a theoretical model for the construction of a synthetic index to measure rural tourism in IAs, starting from 22 basic indicators declined in three Pillars: 1) Infrastructural density and touristic fluxes; 2) Economic impact of touristic sector and 3) Agricultural sector support. It was used a multisource approach (statistical and administrative sources, big data), by analysing 8 data sources: survey and administrative sources, Statistical registers and Big Data on Agritourisms. The research focused on the areas of the Campania Region and the complexity of rural tourism in the Peripheral, Ultraperipheral Areas (165 Municipalities) was compared with the Urban Centres and Belt (260 Municipalities) and the Intermediate municipalities (126). The result provides a tool for monitoring tourism sector by stakeholders and policy makers, useful to promote sustainable territorial competitiveness of IAs.

1. Rural tourism and Italian National Strategy for "Inner Areas" (SNAI)

As registered by the World Tourism Organization (WTO), tourism connected to food and countryside is a quickly growing segment. A high percentage of tourists does choose tourist destination on the basis of quality food as well as of the will to

[1] Authors contributions: Antonella Bianchino the revision of the whole article; Daniela Fusco paragraphs 1, 2.3 and References; Paola Giordano paragraph 3, related subparagraphs and Summary; Maria Antonietta Liguori paragraph 2.1 and Appendix; Donato summa paragraph 2.2.

gain positive experiences by tasting local products in small locations, renowned for their quality and their strong connection to the territory. In this way, a strong relationship has been created among free time, tourism, food rediscovery, and rural areas' promotion, able to exert an influence on both demand and supply of tourist services (Piñeiro *et al*., 2019).

The rural areas are also affected by profound changes mainly in the shift of economic activity and population to urban areas. These phenomena have led to a crisis in the traditional structure and organization of rural areas, exposing them to the risk of economic, social and environmental decline. To combat demographic decline and promote sustainable territorial competitiveness, the Italian National Strategy for "Inner Areas" (SNAI) was born in 2013. The risk is that demographic decline and geographical marginality will be followed by dysfunctional processes that undermine the very provision of basic services. The worrying population loss is confirmed by the data: against a positive demographic dynamic on a national basis between 2001 and 2020 (+3.9%), growth in the central municipalities was even more accentuated (+5.6%) while the inland areas lost the overall population (-1.4%). This decrease is particularly relevant because it is precisely the most marginal municipalities of the Inner Areas (Peripherals and Ultra-peripherals Regions) that have suffered the biggest drops (respectively -4.7% and -9.1%) (ISTAT, 2022).

The identification of the Inner Areas (IAs) starts from a polycentric reading of the Italian territory, characterized by a network of municipalities or aggregations of municipalities (service offering centers) around which gravitate areas characterized by different levels of spatial marginality (Agenzia per la Coesione Territoriale, 2021).

SNAI looks at the entire Italian territory in its articulation at municipal level and identifies Municipalities with a joint offer of three types of services - health, education and mobility - known as Poles/Inter-municipal Poles. It also classifies all the other Municipalities based on their distance from these poles (in terms of actual average travel times on the road), classifying them into four categories at increasing relative distance - Belt, Intermediate, Peripheral, Ultraperipheral - and, therefore, with greater potential discomfort in the use of services. Municipalities classified as Intermediate, Peripheral and Ultraperipheral represent the set of IAs of our country.

In the IAs, the agricultural, pastoral and forestry sectors play a central role as opportunities for economic growth and for the value of care and environmental prevention (Lucatelli and Storti, 2019). For example, agricultural soil management is essential in areas with high levels of landslide risk and hydrogeological disruption; pastoralism contributes to the vitality of mountain giving a contribution to maintaining biodiversity and combating soil degradation.

Rural tourism has a high potential to stimulate local economic growth and social change because of its complementarity with agriculture and other economic

activities (World Tourism Organization, 2023). In the long term, it enhances economic growth by stimulating investment in new infrastructure and human capital, and increases competition, promoting industrial development, creating jobs and thereby increasing income. These positive externalities can stop Inner Areas (IAs) population loss.

The aim of the work is to represent the complexity of rural tourism in IAs, compared with Urban Poles and "Belt" municipalities, analyzing main components and driving forces, using a multisource approach. The study identifies useful indicators for the evaluation of these phenomena by exploiting the opportunity given by using both Big Data and traditional sources.

The research focused on the areas of the Campania Region and the complexity of rural tourism in the Peripheral, Ultraperipheral Areas (165 Municipalities) was compared with the Urban Centers and Belt (260 Municipalities) and the Intermediate municipalities (126).

2. Rural tourism in IAs measurement: a theoretical model

2.1. Multi-source approach

Many National Statistical Institutes (NSIs), especially in Europe, are moving from single-source to multi-source statistics, due to higher quality demands on the statistics produced. NSIs have traditionally produced statistics from a single source, where basically only data from a single data source is used and other data sources are also often used only as ancillary data. In most cases, the single data sources are surveys, although nowadays administrative data are increasingly used as unique data sources and Big Data is also starting to be used. By combining survey data with administrative data and Big Data already available, INSs can reduce data collection and processing costs and reduce the burden on respondents (de Waal *et al.*, 2019).

A multisource approach (statistical and administrative sources, Big Data) was used to meet the need to develop a system of homogeneous, comparable and up-to-date statistics. For the construction of indicators were used 8 data sources: survey and administrative sources (Continuous Population Census, Survey of Museums and similar institutions, Agricultural Census, Capacity of Collective tourist accommodation establishments, Statistical Atlas of the Municipalities), Statistical registers (Statistical register of active enterprises - ASIA, Frame SBS) and Big Data on Agritourisms. Sources are described as below.

• *Capacity of collective tourist accommodation establishments survey*. It is a total survey by Istat, carried out annually (https://www.istat.it/it/archivio/210783). The quantitative survey, at the single municipality level, of the number of establishments,

beds, bedrooms and bathrooms for hotels; of exercises and beds for the other structures. (Data source for indicators 1-5).

• *Continuous Population Census*. Istat makes available the most recent official data on the population in Italian Municipalities deriving from the surveys carried out at the Registry and Civil Status Offices of the Municipalities (http://demo.istat.it/) and from the Population Census (http://dati-censimentipermanenti.istat.it/). Personalized queries (by year, territory, citizenship, etc.) allow you to build the tables of interest and download the data in a reprocessable format. It is also possible to find information on the main demographic phenomena, such as birth and death rates, forecasts of the resident population, the old age index, the average age (Data sources for indicators 1-3).

• *Survey of museums and similar institutions*. The survey is an annual survey of a census nature, conducted by Istat in collaboration with the Ministry of Culture (MiC), the Regions and the Autonomous Provinces (https://www.istat.it/it/archivio/6656). (Data source for indicator 6).

• *Statistical Register of Active Enterprises - ASIA*. The Istat Statistical Register of Active Enterprises (Asia) was established under EC Regulation no. 177/2008 relating to the community coordination of the development of business registers used for statistical purposes (http://www.istat.it/it/archivio/216767). The Register plays a central role in economic statistics: it is used for National Accounting estimates and identifies the reference population for sampling plans and for reporting to the universe of the main business surveys conducted by Istat. (Data source for indicators 7, 9 and 10).

• *FRAME SBS*. The Istat SBS Frame (https://www.istat.it/it/archivio/249448) is an integrated system of administrative and statistical data, created annually by Istat for the estimation of the economic results of companies, starting from the units included in the statistical register of companies (ASIA). It is currently used both for the production of estimates of structural statistics on companies (SBS), and as a relevant source in the estimates of the National Accounts. (Data source for indicator 8).

• *Big Data on Agritourisms*. Istat decided to experiment the possibility to collect information on agritourisms through a web scraping technique, in order to enable the collection of sparse and unstructured information in the Internet, belonging to the vast category of Big Data (Barcaroli *et al*., 2016a). Data source is http://www.agriturismoitalia.gov.it (Data source for indicators 11-15).

• *Statistical Atlas of the Municipalities*. The Statistical Atlas of the Municipalities by Istat (https://asc.istat.it/ASC/) provide a dynamic information, produced both by Istat and by other bodies belonging to the national statistical system. Each series of data (currently around 150 in total) is then enriched by a collection of metadata

which guide the user towards a correct reading and interpretation of the contents (Data source for indicators 4, 11 and 12).

• *Agricultural Census*. Since data at municipality level of the 7th Agricultural Census (2021) are not yet available, we based our analysis on last available data, referred to 2010 Agricultural Census, as a proxy (Data source for indicators 16-22).

2.2. Big Data (web data) analysis on agritourisms: an innovative approach

Usually, the main source of information for these farms is represented by the administrative data of the authorized Agritourisms, collected both by Istat and by Ministry of Agriculture in different surveys. To enhance the data quality (i.e. timeliness, accuracy, punctuality) and to increase the amount of information related, in this work was proposed the use of web scraping techniques.

Web scraping is a process of extracting data from websites using software programs; it has become an increasingly popular technique in recent years due to the growth of the internet and the large amount of data available online. Web scraping can be used for a variety of purposes, including data analysis, research, and business intelligence (Barcaroli *et al*., 2015; Barcaroli *et al*., 2016b)

There are two main types of web scraping: specific and generic web scraping. Specific web scraping involves scraping websites where the structure and content are known in advance. This means that the software program can be programmed to replicate the behavior of a human user visiting the website and extract only the relevant information from the website.

Generic web scraping, on the other hand, involves scraping websites where no prior knowledge of the content is available. This means that the software program must scrape the entire website and then use machine learning or other techniques to infer relevant information.

In this work, the web data acquisition focuses on specific web scraping. Custom software programs have been developed to extract information from the website http://www.agriturismoitalia.gov.it, which is the official website for Italian Agritourisms, including about 25,000 units.

Specifically, the list of Italian official Agritourisms was acquired first, then for each enterprise in the list, detailed information was downloaded. For both activities custom web-scraping programs have been developed from scratch by the team in Python.[2]

The extracted data were stored in a tabular data format and automatically processed using a Python script to produce the dataset used for analysis in this study.

[2] The software used for web-scraping has not yet been made public.

Big Data preparation is the process of cleaning and transforming raw data prior to processing and analysis.

- Normalization of the denominations of the municipalities (122 agritourisms for which it was not possible to match an official municipality on the basis of the name)
- Manual correction of incorrect municipality names
- Search and resolution of duplicates
- Attribution of municipality codes
- Attribution of the number of beds: num_posti_letto_nel_comune = (num_camere_nel_comune * 2) + (num_appartamenti_nel_comune * 4) + (num_piazzole_nel_comune * 4).

2.3. Pillars and indicators proposed

For the assessment of driving forces that affect rural tourism in IAs, this study proposes a theoretical model for building a synthetic index starting from 22 basic indicators declined in three Pillars: 1) Infrastructural density and touristic fluxes; 2) Economic impact of touristic sector and 3) Agricultural sector support. The estimation of a complex phenomenon as tourism, with the use of a synthetic index, summarizes the concept at the highest levels, leaving little space to the analysis of the individual facets, but represents a photograph of the phenomenon, useful for the evaluation of touristic and agricultural ex post policies.

Indicators are described in the Appendix (Table 1).

3. Results and Final remarks

3.1. Results

Focusing on Campania region, the following areas have been joined for the calculation of indicators:

- Group A: Poles/Inter-municipal Poles, Belt (260 municipalities)
- Group B: Intermediate (126 municipalities)
- Group C: Peripheral, Ultraperipheral (165 municipalities).

Pillar 1 (Table 2): the potential of accommodation facilities is evidently very high in Peripheral and Ultraperipheral Areas. For example, the indicator Total accommodation rate (Figure 1) where the value for group C (222.6) is taller than

group A (17.3) and group B (102.4). Exception is Visitor pressure on museum and similar institutions, which, as expected, is higher in the Poles.

Table 2 – *Basic indicators by areas - Pillar 1: Infrastructural density and touristic fluxes.*

Infrastructural density and touristic fluxes	Total accommodation rate	Rate of accommodation		Density of establishments, hotels and accommoda-tion facilities	Incidence of accommodation at municipality level	Visitor pressure on museum and similar institutions per inhabitant
		high-end hotel structures	extra-hotel facilities			
Poles Areas	17.3	6.3	6.8	17.5	1.5	2,527.8
Intermediate Areas	102.4	26.6	58.9	13.6	1.0	1,087.7
Peripheral/ Ultrap. Areas	222.6	63.1	100.0	19.2	1.9	1,969.5
Campania	41.8	12.8	19.1	17.1	4.5	2,351.7

Pillar 2 (Table 3): even the economy of the tourism sector confirms the potential of Peripheral and Ultraperipheral Areas, particularly the Added value per capita of the tourism sector registers in the group C shows a twice value (1,398.8) compared to group B (725.5) and more than quadruple value compared to group A (321.9).

Table 3 – *Basic indicators by areas - Pillar 2: Economic impact of touristic sector.*

Economic impact of touristic sector	Incidence of employment in the tourism sector	Added value per capita of the tourism sector	Incidence of employment in the tourism-related entertainment sector	Localization quotient of local tourist unit employees
Poles Areas	7.7	321.9	1.2	0.9
Intermediate Areas	11.5	725.5	1.0	1.3
Peripheral/ Ultrap. Areas	17.2	1,398.8	1.4	2.0
Campania	8.7	446.1	1.2	1.0

Pillar 3 (Tables 4a,b): the density of Agritourisms (Figure 2) is lower in group C (0.05). The used source of Big Data, by its nature, excludes a part of the Agritourisms, in particular those not in possession of the certification. (Agritourisms with certification can use the label on the right). In fact, thanks to the "Agriturismo Italia" brand by Ministry of Agriculture, tourists and professional operators can easily distinguish officially accredited businesses. This distinction is very important, especially for the international market where the Italian agritourism reality is not always perfectly known and the various operators could easily be disoriented by

other forms of hospitality, equally present in the rural area. Therefore, it cannot be excluded that a part of the Agritourisms is missing, even if they are located in the Peripheral and Ultraperipheral Areas, but not in possession of the certification.

This is caused by uneven territorial marketing throughout the territory and the lack of associations among the structures. Agricultural holding data source refers to Agricultural Census 2010, since data at municipality level of 2021 Census are not yet officially released. In this case we used this source to calculate basic indicators for Pillar 3, but we'll update these indicators. Moreover, 2010 used data refer to the location of the holding, in most cases the legal address, instead of the location of lands (cadastral parcels) or livestock.

Table 4a – *Basic indicators by areas - Pillar 3: Agricultural sector support.*

Agricultural sector support	Density of Agritourisms	Agritourisms with accommodation	Share of Agritourisms with catering services	direct sale	other gainful activities except direct sale
Poles Areas	0.06	0.51	88.38	41.90	63.73
Intermediate Areas	0.06	0.64	91.74	40.87	56.09
Peripheral/ Ultrap. Areas	0.05	0.41	77.22	34.60	36.71
Campania	0.06	0.51	85.89	39.28	52.86

Table 4b – *Basic indicators by areas - Pillar 3: Agricultural sector support (2).*

Agricultural sector support	Share of holding with grapes for PDO/PGI wines	organic farming UAA	organic farming livestock	wooded area	permanent crops	permanent pasture and meadows	short rotation coppices
Poles Areas	27.03	0.95	1.29	15.05	77.13	4.78	0.82
Intermediate Areas	18.73	1.55	1.26	24.88	85.19	9.48	1.55
Peripheral/ Ultrap. Areas	8.89	1.52	0.97	33.85	81.34	20.58	1.62
Campania	18.92	1.30	1.16	23.59	80.91	10.83	1.28

Figure 1 – *Total Accommodation rate*.

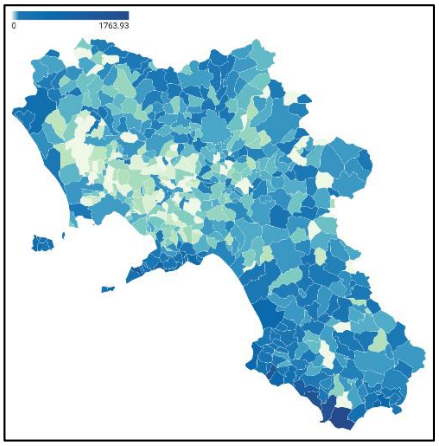
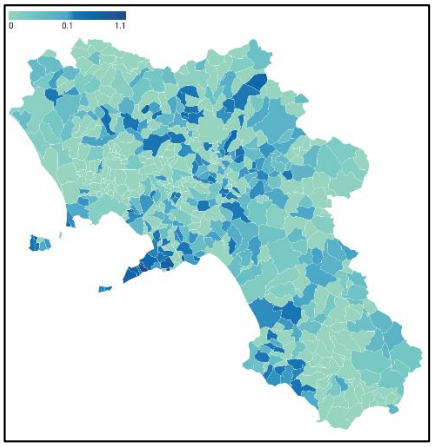


Figure 2 - *Density of Agritourisms.*



3.2. Final Remarks

The rural tourism of Inner Areas, to a certain point, depends on the agricultural sector. The construction of a new offer (diversification of activities) and the construction of an integrated and organized offer of high-typical goods and services (specialties and integrated specialties) in synergy with tourism is necessary for the sector development.

The decline in agricultural and other forms of rural employment in many countries has created a need for a diversified range of rural businesses. In most cases, rural tourism has become an important element of the diverse activities and development in rural areas.

Switching from single source to multisource statistics seems like the way to go. However, this transition is not easy. Multisource statistics come with new problems that need to be overcome before the resulting output quality is sufficiently high and before those statistics can be produced efficiently (de Waal *et al*., 2019).

The limit of the use of multi-source statistics is that supporting data come in many different varieties as data sets can be combined in many different ways. Every data source has its own problems (*i.e.* time lag, different target) for which customized solutions are needed.

Despite these limits, the described tool is important to monitor the trend of rural tourist supply and demand in internal areas with the use of indicators in order to guide the stakeholders in the political choices about sustainable IAs.

The next step will be to calculate and to evaluate the strength of synthetic indices and calculate them for all Italian regions.

Appendix

Table 1 - *Pillars, Basic indicators, Algorithms and related Data sources. (segue)*

Pillar	Indicator	Algorithm	Data source
Infrastructural density and touristic fluxes	Total accommodation rate	Number of total beds per 1,000 inhabitants	Istat, Capacity of collective accommodation establishments survey; Istat, Continuous population census
	Rate of accommodation of high-end hotel structures	4-5 star hotel beds per 1,000 inhabitants	idem
	Rate of accommodation of extra-hotel facilities	Extra-hotel beds per 1,000 inhabitants	idem
	Density of establishments, hotels and accommodation facilities	Total beds per km^2	Istat, Capacity of collective accommodation establishments survey; Istat, Statistical Atlas of the Municipalities
	Incidence of accommodation at municipality level	Total beds of the Municipality/Total beds at national level (%)	Istat, Capacity of collective accommodation establishments survey;
	Visitor pressure on museum and similar institutions per inhabitant	Visitors of museum and similar institutions per 1,000 inhabitants	Istat, Survey of museums and similar institutions
Economic impact of touristic sector	Incidence of employment in the tourism sector	Employees of tourist Local Unit/Total employees of Local Units of Municipality	Istat, Statistical Register of Active Enterprises (ASIA)
	Added value per capita of the tourism sector	Added value of tourist Local Units per inhabitant	Istat, FRAME SBS
	Incidence of employment in the tourism-related entertainment sector	Employees of tourism-related entertainment Local Units/Total employees of Local Units of Municipality	Istat, Statistical Register of Active Enterprises (ASIA)
	Localisation quotient of local tourist unit employees	[Employees of the Tourist Local Unit of Municipality/ Employees of the total Local Unit of Municipality]/[Employees of the Tourist Local Unit Italy/ Employees of the total Local Unit Italy]	idem

Table 1 - *Pillars, Basic indicators, Algorithms and related Data sources. (continua)*

Pillar	Indicator	Algorithm	Data source
Agricultural sector support	Density of agritourisms	Agritourisms per km^2	Big Data on agritourisms; Istat, Statistical Atlas of the Municipalities
	Density of agritourisms with accomodation	Beds of agritourisms per km^2	idem
	Share of agritourisms with catering services	Agritourisms with catering services/Total agritourisms	Big Data on Agritourisms
	Share of agritourisms with direct sale	Agritourisms with direct sale/Total agritourisms	idem
	Share of agritourisms with other gainful activities except direct sale	Agritourisms with other gainful activities except direct sale/Total agritourisms	idem
	Share of holdings with grapes for PDO/PGI wines	Holdings with grapes for PDO/PGI wines/Total holdings with vineyard	Agricultural Census
	Share of holdings with organic farming UAA	Holdings with organic farming UAA/Total holdings with UAA	idem
	Share of holdings with organic farming livestock	Holdings with organic farming livestock/Total holdings with livestock	idem
	Share of holdings with wooded area	Holdings with wooded area/Total holdings	idem
	Share of holdings with permanent crops	Holdings with permanent crops/Total holdings with UAA	idem
	Share of holdings with permanent pasture and meadows	Holdings with permanent pasture and meadows/Total holdings with UAA	idem
	Share of holdings with short rotation coppices	Holdings with short rotation coppices/Total holdings	idem

Acknowledgements

The authors would like to thank Melania Palma and Alessandro di Giorgio for their support in data quality control.

References

AGENZIA PER LA COESIONE TERRITORIALE. 2021. Le aree interne: di quali territori parliamo? Nota esplicativa sul metodo di classificazione delle aree. *Methodological note*. Cohesion Policy Department DPS. 17 pp.

BARCAROLI G., FUSCO D., GIORDANO P., GRECO M., MORETTI V., RIGHI P., SCARNÒ M. 2016a. ISTAT Farm Register: Data Collection by Using Web Scraping for Agritourism Farms. In *Proceeding of International Conference on Agricultural Statistics. ICAS VII,* FAO, Rome, 26-28 October 2016.

BARCAROLI G., SCANNAPIECO M., SUMMA D. 2016b. On the Use of Internet as a Data Source for Official Statistics: a Strategy for Identifying Enterprises on the Web. *Rivista Italiana di Economia Demografia e Statistica* volume LXX N. 4 (Ottobre-Dicembre 2016).

BARCAROLI G., SUMMA D., SCARNÒ M., SCANNAPIECO M. 2015. Using Internet as a Data Source for Official Statistics: a Comparative Analysis of Web Scraping Technologies. In *Proceedings of Conference NTTS2015 - New Techniques and Technologies for Statistics.* Brussels, 9-13 March 2015.

DE WAAL T., VAN DELDEN A., SCHOLTUS S. 2019. Multi-source statistics: Basic situations and methods. *International Statistical Review*, 88(1), 203-228.

ISTAT. 2022. La geografia delle Aree Interne nel 2020: vasti territori tra potenzialità e debolezze. *Istat Statistiche focus*, 20 luglio 2022.

LUCATELLI S., STORTI D. 2019. La strategia nazionale aree interne e lo sviluppo rurale: scelte operate e criticità incontrate in vista del post 2020. *Agriregionieuropa* anno 15 n°56, Mar 2019.

PIÑEIRO M. V., DE SALVO P., GIOMMI F. 2019. Rural Tourism and Territorial Development in Italy. In M.J. BASTANTE-CECA (ed.). *Sustainability Assessment at the 21st century*. IntechOpen, 198 pp.

WORLD TOURISM ORGANIZATION. 2023. Tourism and Rural Development: A Policy Perspective, UNWTO, Madrid, 2023.

Antonella BIANCHINO, Istat, bianchin@istat.it
Daniela FUSCO, Istat, dafusco@istat.it
Paola GIORDANO, Istat, pgiordano@istat.it
Maria Antonietta LIGUORI, Istat, liguori@istat.it
Donato SUMMA, Istat, donato.summa@istat.it

A MULTIDIMENSIONAL ANALYSIS FOR MONITORING ENVIRONMENTAL HEALTH IN ITALY[1]

Giuseppe Lecardane, Flora Fullone, Monica Carbonara

Abstract. In the last few years, attention and sensibility towards problems relating to environmental issues has grown. Consequences of climate change, reduction of water resources, urban pollution and other manifestations of environmental crisis transversally affects the life and activities of human beings. In this work some main indicators for the assessment of environmental quality in the national territory are analysed. The complex and multidimensional nature of environmental phenomena has required identification, selection and measurement of a series of indicators relating to 6 macro areas (geological and natural risks, consumption, emissions, waste, protected areas, sustainability and environmental certification) to the implementation of more effective and incisive information and operational programs in the area. Given multidimensionality of the phenomena, having analyzed and described the individual dimensions, a multivariate synthesis analysis was carried out to compare the state of environmental health at a regional level. The results show that the negative impact on the environment is in all Italian regions but is stronger in southern Italy. It is a study that offers in-depth insights to subjects engaged in development policies in compliance with environmental sustainability.

1. Introduction

The aim of this work is to analyse the impact of human activities on the Italian regions. The environmental analysis presents some critical issues since there are several aspect to identify and sometimes they are not homogeneous. A set of objective indicators has been selected to express the different environmental pressures caused by human and economic activities, moreover a subjective sentiment of environmental satisfaction has also been included. The indicators are evaluated in

[1] The paper is the result of the common work of the authors. In particular: sections are attributed as follows: F. Fullone paragraphs 1 and 2, M. Carbonara paragraph 3 and G. Lecardane paragraphs 4 and 5.

relation to the regional area and resident population, and then different multidimensional analyses are carried out.

2. Factor analysis

There are many public indicators, provided by the research institutes Istat and Ispra, regarding different aspects of the environmental dimension at regional level. The research hypothesis to be tested is whether there are a few latent factors that influence the observed indicators.

A factor technique has been used to carry out an exploratory analysis of the observed indicators of the environmental phenomena. The assumption is that the analysed variables represent the effects of one or a few latent variables (factors), that are not directly observable and measurable (reflexive model). The indicator's long list is represented in Table 1.

Correlation analysis shows that some indicators are highly correlated (Pearson correlation factor >0,9) for example (nitrous oxide emissions, carbon dioxide emissions). In this case they seem to provide the same information, so a short list has been selected (Tab. 2).

Table 1 - *Environmental indicators, long list.*

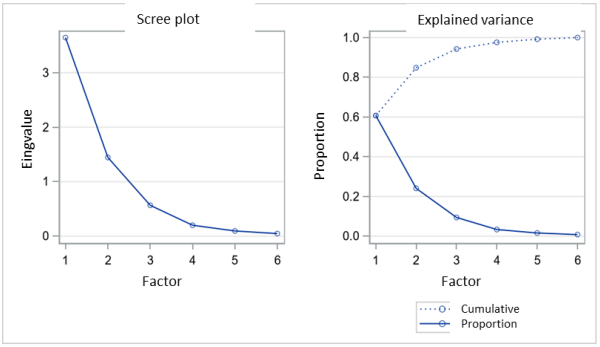
Environmental Issue	Environmental indicators	*Source*	*Year*
Soil	Soil consumption	*ISPRA*	*2021*
	Amount of fertilizer used	*ISTAT*	*2021*
Raw material	Raw material taken	*ISTAT*	*2020*
Population	Population density	*ISTAT*	*2021*
Companies in the environmental area	Active companies with 3 or more employees that reduce the environmental impact of their activities (2018 Census)	*ISTAT*	*2018*
	Environmental quality UNI-EN-ISO 14001 certifications for company and production sites	*ISPRA*	*2020*
Waste Sorting	Urban waste collected separately	*ISPRA*	*2020*
	Special and hazardous waste incinerated and energetically recovered	*ISPRA*	*2020*
Atmospheric emissions	Nitrous oxide emissions N20	*ISPRA*	*2020*
	Carbon dioxide emissions CO2	*ISPRA*	*2020*
Environmental quality Perception	Environmental Satisfaction	*ISTAT*	*2020*

Table 2 - *Environmental indicators, short list.*

Indicators	*Dimension*
Final consumption of electricity (agr, industrial, tertiary, residents) per unit area	*GWh per km²*
Soil consumption	*% of regional area*
Raw material taken with respect to the regional area	*1000 tons per km²*
Population density	*residents per km²*
Percentage of people very or fairly satisfied with the environmental situation (air, water, noise) of the area in which they live.	*% of residents aged 14 and over*
Active companies with 3 or more employees that reduce the environmental impact of their activities (2018 Business Census) compared to the total regional population	*n. companies per residents population*

The selected indicators describe the impact of human activities on the various regional areas, characterised by artificially covered soil, energy consumption, raw material extraction, and population density. At the same time, we will consider the number of companies that reduce the environmental impact of their activities, in relation to the resident population, and the percentage of people who express a positive assessment of the environmental situation of the area in which they live. The factor extraction method used is the main components, which aims to identify a linear combination of the observed indicators, able to explain most of their variance.

Figure 1 - *Factor analysis: scree plot, eingvalues.*

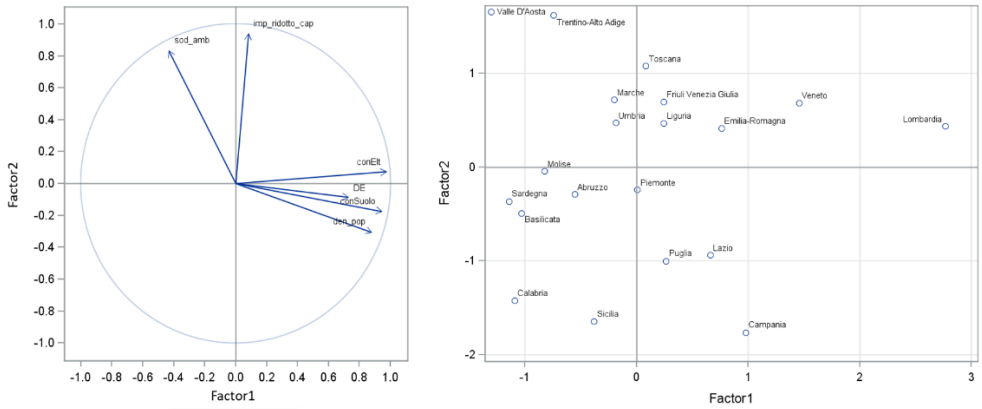


Kaiser measure of sample adequacy: MSA total = 0.7

	Eingvalue	*Proportion*	*Cumulative*
1	3,64	0,61	0,61
2	1,44	0,24	0,85
3	0,57	0,09	0,94
4	0,20	0,03	0,98
5	0,10	0,02	0,99
6	0,05	0,01	1,00

From the six indicators described in the short list, two factors have been extracted, that explain 85% of the total variance (Fig. 1). However, the variance explained by the two factors for each variable is not less than 65%.

Figure 2 - *Rotated factor pattern, left image; factor score of the regions on the right.*



The factors are easy to interpret on the basis of figure 2; the first, which explain the 65% of the total variance, is highly correlated with the consumption of resources such as soil and energy, the extraction of raw material and the environmental impact of the population on the territories. The second factor, on the other hand, reflects the reduced impact of economic activities in relation to the resident population and a general sentiment of environmental satisfaction.

The factor score of the regions (Fig.2) gives the representation of each region on the factorial coordinates. In the first quadrant of the Cartesian plane we can observe the central-northern regions where consumption and environmental pressures are high, and at the same time, there is a moderately positive perception of the environmental situation.

In the second quadrant, factor1 which reflects consumption and environmental pressure on the territory, is moderately low, and there is a medium-high environmental perception above all in the Alpine regions. In the third quadrant, consumption and environmental pressures are not high, but the environmental perception is not positive. In the fourth quadrant the large southern regions are represented, where environmental pressures are moderate, but environmental perception is medium-low, especially in Campania.

3. Cluster analysis

In order to explore the main and most significant similarities and differences in the state of environmental health among the Italian regions, the Cluster analysis was carried out.

Ward's hierarchical method was used for the analysis and the quadratic Euclidean distance was used to measure the dissimilarity between the statistical units.

For the analysis, the variables considered in the factor analysis previously described in paragraph 2 were used:

- Final consumption of electricity (agriculture, industry, tertiary, residual)
- Consumed soil
- DE Raw material taken with respect to the regional area
- Population density
- Environmental satisfaction
- Active companies with 3 or more employees that reduce the environmental impact of their activities

The cluster analysis made it possible to identify five different groups' number and territorial distribution (Fig. 3):

Cluster 1: Abruzzo, Molise, Basilicata, Calabria, Sardegna.

Cluster 2: Piemonte, Friuli Venezia Giulia, Liguria, Toscana, Umbria, Marche.

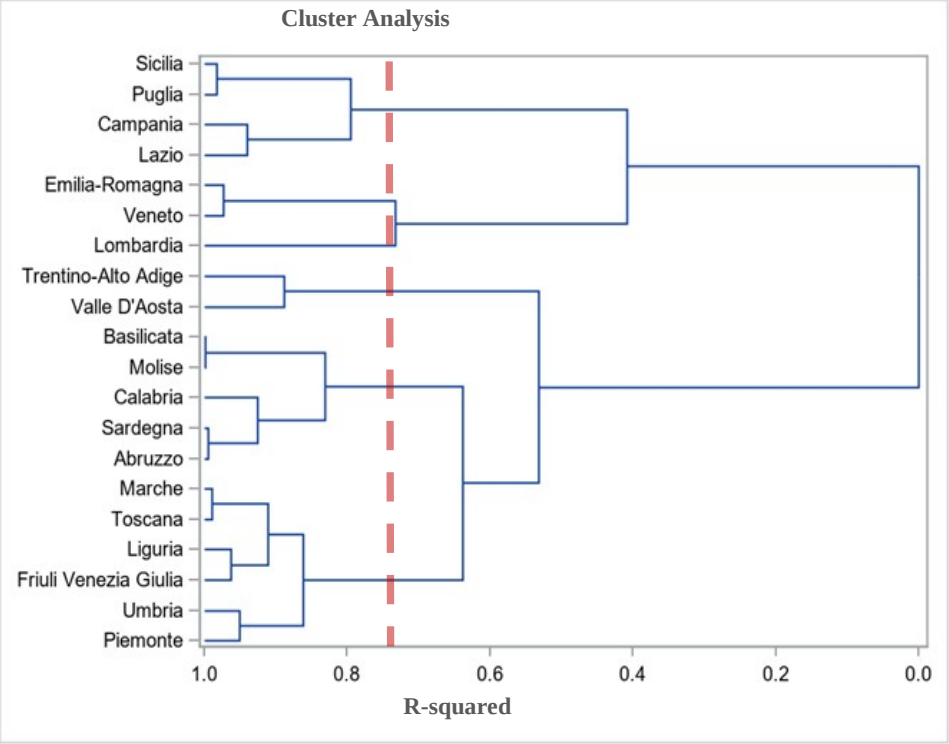
Cluster 3: Campania, Lazio, Puglia, Sicilia.

Cluster 4: Lombardia, Veneto, Emilia-Romagna.

Cluster 5: Valle D'Aosta, Trentino-Alto Adige.

R-squared equal to 0.73 indicates the validity of the chosen partition.

Figure 3 – *Dendogram.*



The following graph and table (Fig.4 and Tab. 3) illustrate their spatial distribution and composition.

Cluster 5, which includes Valle d'Aosta and Trentino-Alto Adige, is the one that is best positioned with a reduced environmental impact on the territories, while the number of companies that reduce the environmental impact of their activities and environmental satisfaction they are taller.

Cluster 1, which includes Abruzzo, Molise, Basilicata, Calabria and Sardegna, also has a good advantage situation with a high level of satisfaction.

In Clusters 2 and 4, which include the central-northern regions, consumption and environmental pressures are higher and there is a moderately positive perception of the environmental situation.

In Cluster 3, where the large southern regions are represented, environmental pressures are higher but the perception and number of companies that reduce the environmental impact of their activities are lower.

Figure 4 – *Representation of clusters on factors.*

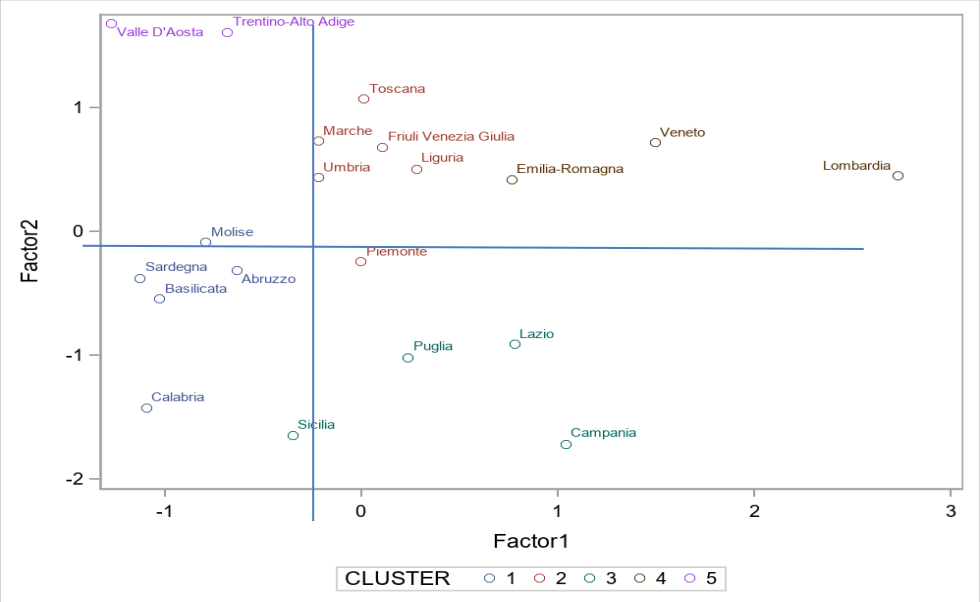


Table 3 – *Average values of variables calculated within each cluster.*

Cluster	DE	Final consumption of electricity	Consumed soil	Population density	Active companies that reduce the environmental impact of their activities	Environmental satisfaction
5	0,5	0,4	3%	58,7	16,1	86,9
1	0,8	0,4	4%	86,0	10,0	75,8
2	0,8	0,9	7%	170,9	13,0	78,7
3	1,1	1,0	8%	284,4	9,4	62,8
4	1,6	1,9	11%	294,4	13,3	71,8

4. A multidimensional measurement of the environmental phenomenon in Italy with synthetic index

The objective is to provide a synthetic tool the state of the environmental phenomenon starting from a set of selected indicators considered relevant in their impact with the territory and subsequently comparing performances at the regional level. Based on the latest data provided by Istat and Ispra on the environmental quality of our Italian regions, a set of elementary indicators were selected and combined (Tab. 4). Indicators present a rather pronounced variability and little correlation with each other, characteristics for the purposes of the objective being proposed. It's the basis for the aggregation process through the construction and comparability of some main composite methods.

Table 4 – *Environmental indicators selected*

Environmental issues	Environmental indicators			
	Polarity (+/-)	Indicators	Year	Source
Energy	(-)	a1. Final energy consumption (residents, industry and tertiary sector) *(GWh/km2)*	2021	ISPRA
Soil	(-)	a2. Soil consumption (*% of regional area)*	2021	ISPRA
Raw material	(-)	a3. Raw material taken (*1,000 tons per km^2)*	2020	ISTAT
Population	(-)	a4. Population density *(residents per km2)*	2021	ISTAT
Perception of environm. quality	(+)	a5. Environmental Satisfaction *(% residents aged 14 and over)*	2020	ISTAT
Companies in the environm. area	(+)	a6. Active companies with 3 or more employees that reduce the environmental impact of their activities *(n. companies per residents population)*	2018	ISTAT

Elementary indicators have been normalized and standardized to obtain data purified from units of measurement and comparison process. Standardized deviation in the composite index allows the construction of a robust measure and not very sensitive to remove a single elementary index (Mazziotta M. and Pareto A., 2013).

In addition, *polarity* (positive or negative) of the relationship between indicator and phenomenon was specified.

Finally, standardized indicators were aggregated. The choice of aggregation process follows the Mazziotta-Pareto Index (MPI) methodology as the MPI penalises the simple average with horizontal variability. This approach is of absolute importance when dealing with the multidimensional phenomenon; combination of domains that must be as homogeneous as possible. Following, steps to calculate composite index by comparing the following methods.

Given the matrix $X=\{x_{ij}\}$ with n rows (units) and *m* columns (indicators), composite methods have the following mathematical properties:

Adjusted MPI *(AMPI)*

$$MPI_{ci}^{\pm} = M_{ri} \pm S_{r_i} cv_i$$

$$\text{con } r_{ij} = \begin{cases} \frac{(x_{ij} - Min_{xj})}{(Max_{xj} - Min_{xj})} 60 + 70 \text{ if the indicator } j \text{ has positive polarity} \\ \frac{(Max_{xj} - x_{ij})}{(Max_{xj} - Min_{xj})} 60 + 70 \text{ if the indicator } j \text{ has negative polarity} \end{cases}$$

$$M_{r_i} = \frac{\sum_{j=1}^{m} r_{ij}}{m} \qquad S_{r_i} = \sqrt{\frac{\sum_{j=1}^{m} (r_{ij} - M_{ri})^2}{m}} \qquad cv_i = \frac{S_{ri}}{M_{r_i}}$$

The *AMPI* is a non-compensatory (or partially compensatory) composite index and allows min-max standardization of the indicators j and aggregation with the arithmetic mean penalized by the "horizontal" variability of the indicators themselves. Normalized values are approximately in the range (70; 130), where 100 is the reference value[2]. From the exploratory data analysis, indicators show a pronounced variability and little correlated with each other, except values corresponding to a1, a2 and a4, characteristics suitable to achieve the aims (Tab. 5).

Figure 5 shows cartograms of MPI^{+} and MPI^{-} method calculated in the Italian regions. Result of the regional ranking with the *composite index* method is like groupings obtained with the *cluster* and *ACP analysis.*

[2] In the Bienaymé-Cebycev theorem, terms of the distribution within the interval (70; 130) constitute at least 89 percent of the total terms of the distribution.

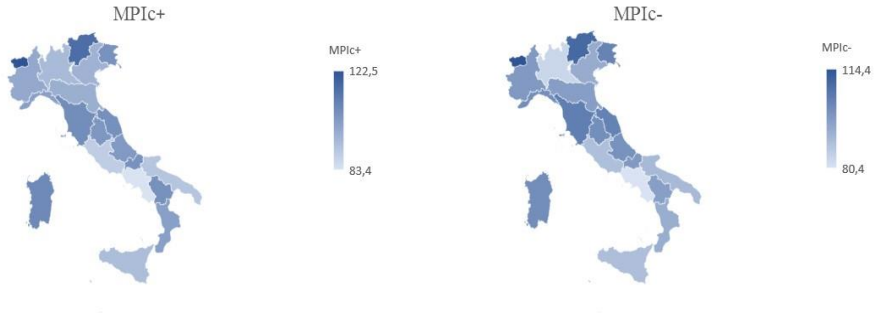
Table 5 – *Average and variability measures of environmental indicators.*

	a1	a2	a3	a4	a5	a6
M	0,90	0,06	0,97	179,66	64,59	11,87
S.q.m.	0,59	0,04	0,62	109,68	47,91	8,48
C.V.	0,66	0,68	0,64	0,61	0,74	0,71

Source: Istat and Ispra data processed

Ranking, based on the assumed values of the two synthetic indicators, shows the positioning of Italian regions according to the state of environment health that decreases as one moves towards the higher ranks (Tab. 6). The positioning, in terms of pressure on the environment for the issues considered, distinguishes Valle D'Aosta and Trentino Alto Adige as the region with the best perception environmental performance. Instead, Sicilia, Puglia, Lazio and Campania the environmental perception is not positive although consumption and environmental pressures are not high. Also noteworthy is the positioning in the lower ranks of Emilia-Romagna, Veneto and Lombardia where environmental consumption and pressures are high and the perception in terms of household satisfaction is rather negative.

Figure 5 – *Map of the MPI$^+$ and MPI$^-$ composite indices.*



Source: Istat and Ispra data processed

Table 6 – *Regional ranking between MPI⁺ and MPI⁻ composite indices.*

Region	MPIc+		MPIc-		Region	MPIc+		MPIc-	
	N.	Rank	N.	Rank		N.	Rank	N.	Rank
Valle D'Aosta	122,5	1	114,4	1	Liguria	101,9	11	100,5	8
Trentino-A. A.	115,7	2	109,6	2	Calabria	101,8	12	93,2	15
Sardegna	107,8	3	100,7	6	Piemonte	99,2	13	98,1	10
Toscana	106,7	4	104,4	3	Emilia-R.	97,8	14	97,0	12
Marche	106,0	5	103,5	4	Veneto	96,7	15	93,5	14
Basilicata	105,7	6	96,8	13	Lombardia	94,5	16	83,6	19
Friuli-V. G.	105,6	7	103,3	5	Sicilia	93,7	17	89,3	17
Molise	105,6	8	98,1	11	Puglia	91,5	18	90,3	16
Umbria	104,1	9	100,6	7	Lazio	89,3	19	88,9	18
Abruzzo	103,5	10	99,7	9	Campania	83,4	20	80,4	20

Source: Istat and Ispra data processed

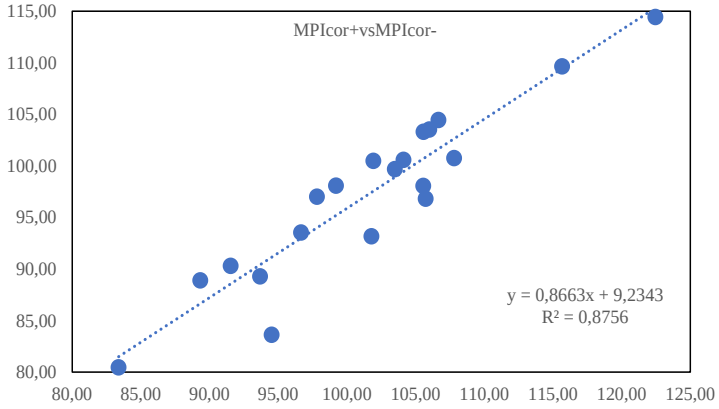
Table 7 shows rank differences compared by means of the absolute difference and Spearman's rank correlation coefficient. Sensitivity analysis shows similar results in the comparison between MPI⁺ and MPI⁻ method with absolute average rank differences 1.90 positions respectively with a strength of the relationship directly proportional and close to 1 (0.91). Linear relation is also high with R^2 equal to 0.87 (Fig. 6).

Table 7 – *Sum of ranking differences between MPI⁺ and MPI⁻ composite indices*

Measures	MPIc+-MPIc-
Absolute average rank difference	1,90
Cograduation index ρ	0,91

Source: Istat and Ispra data processed

Figure 6 – *Linear relationship and R^2 between MPI methods.*



Source: Istat and Ispra data processed

5. Conclusions

The study on environmental aspects of Italian regions through the comparison of multidimensional *acp*, *cluster* and *compositive methods* led to a convergence judgment of results and offer an important contribution to the interpretation of the phenomenon.

The geography of environmental status, anthropogenic pressure and perception once again shows a unbalanced configuration to the detriment of most of the central and southern regions. At the other end of the distribution, greater satisfaction and environmental performance are recorded in the North where investments in development projects and awareness of environmental issues are growing.

A good result for those who carry out a study of the phenomenon and must give an interpretation that is as representative as possible of the environmental reality.

References

DELVECCHIO, F. 2010. *Statistica per l'analisi di dati multidimensionali*. Cooperativa Libraria Editrice Università di Padova (CLEUP).

ISTAT. 2020. *Dati ambientali nelle città. Qualità dell'ambiente urbano*. Statistiche Report. I.Stat. Roma.

LECARDANE G. 2022. *Environmental monitoring in Italian cities. A multidimensional analysis with composite indices.* Rivista Italiana di Economia Demografia e Statistica, Sieds. Vol. LXXVI n. 4, pp. 109-119. Roma.

LECARDANE G. 2013. *Lo Stato dell'Ambiente. Misure statistiche e Indicatori di qualità ambientale*. E-BOOK ISTAT. Roma.

MAZZIOTTA, M., PARETO, A. 2013. *A Non-compensatory Composite Index for Measuring. Well-being over Time*. Cogito. Multidisciplinary Research Journal, Vol. V, 4.

OECD. 2008. *Handbook on Constructing Composite Indicators*. OECD Publications. Paris

---

Giuseppe LECARDANE, Istat, lecardan@istat.it
Flora FULLONE, Istat, ffullone@istat.it
Monica CARBONARA, Istat, mocarbon@istat.it

SOCIETÀ E RIVISTA ADERENTI AL SISTEMA ISDS
ISSN ASSEGNATO: 0035-6832

Direttore Responsabile: CHIARA GIGLIARANO

Iscrizione della Rivista al Tribunale di Roma del 5 dicembre 1950 N. 1864

Associazione all'Unione Stampa Periodica Italiana

TRIMESTRALE

La copertina è stata ideata e realizzata da Pardini, Apostoli, Maggi p.a.m.@tin.it – *Roma*

Stampato da CLEUP sc
"Coop. Libraria Editrice Università di Padova"
Via G. Belzoni, 118/3 – Padova (Tel. 049/650261)
www.cleup.it

ATTIVITÀ DELLA SOCIETÀ

A) RIUNIONI SCIENTIFICHE

XXXVII	La mobilità dei fattori produttivi nell'area del Mediterraneo (Palermo, 15-17 giugno 2000).
XXXVIII	Qualità dell'informazione statistica e strategie di programmazione a livello locale (Arcavacata di Rende, 10-12 maggio 2001).
XXXIX	L'Europa in trasformazione (Siena, 20-22 maggio 2002).
XL	Implicazioni demografiche, economiche e sociali dello sviluppo sostenibile (Bari, 15-17 maggio 2003).
XLI	Sviluppo economico e sociale e ulteriori ampliamenti dell'Unione Europea (Torino, 20-22 maggio 2004).
XLII	Sistemi urbani e riorganizzazione del territorio (Lucca, 19-21 maggio 2005).
XLIII	Mobilità delle risorse nel bacino del Mediterraneo e globalizzazione (Palermo, 25-27 maggio 2006).
XLIV	Impresa, lavoro e territorio nel quadro dei processi di localizzazione e trasformazione economica (Teramo 24-26 maggio 2007).
XLV	Geopolitica del Mediterraneo (Bari, 29-31 maggio 2008).
XLVI	Povertà ed esclusione sociale (Firenze 28-30 maggio 2009).
XLVII	Un mondo in movimento: approccio multidisciplinare ai fenomeni migratori (Milano 27-29 maggio 2010).
XLVIII	150 anni di Statistica per lo sviluppo del territorio: 1861-2011. (Roma 26-28 maggio 2011).
XLIX	Mobilità e sviluppo: il ruolo del turismo. (San Benedetto del Tronto, 24-26 maggio 2012).
L	Trasformazioni economiche e sociali agli inizi del terzo millennio: analisi e prospettive (Università Europea di Roma, 29-31 maggio 2013).
LI	Popolazione, sviluppo e ambiente: il caso del Mediterraneo (Università Federico II di Napoli, 29-31 maggio 2014).
LII	Le dinamiche economiche e sociali in tempo di crisi (Università Politecnica delle Marche, 28-30 maggio 2015).
LIII	Mutamento economico e tendenze socio-demografiche tra sfide e opportunità (Università degli Studi Internazionali di Roma, 26-28 maggio 2016).
LIV	Mobilità territoriale, sociale ed economica: modelli e metodi di analisi (Università degli Studi Internazionali di Catania, 25-26 maggio 2017).
LV	Coesione sociale, welfare e sviluppo equo e sostenibile (Università degli Studi dell'Insubria, Varese 24-25 maggio 2018).
LVI	Benessere e Territorio: Metodi e Strategie (Università Politecnica delle Marche, Ascoli Piceno 23-24 maggio 2019).

LVIII Tra marginalità e sviluppo. La sfida della sostenibilità in una prospettiva mediterranea (Università LUMSA, Palermo, 26-27 maggio 2022).

LIX Aspetti economici e sociali dell'invecchiamento demografico (Università degli Studi di Napoli Federico II Napoli, 25-26 maggio 2023).