IASSC Technical Reports

Quality Profile

Questionnaires, Fieldwork, and Data Preparation

ITALIAN LIVES

Mario Lucchini, Gianluca Argentin, Davide Bussi, David Consolazio, Giovanna De Santis
Tiziano Gerosa, Giovanni Guidi, Serafino Negrelli, Carlotta Piazzoni, Maurizio Pisati
Chiara Respi, Egidio Riva, Emanuela Sala, Domingo Scisci, Marco Terraneo

IASSC
INSTITUTE FOR ADVANCED STUDY OF SOCIAL CHANGE

IASSC Technical Reports No. 1
April 2023

Quality Profile

Questionnaires, Fieldwork, and Data Preparation

Mario Lucchini, Gianluca Argentin, Davide Bussi, David Consolazio, Giovanna De Santis
Tiziano Gerosa, Giovanni Guidi, Serafino Negrelli, Carlotta Piazzoni, Maurizio Pisati
Chiara Respi, Egidio Riva, Emanuela Sala, Domingo Scisci, Marco Terraneo

IASSC
INSTITUTE FOR ADVANCED STUDY
OF SOCIAL CHANGE

The **Institute for Advanced Study of Social Change (IASSC)** is a permanent observatory on social change based in the Department of Sociology and Social Research of the University of Milan-Bicocca (Italy).

The **IASSC Technical Reports** series presents technical and scientific information from projects of the Institute for Advanced Study of Social Change.

Please cite this publication as:
Lucchini, M., Argentin, G., Bussi, D., Consolazio, D., De Santis, G., Gerosa, T., Guidi, G., Negrelli, S., Piazzoni, C., Pisati, M., Respi, C., Riva, E., Sala, E., Scisci, D., Terraneo, M. (2023) *Quality Profile: Questionnaires, Fieldwork, and Data Preparation*. Milano: Institute for Advanced Study of Social Change.

Copyright © 2023 Mario Lucchini, Gianluca Argentin, Davide Bussi, David Consolazio, Giovanna De Santis, Tiziano Gerosa, Giovanni Guidi, Serafino Negrelli, Carlotta Piazzoni, Maurizio Pisati, Chiara Respi, Egidio Riva, Emanuela Sala, Domingo Scisci, Marco Terraneo

Published by the Institute for Advanced Study of Social Change (IASSC), Via Bicocca degli Arcimboldi 8, 20126 Milano, Italy, https://iassc.unimib.it/en.

Typeset in LaTeX using the su□esi class by Ivan Valbusa.

ISBN (PDF): 979-12-80999-01-6

contents

1

introduction

The Italian Lives (ITA.LI) is a survey, namely a systematic way for gathering information by asking people questions from a sample of households and individuals for the purposes of constructing descriptive and analytic statistics as regards the attributes of the population. Data coming from ITA.LI can be used to understand Italian society and to test theories of behaviour and social change.

ITA.LI is promoted by the *Institute for Advanced Study of Social Change* (IASSC), the permanent observatory on social change in Italy at the Department of Sociology and Social Research of the University of Milano-Bicocca. ITA.LI takes its inspiration from the *Italian Longitudinal Household Panel (Indagine Longitudinale sulle Famiglie Italiane - ILFI)*[1] , a panel study that began in 1997 and ended in 2005. Likewise ILFI, ITA.LI aims to collect a set of basic information on the current conditions of the sampled subjects and households (composition, demographic characteristics of members, perceptions, attitudes and behaviours related to daily life) and to study social change in terms of residential mobility, schooling, working career, forms of family cohabitation and marriage. Such phenomena can be studied from life course data, that is, by reconstructing individual-level trajectories in their entirety. The information collected along the time axis, in the form of “spells” and “repeated occasions”, allows the passage from a static image to a dynamic representation of the phenomena, as well as the application of analytical strategies with a high investigative power.

1 https://periodicounitn.unitn.it/periodicounitn.unitn.it/archive/periodicounitn/numero16/indagine.html

ITA.LI's survey design allows researchers to study the timing of key life cycle events (e.g., formation of a new family, birth of a child, job loss, career progression, retirement from the workforce, onset of a disease, etc.). These events are the result of processes and take the form of transitions, status changes, rites of passage.

Like other panel studies, ITA.LI can be metaphorically described as a collection of family photo albums where each photo points to something interesting. However, a deeper understanding is achieved by examining the sequence over time of the images and the interactions between the people in them (Rose *et al.* 1994). Scrolling through the album, the story that emerges is more than the sum of the individual photos.

The richness of panel data, together with the techniques employed for life course analysis, offer the possibility to describe individual trajectories of change within specific spatial and temporal contexts; to approximate estimates of causal effects while controlling for unobserved heterogeneity; to estimate the time it takes before an event occurs and the factors responsible for it; to discern the cohort effect from the age effect and the period effect; to test for interdependencies between events that belong to different domains; and to study in depth the interactions and mutual influences between the life courses of members of the same household. Therefore, ITA.LI constitutes a rich database that is particularly suitable for policy-oriented studies. The opportunities for research and in-depth study offered by ITA.LI are particularly relevant in a country like Italy, which currently lacks surveys of this type. The ITA.LI survey is a valuable tool for planning and monitoring social policy interventions. The research design makes it possible to detect the inter and intragenerational social mobility, the risks of poverty and vulnerability of families, the difficulties encountered by individuals in school learning and job placement, delays in the transition to adulthood, and ageing-related risks. The objective of this study is to recruit a dynamic, self-regenerating sample, i.e., to collect longitudinal information about the initial sample members, their children, and those who live with the original members. Collecting information of this kind requires a substantial organisational effort and financial resources. Another critical issue is that sample members and households are subject to mortality and attrition. Finally, compared to standard approaches, the analysis of individual change requires greater expertise in building the database, implementing the analysis models, and substantively interpreting the estimated parameters.

The reconstruction of life courses is the best way to capture the processual and multidimensional aspect of phenomena, to conceptualise biographical paths as the outcome of the interaction between micro, meso and macro dynamics. Moreover, panel data allow the application of analytical strategies through which causal inferences can be drawn. The presence of time-varying variables makes it possible to implement special regression models that can eliminate the effects of unobservable characteristics that remain stable over time (such as innate ability, willpower, genetic endowment, etc.). Such confounders are a source of concern in observational studies, as they introduce bias in the estimates of causal effects.

Panel surveys with retrospective and prospective modules are a valuable tool for "looking back", that is, for gathering information about accumulated experience and capturing potential explanatory factors for events that occur along the life course (Butz and Torrey 2006). For example, the birth of a child can lead to a woman's withdrawal from the workforce, just as a previous experience of unemployment can increase the risk of being trapped in new unemployment events (path dependence); the break-up of a marital relationship can have negative consequences on the mental health of the children; a deterioration in the physical health of a parent can lead to a risk of social exclusion of their family.

Finally, ITA.LI has been designed taking into account the most recent methodological innovations adopted by the main international panels. The data collected are usable by a wide audience of experts working in different disciplinary fields from all over the world.

The ITA.LI database is indeed part of the *Cross National Equivalent File* (CNEF), an international project coordinated by Ohio State University that currently contains harmonised data from the following panels: *Understanding Society* (US), *Household Income and Labor Dynamics in Australia* (HILDA), *Korea Labor and Income Panel Study* (KLIPS), *Panel Study of Income Dynamics* (PSID), *Russia Longitudinal Monitoring Survey* (RLMS-HSE), *Swiss Household Panel* (SHP), *Canadian Survey of Labor and Income Dynamics* (SLID), and *German Socio-Economic Panel* (SOEP).

Many variables contained in ITA.LI appear to be harmonised with those of other panels (especially ILFI), thus facilitating comparative analysis. The availability of harmonised information across different panels makes it possible, on the one hand, to highlight differences and commonalities in educational, employment and family processes and, on the other, to capture the extent to which these dynamics are attributable to local traditions, national policies or macroeconomic factors.

The ITA.LI Research Group (henceforth "RG"), coordinated by Prof. Mario

Lucchini and Prof. Maurizio Pisati, is responsible for the research design and the entire organisational management of the survey. Its members collaborate in planning and coordinating data collection and management activities, dissemination of research products, methodological innovation, and obtaining new funding. The survey agency (henceforth “SA”) in charge of the first wave fieldwork was IPSOS, while in the second wave was Doxa. Data documentation, archiving and publication were managed by UniData - Bicocca Data Archive, Interdepartmental Centre of the University of Milano-Bicocca.

2

the life cycle of ita.li from a quality perspective

Italian Lives (ITA.LI) is part of a longitudinal research project carried out by the Department of Sociology and Social Research of the University of Milan-Bicocca, within the scope of the *Departments of Excellence* project (Italian Law 232 of 11 December 2016).

This document illustrates the procedures inspired by the principles of the survey methodology that have been applied to carry out the first wave of data collection of ITA.LI. RG members have made important decisions in regard to the definition of the sample design, the approach to contact household sampled, the evaluation and testing of the questionnaire and the mode of administration, the training and supervising interviewers, the control for accuracy and internal consistency of the data files, the approaches used to adjust the survey estimates. Each of these decisions taken has cost implications and the potential to affect the quality of the survey statistics. "Quality" is defined within the framework of the total survey error paradigm (Groves *et al.* 2011).

There are two types of inference involved in sample surveys, one from the questions to constructs, the other from the sample statistics to the population statistics (Groves *et al.* 2011). Hence, it is crucial to obtain questions whose answers mirror the constructs, as well as to identify and measure sample units representative of the target population. Notwithstanding the efforts made, such processes are subject to an unavoidable degree of imperfection, leading to the production of survey errors in survey statistics. Validity errors involves the gap between the measures and the construct; measurement errors arise during the application of the measures; editing and processing errors can arise during the phase of data preparation for statistical analysis; coverage errors arise when enumerating the target population using a sampling frame; sampling errors derive from surveys involving only a subset of the frame

population; non-response error arise from the impossibility to measure all sample persons on all the measures; adjustment errors arise in the construction of statistical estimators to describe the full target population. All the types of error described have the potential to distort the results obtained from the survey and need to be taken into account thoroughly to minimise bias in survey statistics. A graphical representation of the phases in which survey errors can occur during the survey life cycle is available in Figure 2.1

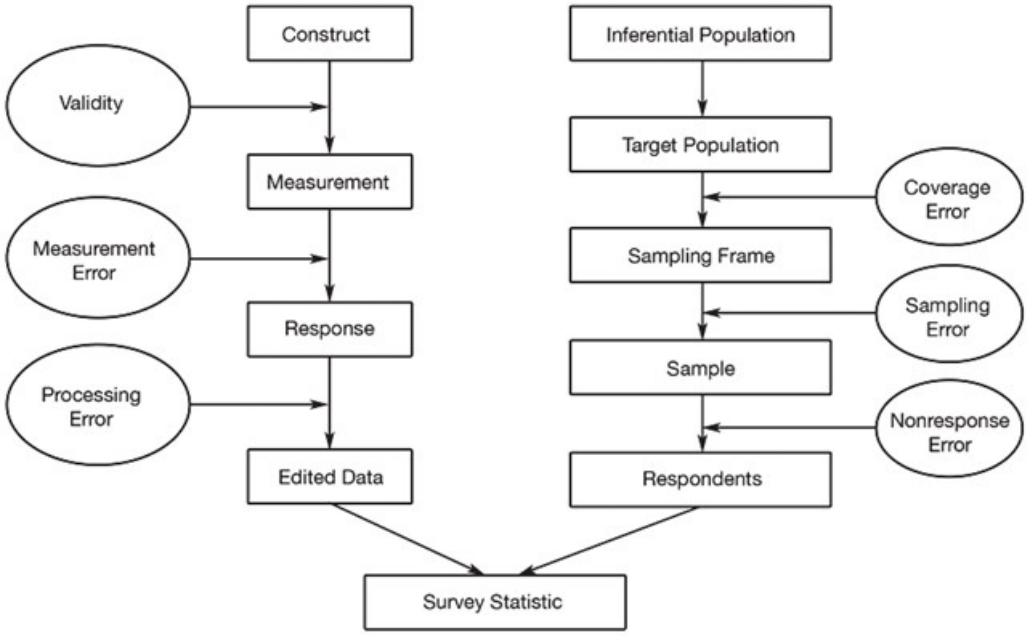

Figure 2.1 Survey life cycle from a quality perspective (Groves *et al.* 2011).

The study consists of a panel survey that involves approximately 5,000 households selected from 278 Italian municipalities using a probabilistic sampling method explained in detail by Pisati (2023, p. 1-24). Its aim is to build a constantly updated dynamic database on social change in Italy o□ering high-quality data to researchers working in several disciplinary fields.

The data are harmonised with those from leading research carried out internationally, in order to make them available to a vast audience of experts interested in cross-national comparative research.

The first wave of ITA.LI, which is covered in this contribution, records the entire life history of its sample members up to the time of the interview, in relation to the following themes: residential mobility, education, working career, marriage or cohabitation, birth or adoption of children. In addition, each individual interview collects repeated information (i.e., gathered in subsequent waves) about perceptions and habits concerning health, quality of life,

deprivation, well-being, resources and debts, household economic supports, Internet use, personality traits, and political participation. A multistage sampling design was developed in collaboration with the *Italian National Institute of Statistics* (Istat)[1], which encompasses a sequential selection of municipalities, addresses, households and individuals (Pisati 2023, p. 1).

The RG began working on the ITA.LI project in 2018, focusing on the survey design, the drafting of the questionnaires and the interviewers' training. Data collection started one year later, in June 2019, and stretched across 20 months, exceeding the timing expected due to delays that occurred as a consequence of the COVID-19 pandemic. The interviews were primarily conducted face-to-face, but alternative data collection methods, specifically telephone interviews, were additionally carried out in order to deal with the advent of the pandemic and maximise survey participation (DeLeeuw 2018).

2.1. *General Population Sample*

The sample strategy adopted in ITA.LI has been described in detail elsewhere by Pisati (2023), and it involved a proportionately stratified clustered sample of addresses. This is a novel selection procedure which had never been used before in the Italian context for national-level surveys: the interviewers, unlike other similar surveys, do not have a direct access to the names of the households to be contacted; instead, they must use a standard protocol to select a household and their members from the address assigned to them. The first stage was to select a sample of 280 municipalities as the primary sampling units (PSUs). Municipalities were selected with probability proportional to the number of residential addresses. The second stage was to select addresses within each sampled municipality from the individual registry.

The number of selected addresses is greater than the one defined in the theoretical sample design to allow the replacement of unusable addresses.

2.2. *Household Selection*

As illustrated in detail elsewhere (Pisati 2023), the probability sample design adopted in the study includes a first stage of random selection conducted on Italian municipalities stratified by region, degree of urbanisation and population size. Within each selected municipality, a random sample of residential addresses was chosen (second stage) and a single household was picked at

1 Claudia de Vitiis and Stefano Falorsi for Istat and Maurizio Pisati for the University of Milano-Bicocca

random from each selected address (third stage). Finally, at the fourth stage, all household members aged 16 and over were considered eligible for the interview.

Household selection takes place at the third sampling stage, which actively involves interviewers in a highly standardised procedure of field listing (Kalton *et al.* 2014). At this stage, each interviewer is assigned a list of sampled addresses and a personal screen media device (tablet) to conduct both the household selection and data collection activities.

Once a new address is reached, the interviewer must verify whether it actually is an accessible residential address or not. An address can be indeed considered eligible only when the interviewer finds a place which could potentially be inhabited (e.g., an uninhabited ruin cannot be considered eligible). Moreover, the address has to be physically reachable without putting the interviewer in danger.

Finding an accessible residential address which could potentially be inhabited by a household places the interviewer in front of two alternative scenarios. The former consists of a single dwelling (e.g., detached house). In this case, no further steps are required, and the interviewer can proceed by contacting the household living there (see paragraph 5.1.1). The latter scenario includes several dwellings at the selected address (eg. semi-detached houses or flats). In this case, the interviewer must randomly select one dwelling in order to proceed with household contact and interviews.

At this stage, the interviewer has to count the overall number of units (e.g., buildings, dwellings, doorbells etc.) located at the address and associate each unit to a progressive number following a standardised procedure (i.e., from top to bottom and from left to right). Once finished, the interviewer asks the data collection software in the tablet to randomly extract a number between 1 and the total number of units. The extracted unit is the one associated with the number proposed by the software.

Based on both the high heterogeneity of housing building types in the Italian context (European Commission and Eurostat 2018) and the need to keep a high degree of standardisation in the sampling procedure, the interviewers were given precise instructions enriched with practical examples on how to proceed at different unit levels.

This procedure should be repeated every time the interviewer comes across multiple units to choose from at a single selection level. Despite being time consuming, this standardised procedure allows the interviewers to account for a variety of situations by simply following the instructions that appear on their tablet.

The realized household sample size in ITA.LI is made of 4,900 units, very

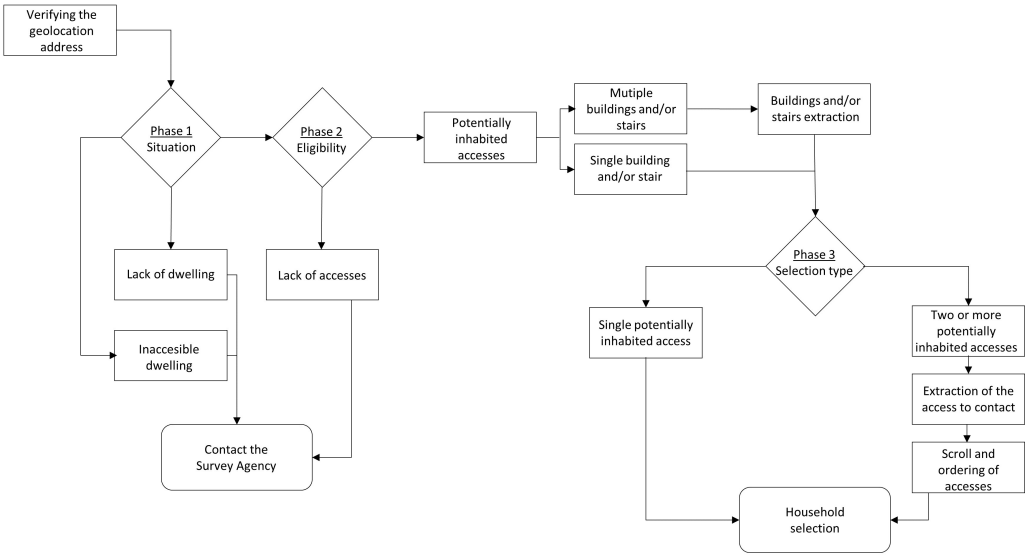

Figure 2.2 Steps leading from sampled address to household selection.

close to the target (5,001). To achieve this result, it was necessary to release 14,264 residential addresses, corresponding to the 37% of the designated sample of addresses (second stage). 13″078 households were selected from the eligible residential addresses, the 79‟6% of which were classified as eligible (10″402). A total of 11″389 individuals were found to belong to the 4″900 responding households, 8″967 of which were e□ectively interviewed, corresponding to a within-household individual response rate of 89%. Most interviews were administered to self-respondents, but a small proportion (2‟1%) were answered by proxy respondents (Pisati 2023, p. 12).

2.3. *Sample Status*

The individuals interviewed are assigned to one of three possible sample statuses:

- Original Sample Member (OSM)
- Temporary Sample Member (TSM)
- Permanent Sample Member (PSM)

OSM are full sample members, whilst the other two statuses are assigned to subjects which do not belong themselves to the sample, but from whom data should be collected given their relationship to one (or more) sample members. Therefore, they need to be included in the fieldwork though not being members of the statistical sample for longitudinal analysis. Rather, the

data they provide can be considered as an attribute of one or more sample members. A more detailed definition of each status is provided below.

2.3.1 Original Sample Member (OSMs)

OSMs are defined as all members of the households sampled at Wave 1, including temporarily absent members living in institutions (e.g., hall of residence, boarding schools). Any child born from an OSM mother and/or father after Wave 1 and observed to be resident with one or both parents at the survey wave following the child's birth is considered to be an OSM. In the case of households composed of a single child, even if not in interview age, if the child moves, his/her co-habitants are eligible for interview; if instead the child moves towards an institution a "split-o household" is created, composed exclusively of the child. From Wave 2 onwards, OSMs of all ages are followed for interview and remain eligible for survey participation as long as they are resident within Italy, potentially for the life of the survey. If an OSM moves house, he/she is followed to his/her new address and those living with the OSM become eligible for interview as TSMs. If, instead, the OSM moves into an institution, just him/her will be enumerated and interviewed, and not other residents of the institution.

2.3.2 Temporary Sample Member (TSMs)

At each survey wave, all members of the household of an OSM who are not themselves OSMs are designated TSMs. TSMs remain eligible for enumeration and interview as long as they are resident in a household that includes at least one OSM or PSM (see below). When a TSM is no longer co-resident with an OSM or PSM, they are not followed and become ineligible for interview. Former TSMs are identified as "re-joiners" if they are subsequently found to cohabit with one OSM or PSM, becoming eligible again.

2.3.3 Permanent Sample Member (PSMs)

TSMs may change to PSMs according to specific circumstances (e.g., substantive research reasons because of the additional contextual information these people may provide for the analysis of OSMs) indicating that attempts should continue to be made to enumerate and interview them, even if and when they no longer live with an OSM. Note that some PSMs will have been enumerated at previous waves with the status of TSM, while others will be enumerated for the first time as a PSM. PSMs remain potentially eligible for enumeration

and interview for the life of survey

2.4. *Following Rules*

In ITA.LI, all persons aged 16 and over who live in a sampled household are considered eligible for the interview. A household is made up of all members residing in the dwelling selected at the tertiary sampling stage.

A household can be then made up of a single person, a family unit or a group of cohabiting individuals without interpersonal (parental or a□ective) ties. This may be the case, for example, of family service personnel (domestic support workers, family workers, caregivers, etc.) who routinely reside in the home in which they serve. In this case, they are considered part of the employers' household as another cohabitant person without romantic ties or ties of kinship or a□nity.

The household also includes those members who, throughout the duration of the survey, have been temporarily absent for valid reasons, such as: military service or voluntary civil service; temporary hospitalisation in health institutions of any kind (hospitals, clinics, etc.); business; o□-site assignments, including attendance at qualifying or refresher courses; seasonal employment; tourism; boarding a navy or merchant marine vessel; detention pending trial. On the other hand, persons temporarily absent from another household are not considered as part of it, as, for example, children in foster care.

A household is considered eligible upon the fulfilment of two specific conditions of both de-facto and legal residence. The condition of de-facto residence requires that at least one member of the household habitually lives in the dwelling, while the legal condition is based on the assumption that at least one member of the household is resident in the municipality where the dwelling is located. If even one of the previous two conditions is not met, the household cannot be considered eligible and, consequently, the address is excluded from the sample.

Household members are considered eligible if they are part of the household (as defined above) and are 16 years of age or older. Interviewers were asked to interview all eligible members in households consisting of 1 or 2 eligible members and at least 2 eligible members in households consisting of 3 or more eligible members. Failure to meet these thresholds results in the exclusion of the household from the final ITA.LI sample. Following Wave 1, from Wave 2 onwards beyond the households and the individuals interviewed in the first round, based on the Contact Sheet (variable "E03") the following household will be contacted:

- Eligible yet not contacted households, namely families with whom the interviewers were not able to establish a direct contact due to absence or unavailability during the first wave
- Eligible and "complete" households, namely families which met the "completeness" criterion, though for whom not all the eligible members were interviewed
- Eligible yet "dropped" households, namely families which did not meet the "completeness" criterion due to a missing contact of one or more members for temporary absence or illness

The household meeting the following criteria are instead excluded from a potential contact for subsequent waves:

- Non-eligible households
- Household for whom it was not possible to assess the eligibility status during Wave 1
- Eligible households which refused to be interviewed (without having ever withdrawn the refusal)

Therefore, as regards individuals, subsequent waves will involve:

- All the OSMs still living in Italy, including subjects never contacted before, subjects temporarily absent during the first wave now available, subjects who were ill at the time of Wave 1 and who are now able to carry out the interview
- All the TSMs, as long as they live with an OSM or a PSM
- Among the PSMs, TSMs fathers (i.e., partners of women OSMs) of children had with an OSM woman following Wave 1, living with such children at the time of Wave 2
- Adopted children of one or both parents in both heterosexual and homosexual couples if aged at least 16 years old

2.5. *Data Collection Methods*

Interviews with members of the selected households were conducted primarily by using the CAPI (Computer-Assisted Personal Interview) method. The decision to have experienced interviewers handing out the questionnaire in a face-to-face mode was primarily motivated by the length and complexity of the retrospective interview (approximately 60 minutes). In addition,

alternative survey methods were arranged to deal with specific cases and situations. The CATI (Computer-Assisted Telephone Interview) method was used with all household members temporarily absent from the house but reachable by telephone and, more generally, to address critical issues in carrying out face-to-face interviews. The CATI method was used more frequently during the COVID-19 pandemic, due to the consequent restrictions and containment measures that made it difficult (sometimes impossible) to reach the interviewers at their locations.

The mediated interview is instead employed for young people aged 16-24 reported by SA as being at high risk of refusal despite being members of a participating household. This type of interview involves carrying out the prospective part of the questionnaire to the participant using a CATI method. The administration of the retrospective part, on the other hand, is delegated to one of his/her parents using the CAPI method.

Finally, the proxy interview is intended for individuals unable to take part in the survey due to health conditions, disabilities, or because they were absent for the entire fieldwork. The latter condition does not include household members who are absent from home intermittently (e.g., workers away from home who return home on weekends) or for short periods (e.g., short vacations, temporary work commitments, etc.). The proxy questionnaire. consists of a shortened version of the main questionnaire which is administered to one of the household members who responds on behalf of the absent person.

2.6. *Survey Timeline*

The first wave of the survey began in June 2019, with the interviewers inspecting sampled addresses and making initial contacts with the households. The innovations introduced by the tertiary sampling stage on households' selection led the RG to avoid defining a priori the closing date of the activities related to data collection.

Another event also contributed to cast doubts on the closing times of the survey. The advent of the COVID-19 global health emergency, approximately seven months after the start of the survey, led both to a suspension of field activities and to a redefinition of contact and interview strategies. The use of the CATI method, initially planned only for some specific cases, was extended to all interviews conducted for the entire duration of the first lockdown (March-June 2020). At this stage, restrictions on spatial mobility and direct contacts were severe, making any form of face-to-face interaction between interviewers and household members impossible (e.g., Sjödin *et al.* 2020).

The easing of restrictions that took place in the following months (June 2020 onwards), allowed the SA to return to using the CAPI method, while continuing to use telephone interviews in higher risk areas of the country and, more generally, to meet the needs of both the households contacted and the interviewers. The adoption of this mixed approach ensured continuity in data collection activities, which ended in January 2021.

As shown in Figure 2.3, between February 2020 and June 2020, the number of interviews carried out was drastically reduced, then the fieldwork gradually resumed. The adoption of a mixed approach had then also a positive impact on the involvement of individuals within households. The chance to administer face-to-face, telephone, or mediated interviews allowed the RG to reduce the number of drop-outs due to failure to meet the complete household criteria.

Overall, 4"900 households participated in the ITA.LI survey, for a total of 8,967 individuals living in 278 municipalities distributed throughout the Italian regions. Only 189 (2"1%) interviews were conducted using the proxy questionnaire, while the remaining 8"658 household members (97"9%) were interviewed using the main questionnaire, including 120 mediated interviews.

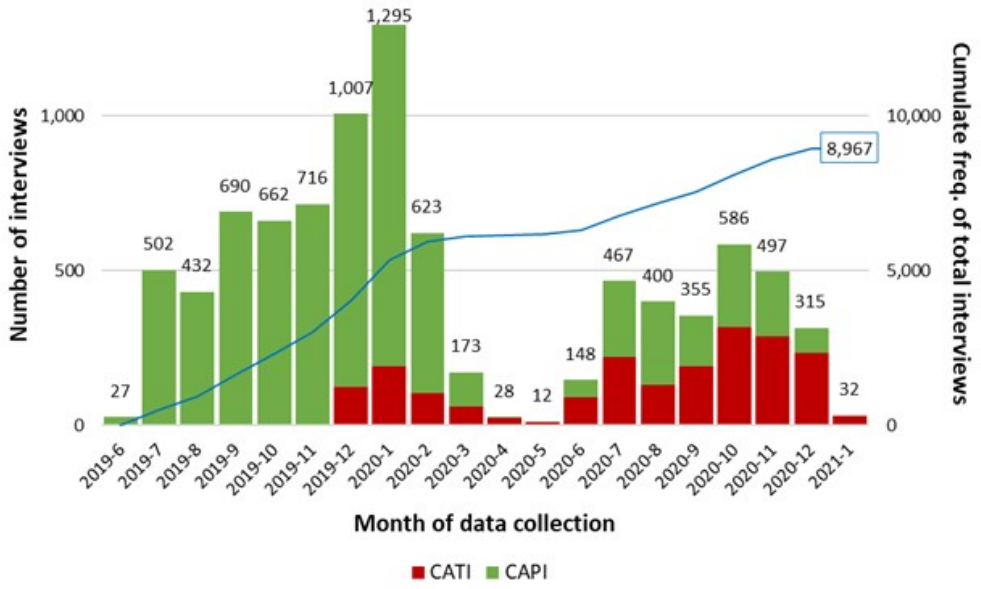

Figure 2.3 Interviews over time by survey method.

The CATI method was used in 22"6% of the interviews conducted via main questionnaire, for a total of 1"954 participants. Comparing socio-demographic characteristics between CAPI and CATI respondents no significant di erences emerge in relation to the method of data collection used. Finally, the signifi-

Table 2.1 Interviewees by number of eligible members in the household.

Number of eligible members	Number of interviewees					Total
	1	2	3	4	5	
1	1”413 (100)					1”413 (100)
2		2”286 (100)				2”286 (100)
3		497 (62“8)	294 (37“2)			791 (100)
4		207 (61“4)	61 (18“1)	69 (20“5)		337 (100)
5		28 (43“1)	11 (16“9)	5 (7“7)	21 (32“3)	65 (100)
6		5 (71“4)	2 (28“6)			7 (100)
7			1 (100)			1 (100)
Total	1”413 (28“9)	3”023 (61“7)	369 (7“5)	74 (1“5)	21 (0“4)	4”900 (100)

cance of the COVID-19 phenomenon prompted the RG to conduct an ad hoc study on the subsample of ITA.LI first wave respondents interviewed before the pandemic outbreak (March 2020). Data collection started in April 2020 and finished at the beginning of September 2020, involving a self-selected subsample of 904 ITA.LI respondents (Respi and Gerosa 2021). The data obtained from this ad hoc module were then linked to those derived from interviews conducted as part of the main survey with a limited loss of respondents, enriching the ITA.LI archive with longitudinal information collected before and after the spread of the pandemic (Gerosa and Respi 2022). The resulting dataset can be thus fruitfully exploited from a quasi-experimental perspective to assess the short-term impact of the pandemic on a wide range of psycho-social domains (e.g., Lucchini *et al.* 2021).

At the end of the survey, 13”078 families were contacted, of which 0“2% were found to be ineligible and 20“2% had unknown eligibility. Of the 10”402 eligible families, 4”900 (47“1%) were complete families, while the remaining 5”502 (52“9%) were not interviewed due to refusal to participate in the survey, drop-out, or unavailability. Out of a total of 10”080 eligible individuals, 8”967

(89“0%) participated in the survey, giving a main questionnaire interview (87“1%) or a proxy interview (1“9%), while individual refusals accounted for 10“4%. Finally, 0“5% of individuals started to answer an interview but did not complete it (partial interviews). The overall household Response Rate 1 (AAPOR 2016) of ITA.LI amounts to 37“1%. More details are presented in (Pisati 2023, p. 12-14).

2.7. *Item Non-Response*

To assess the completeness of the questionnaires filled and the quality of the answers received, the Item Non-response Rate (INR) was calculated, namely a variable-level index of non-response, which formula is:

$$\text{INR} = \frac{\text{Number of units with item non-response}}{\text{All eligible units exposed to the item}}$$

The index was calculated on the interviewees' sample from the main questionnaire, following two distinct procedures. The former allows defining a global measure synthesising the share of missing answers compulsory questions for each individual. Considering the 49 questions administered to all the respondents, which include the answer's options "I prefer not to answer", "I don't know" and/or "I don't remember", it was counted how many times such answers were given dividing by the number of respondents to which such questions were administered. The average of the values obtained by each ratio was 2“25 (SD=0“14).

The latter procedure, instead, focuses on the distribution of missing answers for a set of variables selected due to their relevance in relation to specific issues or with respect to the cognitive mechanisms involved in the answering process. Some have been chosen because of their association with potentially sensitive questions, that is those concerning income, political party (voted and that the respondent is willing to vote in the next election), body weight, self-perceived health, and rejection in the upper secondary training course. Others are of substantive interest, such as the property regime chosen for the couple's union, presence of children, current family situation, place of birth, attendance at kindergarten, occupational status for first job in life, life satisfaction, current non-employment status, public or private primary school attendance, marital status, and satisfaction with the current job. Finally, other variables refer to questions with a possible memory e□ect, such as the father's year of birth, the title of tenure of the house at birth and the mother's job. The INR of each of these variables is reported in descending order in Table 2.2.

Table 2.2 INR for variables of interest, descending order.

Variables	INR
Earned income in the previous month	56“31%
Political party voted for in previous elections	29“15%
Political party that would vote in the next election	25“00%
Body weight	16“16%
Father’s year of birth	13“59%
Matrimonial property regime	9“26%
Home tenure title at birth	3“16%
Had children	1“63%
Current household situation	1“59%
Mother worked when he or she was 14 y.o.	1“19%
Place of birth	1“08%
Attended kindergarten	0“94%
First job position	0“75%
Self-perceived health	0“55%
Overall life satisfaction	0“38%
High school rejection	0“33%
Current non-working condition	0“25%
Private or public primary education	0“24%
Marital status	0“18%
Current job satisfaction	0“15%

2.8. *Sample Representativeness*

In order for a sample survey to produce valid estimates of the quantities of interest, it is required that its realized sample be representative of the target population. One of the most common ways to achieve this goal is the benchmark comparison approach, which involves comparing the distribution of a variable among survey respondents with the corresponding distribution from a benchmark data source, possible derived from o□cial statistics (Groves 2006). A benchmark comparison approach was used to assess the representativeness of the ITA.LI realized sample with respect to some key variables, comparing a small set of standard sociodemographic variables between ITA.LI and the 2019 Italian Permanent Census of Population and Housing data. Some noteworthy deviations were observed in the educational attainment, where there is a slight under-representation of individuals with a maximum

of elementary school education and graduates, and in household composition, with an over-representation of families consisting of two individuals, while families composed of one member or four or more members appear to be under-represented. The results of the comparison are available in (Pisati 2023, p. 15-24), and highlighted the need to adjust the ITA.LI realized sample through appropriate weighting, despite the overall good representativeness of the sample.

3

questionnaire design

ITA.LI questionnaire was designed to fully exploit the potential of longitudinal research analysis, allowing the reconstruction of life courses, thus focusing on change at the individual level dynamics. More specifically, it consists of two main parts aimed at answering two alternative – though potentially interrelated – research objectives. The first part collects the socio-demographic characteristics of the respondents and includes the core themes for the reconstruction of the interviewees' biography from their birth to the time of the interview in relation to residential mobility, education, employment, marriages or cohabitations, birth and/or adoption of children. The second part, on the other hand, discusses cross-cutting themes – such as health, quality of life, quality of job, deprivation, personality traits, new media usage, and the like, thus taking a snapshot at the time of the interview of their attitudes toward relevant topics.

Such information is meant to be collected much more reliably in prospective panel design than in a long retrospective recall. Moreover, the fact that ITA.LI is a household panel allows the collection of data from and about all household members providing information on the family interdependencies and commonalities between individuals (e.g., income, poverty, material well-being, attitudes and values, etc.).

Finally, the main questionnaire is enriched by interviewers' observations reporting the level of understanding, accuracy and interest shown by the interviewee. Moreover, paradata are captured during the data collection process and include call records, date the interview took place, interviewer observations about the type of neighbourhood and accommodation, time stamps, and other data which can be used in constructing unit non-response adjustments but also as substantive variables in econometric analyses (Mudryk *et al.* 2001). There are also some questions about the interview and respondent to which the interviewer is called to answer such as the degree of cooperation

and distrust, and the presence of people who influenced the course of the interview.

Within the questionnaire writing process, all precautions were taken to avoid post-hoc rationalization and the contamination of memory.

3.1. *Questionnaire Structure and Content*

Within the Italian academic research community, a process of extensive consultation was carried out by the community over the data requirements for each area, priority measures critical for longitudinal analysis, and the appropriate balance between the different substantive areas within the questionnaire. This process led to a consultation conference held at the Department of Sociology and Social Research – University of Milano-Bicocca in February, 2019, with the specific aim of reviewing and discussing the conclusions and recommendations of all the advisory groups and making final recommendations for the content and design of the ITA.LI questionnaire. Following the conference, the design work on the questionnaires began, taking forward the recommendations received. Several designated scientific advisory committees were established who met regularly to oversee the content and conduct of the panel. "Topic Champions", experts in topic priority areas such as (1) income, wealth, consumption and expenditure, (2) health well-being and health-related behaviours, (3) employment, (4) education, (5) family, will be designated to consult academic and policy users in their areas of expertise to ensure content in each area keeps abreast of emerging agendas as well as effectively addressing key longitudinal research questions as new data sources and techniques make new approaches possible.

The entire RG was involved in the questionnaire writing process with the aim to include both questions focusing specifically on characteristics that are either expected to be subject to change and questions regarding factors affecting the likelihood of change.

The questionnaire includes mainly questions not specifically developed for ITA.LI. Indeed, all the questions were sourced from prestigious national and international studies. The most important source was represented by the Italian longitudinal analysis ILFI and to a lesser extent from other national and international longitudinal and repeated cross-sectional studies, such as *Understanding Society*, *German Socio-economic Panel Survey* (SOEP), the *Household, Income and Labour Dynamics in Australia* (HILDA), *Swiss Household Panel*, SHARE, *Labour Force Survey* (LFS) and *Multiscope - Aspects of Daily Lives* by ISTAT.

Specialised sub-groups were created, each of which edited a specific section of the questionnaire. Also other scholars, both internal and external to the IASSC, were included to offer additional support in the development of the questionnaire. Once completed, the first draft of the questionnaire was then shared, discussed and revised by the RG as a whole until a final version was reached. It was decided to include questions and standard instruments (such as SF-12, big five personality traits, etc.) from other studies in order to enable comparative research.

Some questions, such as the name of the school attended and/or the address of the workplace were included with the aim to perform a linkage to external data, though this information is not released in the public file and are available upon specific authorised request, due to respondents' anonymity preservation.

The survey questionnaire and materials are available in Italian and English languages, allowing bilingual interviewers to simply switch to the language of choice.

3.2. *Summary of Instruments*

ITA.li can be considered as an interdisciplinary and multipurpose survey, given the wide range of data collection instruments used which make it possible to carry out cutting-edge sociological analysis. Most of these instruments, drawn from its predecessor ILFI and from other household panels (such as US, SHP, SOEP, etc.), include very important features on key domains of life, i.e., health, income, employment, attitude and behaviours. Such information is displayed at different levels of a social hierarchy (with occasions nested within individuals, household, geographical and institutional contexts), making possible to study the impact of individual and contextual effects simultaneously, as well as the effect of cross-level interaction terms. A household questionnaire is administered to the Household Reference Person and a number of questions about the place of living, household resources and debt, benefits and supports, deprivation and well-being were worded with reference to this person. The household questionnaire took about an hour to complete on average.

The key instrument of the design survey is the individual questionnaire that is administered to each member of the household aged 16 or over. As will be illustrated, questions are arranged in modules that cover a broad range of topics (residential mobility, education, employment, unions, children, quality of life, internet, personality traits, health, household resources and debts, benefits and supports, political participation) and it takes around 60 minutes to complete on average.

Finally a "proxy schedule", that's to say a shortened version of the individual questionnaire is used to collect information about household members absent throughout the field period or unable to complete the interview themselves due to physical or mental health problems, that is administered to another member of the household who is able to give the answers on his behalf.

3.3. *Structure and Content*

The questionnaire consists of a total of 13 themed sections, structured as follows:

- Personal information
- Residential mobility
- Education
- Employment
- Family
- Quality of life
- Internet
- Deprivation and well-being
- Personality traits
- Health
- Household resources and debt
- Subsidies and supports
- Political participation

Some sections were administered to all eligible members of the household, while others were intended only to the Reference Person identified by the interviewer during the contact phase. Each section of the questionnaire can be made up of one or more sub-sections.

3.3.1 Questionnaire Structure

The interview for each adult lasts 60 minutes on average, with an additional short household level questionnaire for one individual in the household. The questionnaire includes three main components:

- Core questions repeated at each wave
- Varying component questions

- Interviewers' observations

Questions are arranged in topic modules and cover, among others, individual demographics, education and training, health and caring, current employment and earnings, values and opinions, environmental behaviours, transport, and parenting. About half of the questionnaire content is collected biennially, with additional modules collected at different intervals (e.g., media usage and personality traits)

The remainder at each wave included the varying component. The variable component was designed for:

- Questions which needed to be asked less frequently than core items
- New questions engendered by changing policy and research issues, as in the case of the COVID-19-related questions or religious believes
- Questions to elicit retrospective data on panel members' life history before the first interview

3.3.2 Main themes

The core part of the questionnaire is based on the reconstruction of respondents' "life histories" referring to the following key areas: residential mobility, education, employment and family. For these sections, data is collected in the form of spells.

Personal Information

This is the first section of the individual questionnaire and is handed out to all eligible household members. The purpose of this brief section is to collect information on respondents' sex, date of birth, place of birth, and marital status.

Residential Mobility

The purpose of the "Residential Mobility" section is to reconstruct, as accurately as possible, all the events that make up the residential history of each respondent. More specifically, a housing spell is defined by living continuously in the same dwelling for a period of at least one month. Starting from the house where the interviewee lived at the time of birth, all housing spells are then recorded, up to the current one (i.e., the municipality and the house where the interviewee lives at the time of the interview).

Table 3.1 Main themes: sections and sub-sections.

Sections	Sub-section
Residential mobility	
Education	Education spells
	Military/civil service
Employment	Current employment
	Work spells
	Spells of non-employment and/or interruption of employment
Family	Family of origin
	Current family: unions, children, care-giving

For each spell the following information is collected: the start and end month and year, the Italian municipality or foreign country where the dwelling is located, the tenure of the dwelling (ownership, rent, other title) and some more detailed information on the dwelling, such as the surface area in square metres. Periods in which the respondent lived in housing that cannot be defined as a house (e.g., trailer, RV, tent, container, shack, etc.) or in a group living facility (e.g., retirement home, military or university housing, hospital, orphanage, group home, prison, etc.) are also considered housing spells. For these spells, information is collected only on the start and end dates (month and year) and location.

Education

The "Education" section consists of two sub-sections. The first is called "Education spells" and aims to reconstruct as accurately as possible all the events defining the entire education career of each respondent. The "Military/civil service" sub-section is instead intended to describe any experience of the respondent with military conscription or replacement civil service, and it is linked to the Education spells because it usually takes place during school/university years.

Education spells. This sub-section is made of five ordered domains, aimed at capturing the major milestones in the respondent's educational career. Educational careers are those pursued within the institutional educational system and consist of the following levels:

- Preschool education, that includes daycare, typically up to age 3, and preschool (up to ISCED 0)
- Primary education (ISCED 1)
- Lower secondary education (ISCED 2)

- Upper secondary education (ISCED 3-4), that includes: high schools, technical schools, vocational schools, teacher training schools, etc.; vocational training courses leaving certificate of two-year, three-year or four-year duration run by the Regions or Autonomous Provinces of Trento and Bolzano. Such training courses typically take place in Vocational Training Centres (Centri Formazione Professionale - CFP) or in Vocational Education and Training (Istruzione e Formazione Professionale - IeFP), and do not give access to tertiary education
- Tertiary education (ISCED 5-7), that includes: university leading to a bachelor's degree (three- or four-year undergraduate programs, master's or PhD programs, etc.); tertiary courses leading to another qualification rather than a degree (university certificates, ITS, IFTS, etc.); post-graduate courses requiring a bachelor's degree to be attended (postgraduate degrees, first- and second-level master's degrees, PhDs, etc.)

Information is collected on all the educational spells experienced by the respondent. An educational spell begins with enrolment in a school, in a vocational training centre, or a university, including post-graduate pathways. For each spell, respondents are requested to provide information on course attendance, duration, contents, completion and final degree.

A spell is recorded even if the respondent has interrupted the course of study without completing it or if something relevant changed, such as the educational institution or its location. It is the case of a respondent moving from one school to another regardless of changes in the course of study. On the contrary, a change in the course of study within the same school (e.g., from a scientific high school to a linguistic high school) does not constitute a new spell. The questionnaire provides an ad hoc pathway for this case.

Military/Civil Service. This subsection records the eventual participation of the interviewee in compulsory military service (even in wartime) or in substitute, but still compulsory, civil service. Participation in the "voluntary national civil servic" should instead be reported in the "Employment" section.

Employment

The purpose of the "Employment" section is to reconstruct each respondent's employment career as accurately as possible. Employment history is defined as a sequence of work events and interruptions occurred between the time the respondent began working and the time of the interview. It is organised in the three separate subsections of current employment, work spells, and spells of non-employment/interruption of employment.

Current Employment. This subsection is designed to identify the current work position of the respondent (currently working; if not currently working, ever worked in the past; if not currently working, whether looking for work, etc.) and to capture some additional information about the characteristics of the current employment situation (e.g., working hours per week, job quality, earnings).

Work spells. The goal of this subsection is to collect a wide range of information about each work spell experienced by the respondent during his/her career (e.g., profession, type of contract, full time or part time work).

Spells of non-employment/interruption of employment. This subsection is aimed at recording some basic information about each spell of temporary or permanent interruption in the respondent's work career (including leaves of absence, periods of layo , etc.).

Current Employment is the rst subsection to be administered. If the respondent has had at least one previous work experience, then the Work Events subsection is introduced and all the information on the rst work experience is collected. If that work experience is still ongoing at the time of the interview, the Employment section is over. Otherwise, the end date of the rst work spell is recorded. If, at the end of the rst work spell, the respondent immediately started a new job, then a new Work Spell is registered. Otherwise, the Spells of non-employment/interruption of employment subsection is collected.

Family

The purpose of the Family section is to reconstruct the family history for all the interviewed household members from their birth to the time of the interview.

This section consists of two subsections.

Family of origin. This subsection is aimed at delineating the characteristics of the household members at two time points in the respondent's life (i.e., at birth and at age 14).

Current family. This subsection is designed to record any partnership experienced by the respondent from the time of leaving the family of origin or forming the rst union until the time of the interview.

Again, as in the Residential Mobility and Employment sections, the concept of spell is used to de ne each partnership experienced over the course of the respondents' lifetime. A spell refers to any kind of union, which may be more or less formalised. The end of a spell may correspond to the termination of a given relationship (e.g., separation or death of a partner) or to a change

in the nature of the relationship between the same two people. For example, cohabitations that became marriages or registered civil unions are considered as transitions from a spell to another. In other words, cohabitation should be indicated as the first event, and marriage (or civil union registration) as the following event. Moreover, the Current family subsection includes information on respondents' partners, children, and events of care-giving provided to other household members.

3.3.3 Cross-Cutting Themes

This part of the questionnaire collects personal information as well as respondents' attitudes and behaviours in several domains such as quality of life, Internet use, personality traits, health, and political participation. In addition, household-level information on economic and material deprivation, well-being, resources, debts, benefits and supports are retrieved from one person per household (i.e., the Reference Person).

Quality of Life

The purpose of the Quality of Life section is to collect some basic information about individuals' perceptions about:

- Life satisfaction
- Resources and environmental problems in the neighbourhood where they live

Respondents' satisfaction with life was first measured with a global scale asking respondents to evaluate their lives as a whole on a single item (e.g., Cheung and Lucas 2014) ranging from 0 Not at all satisfied to 10 Completely satisfied . Satisfaction toward specific aspects of their lives was additionally investigated through a battery of 6 items, looking at their relationships with family and friends, their leisure time, the house and the neighbourhood where they live in, and their financial situation.

Neighbourhood quality was instead measured looking at respondents' perception toward the physical-urban, social, cultural, psychological and environmental quality of the neighbourhood (Fornara et al. 2010).

Internet

The Internet section focuses on basic information about respondents' access to and use of information technology and new media. Specifically, the

Table 3.2 Cross-cu ing themes: sections and target respondent.

Sections	Target respondent
Personal Information	All eligible household members
Quality of Life	All eligible household members
Internet	All eligible household members
Deprivation and well-being	Reference person
Personality traits	All eligible household members
Health	All eligible household members
Household resources and debts	Reference person
Bene ts and supports	Reference person
Political participation	All eligible household members

following issues are investigated:

- Internet access
- Use of network-related technologies
- Possession of mobile technology
- Smartphone use
- Social media access and frequency of use

Internet access is investigated in three areas of daily life through the use of di erent technologies: family environment; work environment, only for those who are in employment at the time of the interview; other areas. These questions aim to reconstruct the respondent's actual opportunities to access the network at di erent places and times of day.

Smartphone usage habits were instead measured using an extended version of the Smartphone Pervasiveness Scale (SPS-A) (Gerlo 2022). The SPS-A consists of a set of items asking respondents how frequently they use their personal device in di erent moments of the day. In its original version, the SPS-A was focused on seven daily-life moments particularly relevant to individuals' psychosocial wellbeing that, according to the literature, could be negatively a ected by excessive smartphone use. Here, the scale was enriched with 3 additional items, all on a 5-point scale (From 1 Never to 5 Very often) in order to cover the three relevant sub-domains of smartphone pervasiveness in social relations, on sleep cycle and while carrying out other activities.

Finally, social media access and frequency of use were collected by rst asking respondents whether they have a personal pro le on one or more of the

social media or apps presented in a prede ned list (e.g., Facebook, Instagram, Twitter, etc.). Those who con rmed they had one or more pro les at the time of the interview were asked how frequently they used them on a 6-point scale, ranging from 1 "Everyday or almost" to 6 "Never" (Eurostat 2018).

Deprivation and Well-Being

The "Deprivation and Well-Being" section is administered exclusively to the Reference Persons, asking them to report on three main areas of family deprivation (Fusco et al. 2013). The rst area encompasses economic strains, asking respondents whether their household cannot a ord to face unexpected expenses, one week annual holidays, a meal with meat or sh every second day, keeping home adequately warm, and visiting the dentist. The second area focuses on enforced lack of durables, such as dishwashers, personal computers, cars, etc. Finally, housing issues are faced by asking respondents whether their dwelling su ers from structural problems (e.g., roof to be repaired, inadequate toilet) and shortage of living space.

Personality Traits

Respondents' personality traits are collected using the GSOEP Big Five Inventory (Hahn et al. 2012; Lang et al. 2011) translated and adapted to Italian. It consists of a battery of 15-item on a 5-point scale measuring the following ve domains of personality:

- Extraversion (tendency to be con dent and enthusiastic in interpersonal relationships)
- Agreeableness (tendency towards altruism and looking after others)
- Conscientiousness (tendency to be precise, accurate, and persistent)
- Neuroticism (tendency towards control of emotional states and impulses)
- Openness (openness to new ideas, to the values of others, and to one's own feelings)

Health

The purpose of the "Health" section is to collect information about respondents' perceptions of their own health conditions and some of their health-related behaviours and lifestyles. It is organised in three parts according to the contents it deals with.

The first questions belong to the SF-12 scale (Ware et al. 1996). The SF-12 is the shorter version of the most popular generic measure of patients' outcomes, the 36-Item Short-Form Health Survey (SF-36), which covers eight dimensions of health status: physical functioning, role limitations due to physical health problems, bodily pain, social functioning, general mental health, role limitations due to emotional problems, vitality, and general health perception (Ware and Sherbourne 1992). Two summary measures can be obtained from the eight dimensions without loss of information, generating a measure concerning physical health (Physical Component Summary Scale Score - PCS) and another one concerning mental health (Mental Component Summary Scale Score - MCS) (Ware et al. 1996). The decision to reduce the number of items from 36 to 12 has been made in order to reduce the room of questions required in a questionnaire and ease respondents' burden. It has been demonstrated that the 12-item sub-set of the original 36 items, which includes one or two items for each of the eight dimensions, can be a valid shorter version. Moreover, SF-12 produces the two summary scales PCS and MCS closely replicating those obtained through the original SF-36 (Jenkinson and Layte 1997; Ware et al. 1996). Additionally, Gandek et al. (1998) validated the SF-12 for Italy, asserting that it provides good replications of the SF-36 summary measures. Thus, these measures can be reliably used on the data available from the ITA.LI survey.

In Table 3.3, the two summary measures and their respective items, which correspond to the questions asked to individuals who participated in the ITA.LI survey, are reported.

In details, the 6 items concerning the questions on physical health administered in the questionnaire were the following: General health (GH1) was the question about the self-perceived health asking whether the individuals' health was Excellent, Very good, Good, Fair, or Poor; Moderate activities (PFP2) determined whether individuals' health limited them in performing moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or cycling (extremely, partially, or not at all); Climb Several flights (PFP4) asked individuals whether their health limited them in climbing several flights of stairs (a lot, partially, or not at all); Accomplished less (RP2) asked individuals whether they accomplished less than they would have liked with their work or other daily activities, as a result of their physical health (yes or no); Limited in kind (RP3) assessed whether individuals were limited in the kind of work or other activities, as a result of their physical health (yes or no); finally, Pain-interfere (BP2) determined how much pain interfered with individuals' everyday work, including both work outside the home and housework (not at all, a little bit, moderately, quite a bit, extremely). Furthermore,

Table 3.3 12-Item Short-Form Health Survey Items.

Summary measure	Items
Physical Component Summary Scale Score - PCS	General health (GH1) Moderate activities (PF02) Climb several flights (PF04) Accomplished less (RP2) Limited in kind (RP3) Pain-interfere (BP2)
Mental Component Summary Scale Score - MCS	Accomplished less (RE2) Not careful (RE3) Energy (VT2) Peaceful (MH3) Blue/sad (MH4) Social-time (SF2)

the 6 items concerning the questions on mental health administered in the questionnaire were the following: Accomplished less (RE2) asked individuals whether they accomplished less than they would have liked with their work or other daily activities, as a result of emotional problems, such as feeling depressed or anxious (yes or no); Not careful (RE3) assessed whether individuals did work or other activities less carefully than usual, as a result of any emotional problems, such as feeling depressed or anxious (yes or no); Energy (VT2) asked individuals how much of the time during the previous 4 weeks they have had a lot of energy (none of the time, a little of the time, some of the time, a good bit of the time, most of the time, all of the time); Peaceful (MH3) asked individuals how much of the time during the previous 4 weeks they have felt calm and peaceful (none of the time, a little of the time, some of the time, a good bit of the time, most of the time, all of the time); Blue/sad (MH4) asked individuals how much of the time during the previous 4 weeks they have felt down hearted and blue (none of the time, a little of the time, some of the time, a good bit of the time, most of the time, all of the time); finally, Social-time (SF2) determines how much of the time physical health or emotional problems interfered with individuals' social activities, like visiting friends, relatives, etc. (all of the time, most of the time, some of the time, a little of the time, none of the time).

To compute the two summary measures, procedures recommended by the developers (Ware et al. 1995) can be used. First, all the variables considered

must have the same coding so that higher scores represented good health. Second, for each variable, dummy indicator variables must be created for all but one response choice category. Therefore, out of 47 total response categories among the 12 items, 35 indicator variables are created. Third, the indicator variables must be weighted. This step can be implemented using coefficients from the general US population (Ware et al. 1995). Gandek et al. (1998) recommended employing standard U.S.A.-derived scoring of the SF-12 summary measures, in order to be the data comparable and interpretable across countries in relation to standard benchmarks, i.e., scores with a mean of 50 and standard deviation of 10 in the U.S.A. general population. Calculation of PCS can be then achieved by multiplying each indicator variable by its physical weight and by summing these 35 products. Accordingly, MCS can be computed by multiplying each indicator variable by its mental weight and summing these 35 products. Finally, the sum of the products is added to the respective constant from the general U.S.A. population (Ware et al. 1996) to obtain two continuous variables, one concerning physical health (PCS) and one concerning mental health (MCS). The second part of the Health section collects information on:

- The presence of chronic or long-term disabling diseases
- Areas of life where health creates difficulties
- The severity of the problems indicated

Finally, the third part of the section collects information about respondents' weight, height, insomnia, and physical activity.

Household Resources and Debt

The purpose of this section is to collect basic information about the economic and financial situation of the entire household. Questions belonging to this section are administered exclusively to the Reference Person, who is asked to give information about:

- The value of the home owned
- The amount of the home loan that the household is still paying
- The average monthly household income
- Income sources (employment income, pension income, and real estate income) expressed as a percentage of the total income
- The household's ability to save

- The total amount of debt the household has incurred to banks or nance companies

Subsidies and Supports

This section, which is reserved to the Reference Person, is aimed at detecting the total amount of economic resources and cash aid that respondents received during 2018from relatives, friends and/or from the public system (social allowance, Inclusion Income support, Citizenship Guaranteed Minimum Income, minimum living wage, food minimum). Additional information is also collected on the total amount of economic resources provided by the Reference Person or some other household member for the bene t of other non-cohabiting relatives or friends.

Political Participation

The purpose of this section is to collect information about:

- Respondent's political orientation
- Intention to vote in the next election
- Voting behaviour in past elections

Political orientations were collected drawing on a single left-right item (e.g., Barnes 1971), asking respondents to place themselves on a 9-point scale (from 1 Far left to 9 Far right). They were left with a midpoint anchor without forcing them to take sides on the scale, as suggested by the methodological literature (e.g., Kroh 2007; Saris 1988).

Respondents' voting in past elections and intentions to vote in the future were instead measured asking them to choose from the complete list of parties, movements and coalitions updated at the time of the interview.

3.3.4 Interviewer's Observations

At the end of each of the main themes and after the Health and Political Participation sections, the interviewer is asked to report about the presence of other people during the interview and their potential in uence on the answers given by the interviewee. In addition, at the end of the questionnaire the interviewer must answer additional questions about the interview as a whole (e.g., the context of the interview; the likelihood that the respondent will participate in the next survey). Finally, at the beginning and at the end

of the questionnaire, interviewers are asked to give their assessment of the physical appearance of the people they interviewed. This is because the facial attractiveness of the respondents may have in uenced the data collection process as well as their life history up to the time of the interview (e.g., Sala et al.2013).

3.4. Proxy Questionnaire

The proxy interview is based on a shortened and simpli ed version of the main questionnaire constructed by selecting only some of its questions. It is handed out to another member of the household with the aim of collecting information on the member unable to take part in the study. The proxy questionnaire consists of an introductory section aimed at gathering information about the reasons for the designated household member's absence/inability to respond. This is followed by a shortened version of the main theme sections (residential mobility, education, employment, and family). The focus is only on the current condition and on some of the past life experiences of greater importance (e.g., degrees and years of achievement) and detectable with su cient accuracy. Information about the cross-cutting themes is not collected, since the questions that make up these sections are mainly focused on detecting perceptions and attitudes whose expression cannot be delegated to others.

4

pre-field activities, pre-testing and piloting

Prior to the rst wave, the RG carried out several pre- eld activities in collaboration with the SA. The quality of the survey instrument and the proper functioning of the data collection procedure were rst assessed. Questionnaires were validated using a three-stage procedure that encompassed pre-tests by means of cognitive interviews, mock interviews and informal testing, and pilot testing

4.1. Pre-Testing (Cognitive Interviews)

Pre-testing the questionnaire is aimed at increasing the validity and reliability of survey evidence; it is an essential step in the survey development process (Willis 2018). This activity focuses on how people are answering the questions proposed, since a number of di erent processes are involved in the answering experience, and it is necessary to ensure that respondents interpret and answer questions in the same way in which these are intended by the researcher. This phase will thus allow determining if respondents understand the questions as well as if they can perform the tasks or have the information that questions require. Cognitive interviewing (CI) (Alcser and Conrad 2007). The method is not limited to the evaluations of individual survey questions, being also useful to pre-test other materials developed to convey or request information from survey respondents, such as recruitment letters or scripts, instructions, and consent forms (Willis 2018). The method's usefulness lies in the possibility to identify before the eld questions that pose a challenge to respondents, allowing to revise the questionnaire before the de nitive data collection. If some questions are perceived as ambiguous by the respondent, resulting in di culty to be answered, it may indicate that a rephrasing or a better speci cation of answer options is necessary, as the current version will not

su ce for the proposed analyses. Survey questions perceived as confusing by the respondent may also trigger the presence of interviewer error, as a consequence of possible attempts of clari cations and rewording by the interviewer which can lead to further measurement error. There are several ways to conduct CIs, with a distinction between concurrent and retrospective methods. The former method involves asking verbal probes or by asking respondents to `think aloud' as they formulate an answer to the questionnaire, whilst the latter method asks verbal probes after completing the questionnaire.

A pre-test of the questionnaire was conducted in December 2018. Ten CIs were performed to assess the goodness of the questions, which helped in xing the structure of the questionnaire as a whole. We relied on a retrospective approach, which allowed observing respondents to navigate and react to the questionnaire without interruption, permitting to measure the amount of time respondents needed to complete the questionnaire more accurately. The RG carried out a series of interviews involving individuals with di erent characteristics in terms of sex, age, educational attainment, family and employment status. The pre-test allowed the RG to assess the understanding and correctness of the logical ow of survey questions and to estimate the completion time for each section of the questionnaire.

A total of six cognitive interviews were conducted with potential respondents in January 2019. Participants were selected to ensure a su cient degree of heterogeneity in their socio-demographic characteristics and were individually interviewed at the SA headquarters. The goal was to assess the individuals' comprehension of both questions and response categories, especially looking at the more sensitive ones (Beatty and Willis 2007).

4.2. Mock Interviews and Informal Testing

Mock interviews constitute an informal way to test the questionnaire, allowing testing the questions with a close colleague to identify potential di culties in the question-answer process. Likewise, researchers can simulate a self-interview in which they put themselves in the position of the respondent and try to answer each question. These are informal ways of testing the research instruments that can be very helpful in identifying critical issues that may not be evident at rst sight. The RG performed repeatedly these testing, to ameliorate the questionnaire.

Simulations of the interview were conducted by the GR in April 2019, after the SA nished implementing the questionnaire in its computerised version. At this stage of work, the RG assessed the entire data collection process by

performing the following activities:

- Comparing the draft and the computerised version of the questionnaire, focusing on the text of the questions, the response categories and filters
- Reconstructing and testing all possible interview flow for each section of the questionnaire
- Verifying the correctness of the overall questionnaire flow within and between all its sections, including the contact procedure

At the end of the simulations, the RG communicated to the SA all changes required to optimise the survey tool.

4.3. Pilot Testing

As regards ITA.LI, we carried out a pilot of all data collection instruments and procedures in 50 households, which took place in May 2019. Interviewers were assigned the preliminary task of drawing up the list of all households residing at the sampled address and then to randomly select one of them to be contacted for the interview. This procedure is part of a more articulated probability design based on four sampling stages (see Pisati 2023). Interviewers' role was then crucial not only for the data collection process, but also for the preliminary stage of household selection.

They were indeed trained and supported in selecting the sample household and effectively involving their members in the survey, building their loyalty to the project, and, consequently, collecting quality data over time.

The quality of the survey instrument and the proper functioning of the data collection procedure were both evaluated through the implementation of the activities described.

4.4. Interviewers' Training

A total of 175 interviewers located throughout the country were hired to conduct the first wave of the survey. They were carefully trained to address any communication problems with household members (e.g., cultural barriers). The selected interviewers must have had at least a high school diploma (or equivalent), an excellent knowledge of the Italian language, basic computer skills and previous work experience in the same role. All interviewers were supervised and monitored throughout the study. To this end, each interviewer was provided with a unique and progressive identification code for ex-post monitoring actions on micro-data and para-data files.

Interviewer training was provided through brie ngs held on a regular basis by the SA. The rst set of brie ngs was conducted prior to data collection. New brie ngs were carried out throughout the eld activity both to cope with the turnover that characterises this professional gure and to ensure data quality.

At least two weeks prior to the brie ng, interviewers received all training and eldwork materials. The brie ng was organised to address the main theoretical and technical aspects related to the survey. The theoretical part aimed at training the interviewers about the contents and objectives of the study, the structure of the questionnaire, the meaning of the questions, the peculiarities of longitudinal studies, the code of conduct to be held with the interviewees, as well as the strategies to be activated in order to motivate reluctant respondents. Interviewers followed standardised interviewing principles and received speci c instructions to manage soft refusals. The technical part of the brie ng was instead aimed at illustrating the functioning of the CAPI hardware equipment, the software for managing the electronic questionnaire and the contact procedure management, the system of data acquisition and transmission to the SA servers. In this part of the brie ng, interviewers were also involved in simulations to familiarise with the entire eldwork process.

Brie ngs were conducted in classrooms equipped with a network-connected survey device (tablet) for each of the participating interviewers. To ensure a high level of quality throughout the training process, individual brie ng sessions involved a small number of participants and were delivered only by quali ed trainers with research experience in the eld. For each session, at least two expert trainers separately managed the theoretical and technical parts of the brie ng. The RG actively participated in some of the brie ngs, motivating the interviewers and assessing compliance with the standards of conduct agreed upon with the SA.

The training brie ngs were planned to last two days, during which the following activities took place.

Day 1

- Project presentation, conducted by RG coordinator, introducing the interviewers to the general contents and objectives of the study
- Detailed explanation of the preliminary steps of the data collection process, i.e., household selection and contact
- Simulations of the household selection and contact

Day 2

- Detailed explanation of the remaining steps in the data collection process, including contacting individuals, collecting informed consent, and carrying out the questionnaire
- Simulations of the contacting individuals, collecting informed consent, and carrying out the questionnaire steps

SA conducted the brie ngs in di erent areas of the country at multiple times to ensure an adequate number of active interviewers for the entire duration of the eld. The SA mainly used in-person training delivery modes.

The interviewers were also trained to grant better probabilities to access the eld and maximize response rate. Speci cally, they were informed of the bene ts of getting high response rates within the research process, as a way to a) provide statistically representative results; b) demonstrate that the data are comparable; c) avoid biases in the parameter estimates; d) enable the divulgation of the evidence reached within the scienti c and academic community.

The interviewers received speci c indications to adopt strategies to maximise response rate, such as a) planning and studying the sample entrusted; b) visiting all the addresses as soon as possible; c) planning to reach households at their addresses in the week-end and/or in late afternoon or evening, when it is more likely to meet a potential household reference person at home; d) distributing the visits in di erent times to reach shift workers; e) stressing the importance of the research and make the best use of the communication material available; f) deploying personal capabilities to transform refusal in complete interviews, acting exible and o ering to come back to the household in a di erent times; g) leveraging on the participation incentives.

Di erent topics were suggested to improve participation, highlighting the collective bene ts deriving from the individual contribution to the study, such as the possibility to implement and adopt social policies, also reminding that the contact has no commercial purposes, privacy is safeguarded, there are no mandatory questions and topics perceived as sensitive could be skipped and that the consent to participate can be withdrawn at any time.

Speci c training was provided also concerning the tone of voice and speaking strategies.

4.5. Access to the Field

The RG has planned a series of activities to raise awareness of the contents of the study among local institutions and the target population. Before the data collection started, the RG sent a presentation letter to the local authorities and law enforcement agencies of the selected municipalities, asking the interviewers to contact them for support during the household contact phase in case of need (see survey documentation related to Wave 1 on the ITA.LI website). In collaboration with the University of Milano-Bicocca's press office, the RG also launched a promotional campaign of the study on national and local media. Several articles and news were published, offering interviewers additional material to be shown to potential respondents.

Even the SA planned a series of actions to facilitate the access to the field for all interviewers called upon to work in problematic socio-economic contexts. First, interviewers were enrolled on the basis of their knowledge of the local context. Second, local institutions were involved to inform potential respondents, raise their awareness toward the study, and legitimise the interviewers' work (the documents are sent are available on the ITA.LI website: https://iassc.unimib.it/it/progetti-di-ricerca/itali/documentazione).

5

data collection

The first stage of data collection consisted in contacting the household residing at the sampled address. Once obtaining the cooperation of the household, individual members were included in the sample registry and asked to participate in the survey. At this stage of contact, the interviewer presents the burdens and benefits of participating in the survey to all the eligible household members, together with the request to participate, and, where possible, to proceed with the interview. A system of household incentives was also developed, with the aim of stimulating participation in the study, as well as a helpline for questions or concerns. As known in literature the use of pre-issued incentive has been proven to be an effective strategy for improving recruitment into field studies (Lipps et al. 2019; Smith et al. 2019). All the activities regarding contacts, information, and interviews were continuously monitored by the SA and the RG, with a specific focus on interviewers work and household and individual participation. Data collection procedures were monitored for the entire duration of the field, in order to identify critical issues and promptly implement strategies to tackle them.

5.1. Contacting and Conducting Interviews

The fieldwork was organised in several phases, including: initial contact with the household; contact with all the eligible members; informed consent collection; and interview. While contacting the household, the interviewer fills out the contact form with information about all the members living in the house. After contacting a potential respondent, the interviewer goes through a standardised step-by-step procedure to register the informed consent and to proceed with the interview (the documentation in available on the ITA.LI website: https://iassc.unimib.it/it/progetti-di-ricerca/itali/documentazione).

Each of the above-mentioned activities was handled directly by the interviewers using their personal survey device.

The following sections illustrate the steps underlying the household contact and interview collection procedures.

5.1.1 Household Contact

After completing the household selection procedure, the interviewer proceeds with face-to-face attempts to contact the household by ringing the doorbell, knocking on the door, etc. A minimum number of eight contacts per household is fixed in case of unsuccessful attempts (or unproductive outcomes).

For each contact attempt, the interviewer is required to record the following information: date and time; method of contact (e.g., in-person visit with face-to-face meeting, in-person visit but only by intercom, telephone, non-response, etc.); outcome; any informative materials delivered; reason(s) for any refusal to cooperate and, if possible, characteristics of the refusing household; language with which the interviewer made the contact; language spoken in case of non-native speakers.

Each contact attempt made by the interviewer automatically generates one of the following outcomes: temporary unproductive, final unproductive, and final productive.

Temporary Unproductive Outcome

The temporary unproductive outcome includes the following cases:

- Successful contact appointment. The interviewer was able to speak, face-to-face, with a household member; however, at that time it was not possible to collect information about the household. As a result, an appointment is scheduled at a more convenient time
- Successful contact person unable to provide information. The interviewer was able to speak with someone who, however, was unable to provide information about the household (e.g., minors, members who do not speak Italian, etc.)
- Non-contact. The interviewer is unable to speak to anyone at the selected dwelling

In case of temporarily unproductive contacts or non-contacts, the protocol states that if no member is at home at the first visit, the interviewer will leave a copy of the survey cover letter and/or a contact card, choosing whichever

instrument they feel is most e ective in relation to the speci c context. These put emphasis on the fundamental importance of the participation of the selected family for the success of the investigation, the topics addressed, and length of the interview, as well as the incentives for participation.

In the absence of feedback from the household, the interviewer must visit the address again at least twice a week, including at least once on Saturday or Sunday at di erent times. At least one attempt per week must be made after 7“30pm, until the minimum number of eight attempts is reached. Materials (cover letter and contact card), if not delivered directly to a member, should preferably be left in the mailbox of the selected household. Once the established number of contact attempts has been made without success, the address can be excluded from the sample (nal unproductive contact). If successful, however, the interviewer may consider the contact to be nal productive and proceed as described below.

Final Unproductive Outcome

The nal unproductive outcome includes the following cases:

- Hard refusal. The interviewer receives an immediate refusal to participate, and it is impossible to continue with the completion of the Household Grid or planning any further contact attempts
- Contact attempt invalid address. After the contact attempt, the address turns out to be ineligible as the building is uninhabited or non-residential
- Household withdrawal. This outcome occurs when the respondent (after a previous contact attempt by the interviewer) contacts the SA asking to exclude his/her household from the study. In this case, the SA contacts the interviewer to whom the address is assigned to inform him/her of the refusal and the reasons for it, which should be recorded on the contact form
- Non-contact eight attempts reached. Once the interviewer has made eight attempts without ever being able to make contact with the household, he/she may decide whether to consider that contact as de nitively unproductive or to proceed with further attempts
- Successful contact negative outcome. The interviewer get in touch with someone adequately informed about the household (even if not an household member, e.g., a neighbour), con rming that: the household has moved elsewhere; the household is not available for the entire duration of the eld (e.g., on vacation, hospitalised single member, etc.); no one in the household

speaks Italian; the only member is deceased

Each of the above unproductive outcomes is considered as unconvertible and, therefore, the address is dropped. A special case, however, is represented by non-contact - eight attempts reached , where the choice of making the address de nitively unproductive is left to the interviewer. The interviewer must decide according to their professional experience whether it is worth to try again with further contact attempts or give up.

Final Productive Outcome

The interviewer is able to talk to a household member who has adequate information about the other members. This person becomes the Household Grid Holder. The interviewer proceeds with presenting the study to the Household Grid Holder, asking them for consent for the entire household to participate, and verifying their willingness to provide all information necessary to assess the eligibility of their household.

The Household Grid is lled in by the holder, collecting information on the name and surname, relationship with the holder, sex, marital status, date of birth, citizenship, employment status and education of all members of the household, including children under 16 and those temporarily absent or unable to take part in the interview for health reasons. The rationale for this step is to identify all eligible members of the household and, among those eligible, the most appropriate person to be asked questions about the household's nancial and economic situation (Reference Person).

After recording the nal productive outcome at the household level, the interviewer proceeds with individual contact attempts and interviews of the eligible members.

5.1.2 Individual Contact

Once the household contact process is completed, the interviewer proceeds by contacting all eligible members and lling in the individual contact form. For each contact attempt, the interviewer should include the following information: date and time; method of contact; temporary outcome; reason(s) for refusal; member who refused on behalf of others; help received from other household member(s) in case of a non-native speaker interviewee; nal outcome of the contact. The individual contact form is completed at the end of each visit, when the temporary or nal outcome of the contact attempt becomes known (e.g., full interview vs. partial interview).

As with households, individual contact attempts made by the interviewer may produce di erent outcomes. The following three macro-categories are identi ed: temporary outcome; nal unproductive outcome; nal productive outcome.

Temporary Outcome

The temporary outcome includes the following cases:

- Non-contact. The interviewer is unable to contact the person of interest for that speci c visit
- Successful contact appointment. The interviewer is able to contact the household member, who agrees to participate in the survey. The interviewer then sets an appointment to carry out the interview at a later date
- Successful contact need for language mediation. The interviewer manages to reach the household member, who, however, is not able to perform an interview in Italian. In this case, the interviewer should verify whether there is a household member available to act as linguistic mediator
- Successful contact need for proxy interview. The interviewer catches on that a household member is not interviewable for the entire duration of the survey, due to one of the criteria for which proxy interview is required. In this situation, the interviewer should identify another household member who can provide information on behalf of the non-interviewable member answering the proxy questionnaire

In the case of a temporary outcome, the protocol requires at least ve attempts to contact the respondent, or someone who can potentially act on their behalf. Once the established number of contact attempts have been made without success, the individual can be considered dropped (nal unproductive outcome). If successful, however, the interviewer should consider the outcome of the nal contact productive and proceed as described below.

Final Unproductive Outcome

The nal unproductive outcome includes the following cases:

- Respondent withdrawal. A household member contacts the SA to communicate that he/she does not intend to participate in the survey even though he/she was never contacted by the interviewer. Consequently, the SA informs the interviewer about both the refusal and the reasons for it, which is recorded on the contact form

- Non-contact ve attempts reached. The interviewer has made ve attempts without ever being able to contact the household member. From each subsequent attempt, he/she may decide whether to consider such contact as de nitively unproductive or to proceed with further attempts. The same strategy applies to both the respondent and any person in charge (e.g., language mediation or proxy interview)
- Successful contact negative outcome. The household member turns out not to be interviewable for one of the following reasons: language mediation or proxy interview are not feasible; the individual is deceased; the entire household has moved away
- Partial interview impossible to complete at a later date. The interview is interrupted and the respondent withdraws from survey participation. Therefore, the interview can no longer be completed

Final Productive Outcome

The nal productive outcome includes the following cases:

- Successful contact interview attempt. The interviewer reaches the household member, who is available to respond to the interview. The nal outcome may be a complete interview, a partial interview, or an interruption with no future opportunity to complete the interview
- Successful contact language mediation attempt. A household member is available to respond to the interview on behalf of a non-native speaker individual. The nal outcome may be a complete interview, a partial interview or an interruption with no future opportunity to complete the interview
- Successful contact proxy interview attempt. A household member is available to respond to the proxy interview by providing the information on behalf of the non-interviewable individual. The nal outcome may be a complete interview, a partial interview or an interruption with no future opportunity to complete the interview

5.1.3 Questionnaire Administration

When a nal productive outcome is reached, the interviewer administers the questionnaire to one respondent at a time. The only exception is made for the interview attempt with language mediation, where the designated mediator remains available for the entire course of the interview. The standard

procedure of administration includes a detailed presentation of the research and of the procedures of personal data management. After that, the interviewer proceeds with the collection of the informed consent, the administration of the interview and, nally, the request for consent to be contacted for further research activities.

Presentation of the Research and Processing of Personal Data

At the beginning of the interview, the interviewer hands the respondent two documents: the Participant Information Sheet and the Information on the processing of personal data. The respondent is invited to read both documents in detail, while the interviewer remains at full disposal for any doubts or requests for clari cation. This documentation provides basic information as regards the scopes of the research, the commitment required to participate, the length of the interview, the topics covered, the incentives for participation and the privacy and personal information safeguard (see the Survey documentation related to Wave 1 on the ITA.LI website: https://iassc.unimib.it/it/progetti-di-ricerca/itali/documentazione), and remains available to the respondent even after the end of the entire process, regardless of their choice to join the survey or decline participation.

Collecting Consent/Refusal

Once the respondent has no more doubts or concerns, the interviewer proceeds with the formal request for consent to take part in the study. The record of the respondent's choice of approval/refusal is collected verbally and then formalised on the self-certi cation form available to interviewers.

After compiling the form, the interviewer records the response on the device provided for conducting the interview. In case of refusal, the member is considered dropped and the interview immediately ends with a nal unproductive outcome. Conversely, in the case of approval, the interviewer moves on with the interview.

Administration of the Interview

The interview is administered following the principle of standardisation to ensure stimulus invariance (Fowler Jr and Mangione 1990). The interviewer is required to carry out the questionnaire by reading both questions and response categories as they are written.

However, some questions require a spontaneous answer, as for the case of

Table 5.1 The participation-based incentive plan.

Members interviewed	Voucher value per household
1 or 2	¿ 15
3 or 4	¿ 25
5 and above	¿ 40

the open question referred to the occupation. These additional instructions are reported in the questionnaire as notes for the interviewers.

Consent for Further Activities

The last section of the questionnaire deals with the informed consent for three additional activities: re-contact for new research; social media data linkage; National Institute of Social Security - INPS data linkage. A paper consent form is given to the respondent for each of the above-mentioned activities (see the Survey documentation for Wave 1 on the ITA.LI website). The respondent is then invited to fill the forms out, while the interviewer remains available for any doubts or requests for clarification. Respondent's approval/refusal to each of the additional research activities is finally recorded by the interviewer on the device for data collection.

Despite a relatively high survey participation, a small share of respondents gave consent for data linkage with the INPS database and major social media. The number of consents obtained was not sufficient to ensure a joint analysis of data from these multiple sources, prompting the RG to forego such in-depth activities. In contrast, the request for consent to re-contact for further research activities (not encompassing next survey waves) was more successful among respondents, giving the RG the opportunity to engage approximately 20% of the sample in future research activities promoted by the Department of Sociology and Social Research of the University of Milan-Bicocca.

5.2. Household Incentive Plan

An incremental incentive plan has been designed to encourage survey participation (Lipps et al. 2019). An online shopping voucher is delivered to the household, conditional to the participation in the interview of a predetermined minimum number of eligible members. The nominal value of the shopping voucher varies according to the outline shown in Table 5.1.

5.3. Household Survey Support

A support service has been o ered to all the household members through a toll-free number and an email address to which they could turn to receive answers to questions or concerns regarding the study.

The telephone and email free services remained active throughout the data collection. The operators were trained on the main methodological, organisational and content aspects of the survey, as well as on basic notions of communication techniques to reassure participants of any doubts or questions regarding the survey and to motivate those who expressed reluctance to participate.

Finally, the ITA.LI website includes a speci c page for participants containing detailed information about: study aims, incentives for participation, data processing methods, information materials and contact details.

5.4. Monitoring

The interviewers were monitored by SA on a daily basis and relevant information was shared with the RG on a weekly basis. The entire monitoring process has been crucial to evaluate and implement corrective actions aimed at maximising the response rate and data quality, especially at critical times such as those following the advent of the COVID-19 pandemic.

The SA monitored the performance of individual interviewers using the following indicators: number of complete interviews, number of contacts by sample member, incidence of refusals, and non-response rate. The analysis of these indicators over time allowed the SA to recruit new interviewers in order to cope with the slowdowns in the eld due to both external factors and internal organisational problems. Moreover, the SA continuously monitored the refusal/non-contact indicators activating strategies for hard refusals conversion. Finally, eld coordinators dynamically managed interview assignments or reassignments depending on the individual interviewer's availability and workload.

RG monitored trends in contact outcomes and participation rates, also assessing their homogeneity across di erent segments of the population under study (see Pisati 2023).

6

data

The main objective of the ITA.LI study is to o er a wide audience access to an updated and well-documented source of information that can be used by the international scienti c community for the study of social change in Italy.

To achieve this goal, the RG adopted an approach to documenting the entire data lifecycle (Ball 2012), from study conception to data distribution to the public. It is well known that in today's context it is not possible to assess the quality of a study solely by the content of the information collected, but it is also necessary both to share with future users the choices made during the research process (Förster et al. 2020) and release data that meet the characteristics of Findability, Accessibility, Interoperability, and Reuse (Wilkinson et al. 2016), in full compliance with the Open Science paradigm (Watson 2015).

Within the project, time and resources have therefore been dedicated to the activity of data management and documentation, thanks also to the collaboration with UniData Bicocca Data Archive, specialised in data preservation procedures in the social sciences domain.

6.1. Data Management

Given the complexity of the investigation, it was necessary to adopt a mode of data management that can be summarised in the expression code- rst approach, which consists in focusing attention and work on the code that generates the data as opposed to the data itself, facilitating the activities of documenting the operations performed, transparently and collaboratively (Gentzkow and Shapiro 2014). This approach resulted in the writing of a set of scripts - in the Stata, Mata and R programming languages - aimed at automating the procedures of data creation and checking. The whole code was maintained within a repository managed by version control software

(Git, in this case) that ensured the tracking of all changes made within the scripts during data management activities. This tool was complemented by a repository hosting manager (GitLab) that facilitated collaboration among RG members and the handling of critical issues that arose during data management activities. In particular, through speci c functions (issue tracking), a detailed monitoring of the problems that emerged and of the code produced for their resolution was guaranteed.

The code produced was used for the management of the entire work ow: from the archiving of the original data, received from the SA, to the production of the les for the research addressed to the entire scienti c community. Following an iterative approach, data were made available to the RG through incremental releases from the early weeks of the eldwork. Each release was accompanied by a technical note containing all activities performed, closed and still open issues. Thanks to the feedback received and the internal RG discussion after each release, it was possible to improve the quality of the data.

The data management process followed a structured procedure. Starting from the raw data sent by the SA on a weekly basis, a rst series of checks was carried out with the aim of verifying the e ective correspondence of the data collected with that required by the RG. More speci cally, this phase involved checking the functioning of the survey instruments during the eld phase, with particular attention to the ow foreseen by the lters in the questionnaire, the format of the variables, the range of values allowed (wild codes) and the presence of any duplicate or incomplete records. Next, several datasets were created by separating the substantive data from the respondents' contact process data and further splitting them by topic to obtain a le structure consistent with the type of data represented (individual or household, repeated information or not, etc.). This activity allowed the RG to have adequate data les available to proceed with more in-depth clean-up activities.

6.2. Data Cleaning

Clean-up activities were directed primarily at individual datasets, adopting, as much as possible, shared and consistent strategies within the entire project. In particular, attention was paid to the following aspects:

- Recoding of the strings related to the category other, specify , leading them back to pre-coded categories
- Recoding of system-missing values and attribution of an explicit meaning to each type of missing information
- Deletion of redundant information due to, for example, the ow of spell

collection

- Consistency checks within the individual thematic les (e.g., logical coherence between the spells of the educational career of the interviewees), with the exception of temporal coherence (dates and spell durations) on which it was decided not to intervene
- Streamlining the information collected and the le structure to facilitate analysis. For example, merging variables containing similar information but derived from di erent questions for past and current job spells

Control activities between datasets aimed exclusively to verify the consistency between the information collected in the Household Grid and in the individual questionnaire, relating to the composition of the sample and households, the status of the individuals (eligibility, outcome of the interview, etc.) and the main socio-demographic characteristics.

The nal phase involved the production of the Public Use Files which, with respect to the data available to the RG, will be subjected to a careful assessment of the risk of identi cation and all necessary measures will be taken to protect the identity of the participants.

6.3. Data Release

Currently, only the rst wave of survey collection has concluded; however, given its longitudinal nature, e orts have been made already at this stage to plan strategies that facilitate the cumulative use of the data and related documentation released in each of the waves that will be conducted in the future.

6.3.1 Data Architecture

The data from the rst wave of the survey was organised taking into account the di erent survey instruments used (e.g., Household Grid, main questionnaire, proxy questionnaire) and with the goal of separating thematic information from more general information useful to all analysts, regardless of speci c research interests.

The available datasets and their main characteristics are listed below. For more details on the content and peculiarities of the individual les, please refer to the documentation accompanying the data.

- Personal Data (PD): substantive data regarding individuals interviewed via main questionnaire. Family of origin and cross-cutting topics are also available here

- Residential Mobility (RM): substantive data regarding residential mobility career of individuals surveyed via main questionnaire
- Education (ED): substantive data regarding educational career of individuals surveyed via main questionnaire
- Job History (JH): substantive data regarding employment histories of individuals interviewed via main questionnaire
- Partnership History (PH): substantive data regarding the partnership (marriages/cohabitations) history of the individuals interviewed via main questionnaire
- Caring (CA): substantive data regarding family care spells of individuals surveyed via main questionnaire
- Financial Resources (RS): substantive data regarding the financial information of surveyed households
- Proxy (PX): substantive data regarding individuals interviewed via proxy questionnaire
- Household Grid (HG): data for all households in the sample, including non-responding and ineligible individuals, identifiers, and weights

6.3.2 Question Naming Conventions

When specifying the survey instruments, it is necessary to assign a short name for each question asked, so that they directly correspond to variables in the data files made available to analysts.

Variable names were assigned in a systematic manner, with the goal of providing the user with references to the type of information contained but also to facilitate the cumulative use of variables across multiple waves in the study. The scheme used is as follows:

Where:

- WYY denotes the survey to which the variable belongs, and, specifically, YY denotes the year of the survey for each wave
- DD refers to the domain to which the variable refers and therefore to the thematic dataset in which it has been inserted (see previous list);

- NNN is a three-digit sequence that identifies the variable. The numbering generally reflects the order in which the question was asked within the reference questionnaire from which each variable is taken.

For example, the name of the variable W1_PD_001 consists of W1, prefix indicating that the variable belongs to the first wave (started in 2019), PD indicating the dataset it is part of (i.e., Personal Data), 001, is the progressive number of the variable.

The DD_NNN information pair uniquely identifies a specific variable and, consequently, a specific piece of collected information. This identifier will remain constant over time: in each wave the same variable will have a name given by the same identifier DD_NNN but with a different prefix, referring to the wave to which it belongs (e.g., W1_PD_001, W2_PD_001, etc.).

In contrast, the derived variables follow a partially different pattern. They have been named by keeping the reference to the wave and domain, while the progressive is replaced by letters that refer to the semantic content of the variable in question. For example, the name W1_JH_ISCO is used for the variable belonging to the data from the first wave (W1), related to the respondent's job history (JH) and containing information about the profession held, coded using the standard ISCO classification.

6.3.3 Data Documentation and Access

The following documentation was provided along with the data:

- Questionnaires: main questionnaire, proxy questionnaires, Household Grid questionnaire
- Codebook for each dataset
- User guide
- Quality profile
- Sample design, weighting, variance estimation, and data usage manual (Pisati 2023)
- Methodological notes
- Syntax

The data will be distributed by UniData - Bicocca Data Archive (https://www.unidata.unimib.it) and will be downloadable after registration. Documentation will be freely available both on the survey site (https://assc.unimib.it/it/progetti-di-ricerca/itali) and on the dedicated page

of the UniData website (https://www.unidata.unimib.it/?indagine=italian-lives-ita-li-2019-2021&highlight=ita.li).

6.3.4 Data Access and User Support

ITA.LI survey data is made available free of charge to universities and research institutes for research and teaching purposes in the form of an anonymized microdata file, available in various data formats. Interested users must fill a form and, after approval, the data can be downloaded via a secure data transfer system.

The RG offers different forms of user support, including the share of the syntax file to prepare the raw dataset for longitudinal analyses. Specific requests will be evaluated case by case after contacting the RG.

7

data coding and missing imputation

7.1. Occupation

7.1.1 Manual Coding

In ITA.LI respondents are asked for details of their occupation. This information was collected with open-ended questions because the scope of the possible responses is too broad to capture with a closed-ended question and a list of prede ned response categories (Züll 2016).

Descriptions of the occupations of their former partner's, their parents' and for a person who mainly provided for respondent's sustenance (if this person was not a parent) are collected as well.

These descriptions were then manually re-coded according the International Standard Classi cation of Occupations 2008 (ISCO 08) scheme, developed by the International Labour Organization (ILO) (http://www.ilo.org/public/english/bureau/stat/isco/isco08/). The occupation variables were coded to the 4-digit ISCO 08 version.

Coders had the task of assigning ISCO 08 codes to occupational titles starting from two main elements: a) job title and b) description of duties and tasks. Additional information could be used by coders (only if necessary): respondent's level of education, age, status in employment, rm sector, rm size.

12 coders were trained and instructed by providing basic information about ISCO's logic of classi cation and some examples of occupation coding. However, it is fundamental that coders are familiar with the structure of ISCO to achieve a high level of consistency. Therefore, we gave the coders time to explore the ISCO 08 scheme in detail. For the coders, the most valuable support is the listings of occupational titles and their associated ISCO 08 codes, providing a quick route into the occupational categories. Familiarity with the

Table 7.1 Variables in dataset with ISCO-08 classification by reference person.

Variables	ISCO-08
Respondent's jobs (all occupational events)	1917_JW_ISCO
Former partner's job (at the time of marriage or domestic partnership)	W1917_PH_019
Mother's job (when respondent was 14 years old)	1917_PD_337
Father's job (when respondent was 14 years old)	1917_PD_321
Person who mainly provided for respondent's sustenance if this person was not a parent (when respondent was 14 years old)	W1917_PD_362

Note: in the simplest case, ISCO-08 coding description was equivalent to the one in the answer field

scheme will help reduce the time needed to assign codes, but the possibility of error is likely to remain because there are many instances where a given occupational title and description could quite reasonably be assigned to several codes (some legitimate).

Assigning ISCO codes to job descriptions requires the use of a rule-based system to ensure that it is performed in an accurate, consistent, and efficient manner.

We adopted a double independent coding, i.e., two coders had to classify the same job independently, working without communicating to one another. The next step was comparing the code assigned to the same job description by the two independent coders. If they assigned the same code, it was considered correct, otherwise a third expert coder, i.e., coder with a long experience in this field, revised the job description and chose the suitable code.

Overall, in Wave 1 around 16'000 occupations were coded and about half of them required expert coder intervention to resolve the inconsistencies between original coders.

7.1.2 Automatic Programming

Manual coding occupations is a demanding and time-intensive activity. It is difficult to explain to the interviewees how the ISCO-08 classification scheme is structured and, furthermore, the open-ended question used to gather the description of the job has less influence on the answer than a closed-ended

question. Therefore, it is crucial to find a way to translate an open-ended question to a code belonging to the ISCO scheme. The description of each respondent's job can be translated into ISCO code by manual coding, automatic or semi-automatic methods. Manual coding is not always the best approach. As known in the literature, it could be expensive, time-consuming and non-transparent (Patel et al. 2012). Of course, automatic or semi-automatic methods have some drawbacks too.

First, it is not feasible to automatically code every answer correctly. Second, open answers can be characterised by spelling problems, grammatical problems, or they may not be complete. The computer does not work like the human mind, and it is difficult to classify the meaning of the words and distinguish between singular and plural, masculine and feminine, or verbal tenses. Lastly, the software has to correctly interpret the meaning of the text, i.e., identify which code from the classification best fits a description.

Automatic toolboxes used to code text responses are characterised by an informative base, that's to say a dictionary, with synthetic and unambiguous descriptions associated with the codes, and a search engine able to perform a text standardisation (e.g., stemming, tokenization,sentencing and lemmatization), in order to reduce randomness, variation and wrongness inherent in natural language. Once the dictionary is built, different approaches to automatic coding can be used. The rule-based approach is the simplest one: if the text data meets a prespecified logical condition a specific code is assigned. Another way to employ automatic coding follows the machine learning approach which does not start from a pre-built dictionary but it learns from previous coded information (i.e., Bayesian Multinomial models, support vector machines, neural networks).

In the first wave of ITA.LI a rule-based coding algorithm has been developed in order to translate the job descriptions into ISCO-08 code. Although the occupations were coded entirely manually, this kind of algorithm was usefully employed to check for potential errors.

Automatic coding of occupations turned out to be extremely helpful, since its introduction has been shown to improve the quality of the manual coding by removing characteristic inconsistencies. For the next waves of data collection, the RG will evaluate the opportunity to replace this method with that of manual coding. Since the matching between the two methodologies (manual and automatic) was high (75%) it is expected that the level of discontinuity between the two methodologies would be relatively low.

7.2. Income

In ITA.LI it was asked to each household member to report their net monthly earned income (or salary), specifying the exact amount or choosing a class in a salary range. From this information it is possible to derive the net household income. Both individual and family income are considerably interesting; nevertheless, they present a high number of missing data, making it necessary to adopt specific methods to estimate them from other information collected during the interview.

In Wave 1, 3'990 individuals (45.5% of the respondents) were employed, whereas 1'750 reported their net salary, 70 did not remember, and 210 refused to answer the question. Accordingly, overall, 56.1% of those employed had missing information on net salary. It is interesting to notice that the percentage of non-response varies according to professional condition, as it is 73.3%, 77.4%, and 77.4% for entrepreneur, freelancers, and self-employed, respectively, whilst it is much lower for employees (50.8%) and manual workers (47.7%). Specific indications on how to treat item non-response in ITA.LI data are present in (Pisati 2023).

8

ethical requirements

Participation in the ITA.LI survey is voluntary and can be formalised only after providing potential respondents detailed information on the project characteristics as well as on the burdens and bene ts deriving from personal involvement. After being properly informed, potential respondents can freely give or deny their consent to participate in the study and, consequently, actively take part in the proposed research activity.

Personal data are treated as strictly con dential, in accordance with the EU General Data Protection Regulation 2016/679 and the Italian Legislative Decree 196/2003.

ITA.LI protocols have been screened and approved by the Ethics Committee of the University of Milano-Bicocca on 31 May 2019 (No. 0042665/19).

8.1. Participants' Information and Consent

During the contact phase, the interviewer is required to collect respondents' informed consent to participate in the study and to process their personal data for research purposes.

At the rst contact, the interviewer gives the respondent the Participation Information Sheet and associated Research Consent Form (see the Survey documentation for Wave 1 on the ITA.LI website). Respondents who agree to participate in the research are requested to ll out the Research Consent Form. Once completed, the form is given back to the interviewer and then sent to the SA managers. In the case of refusals, potential respondents are not requested to ll the Research Consent Form. In both cases (consent or refusal), respondents receive the Information Sheet containing all the information needed to contact the RG.

Once respondents give their consent to participate, the interviewer hands over the Personal Data Processing Disclosure Statement and the associated

three data processing consent forms (see the Survey documentation for Wave 1 on the ITA.LI website). These forms are presented at the end of the interview. The rst form aims to collect respondent's consent to be contacted for participation in the ITA.LI qualitative study and/or in additional research projects promoted or carried out by the Department of Sociology and Social Research of the University of Milano-Bicocca. The second form requests access to publicly available information on social pro les (Facebook, Instagram, Twitter), the third concerns the access to administrative data available from the National Institute of Social Security INPS.

At any time, survey participants are free to express their concerns and/or make requests for clari cation face-to-face, via toll-free number or email. A dedicated website with all the information about the study (objectives, methods of participation and data processing) is also at their disposal. A link to the website is included in both the survey cover letter and the information sheets given to respondents.

8.2. Anonymity and Con dentiality

Personal data collected by SA are managed according to strict procedures aimed at preventing the access to unauthorised parties. These data were encrypted, sent to the RG data controller and deleted once the survey was completed. Prior to RG analysis of the data, information allowing direct recognition of participants (rst name, last name, address, social security number, phone number, email, etc.) was removed and each individual was assigned a randomly unique identi cation code (pseudo-anonymisation). The RG has access only to the data containing the unique code and questionnaire responses. Personal data is encrypted and accessible only to the Research Managers. At the conclusion of each survey wave, the RG will proceed with data anonymisation and archiving.

8.2.1 Anonymization Principles

Speci cally, anonymization procedures are implemented to reduce disclosure risks with minimal information loss, preserving data utility, and consists in removing (or modifying) one or more identifying variables (e.g., identi cation numbers, punctual addresses, etc.).

More in detail, these lasts include:

- Direct identi ers, that is variables such as names, addresses, identity card or social security numbers, which allows a direct identi cation of a respondent

but are not needed for statistical or research purposes, being important only in relation to the data collection procedure. Accordingly, such information should be removed from the published dataset

- Indirect identi ers, namely characteristics that may be shared by several respondents, and whose combination could lead to the re-identi cation of one of them. Variables that do not allow to identify individuals per sé, are potentially problematic in relation to privacy concerns when their combination permits to identify speci c subjects, for instance crossing information about age, sex, address, and occupation. Such variables are fundamental for the statistical analysis and should not be removed from the published data le; however, it is necessary to determine which variables could potentially favour identi cation in order to modify their speci city to eliminate the possibility of identi cation (e.g., replacing the address with the neighbourhood, job with occupational class, age with age class, etc.). It is crucial to balance privacy safeguard with information loss, maximizing the former and minimizing the latter

Disclosure risk depends not only on the presence of identifying variables in the dataset, but also on the existence of an intruder, that is subjects that could take advantage from acquiring identi able information for personal or commercial purposes, as well as on the costs of re-identi cation, which is in inversely proportional to the bene ts for a potential intruder.

Accordingly, the published data le does not include any data which allows direct identi cation, such as name and addresses. Such information is stored on a separate secure server accessible exclusively by the Research Managers, who require it to maintain the integrity of the database (e.g., in the case of change of address), and the SA, to enable the sending out of invitations at the next wave of data collection.

8.2.2 Retention of Personal Data

Participants' personal data have been retained for the duration of the research. The research was conclude upon completion of the last wave and associated data cleaning and quality control. Personal data concerning the purpose of participation in new research promoted or carried out by the Department of Sociology and Social Research of the University of Milano-Bicocca have instead been stored inde nitely.

After being anonymised, the survey data have been archived at UniData - Bicocca Data Archive according to standardised procedures shared with other Data Archives for the Social Sciences (in line with the standards de ned by the

European infrastructure CESSDA). The data, properly documented according to FAIR principles (Wilkinson et al. 2016), will be made available to the entire scienti c community.

8.2.3 Security Measures

The SA, in its capacity as External Data Processor, stores and transfers participants' personal information ensuring high safety standards about information encryption and access control. After seven days the data is deleted through an automatic secure wipe. The pseudo-anonymised data is encrypted and shared among RG members with cloud systems that provide access control.

references

Alcser, K. and Conrad, F. G. (2007) *Cognitive Interviews: Putting Guidelines Into Practice Adapted for the World Mental Health Study*. Survey Research Center, Institute for Social Research, University of Michigan. [Cit. on p. 35.]

Ball, A. (2012) "Review of data management lifecycle models (version 1.0)," *REDm-MED Project Document redm1rep120110ab10. Bath, UK: University of Bath* [Cit. on p. 51.]

Barnes, S. H. (1971) "Left, right, and the Italian voter," *Comparative Political Studies* 4(2), pp. 157-175. [Cit. on p. 33.]

Beatty, P. C. and Willis, G. B. (2007) "Research synthesis: The practice of cognitive interviewing," *Public opinion quarterly* 71(2), pp. 287-311. [Cit. on p. 36.]

Butz, W. P. and Torrey, B. B. (2006) "Some frontiers in social science," *Science* 312(5782), pp. 1898-1900. [Cit. on p. 3.]

Cheung, F. and Lucas, R. E. (2014) "Assessing the validity of single-item life satisfaction measures: Results from three large samples," *Quality of Life research* 23, pp. 2809-2818. [Cit. on p. 27.]

DeLeeuw, E. D. (2018) "Mixed-mode: Past, present, and future," in *Survey Research Methods*, 2, vol. 12, pp. 75-89. [Cit. on p. 7.]

European Commission and Eurostat (2018) *Living conditions in Europe: 2018 edition*. Publications O□ce. [Cit. on p. 8.]

Eurostat (2018) *Community Statistics on Information Society (CSIS)*. Luxembrug: Publications O□ce of the European Union. [Cit. on p. 29.]

Fornara, F., Bonaiuto, M., Bonnes, M., *et al.* (2010) *Indicatori di qualità urbana residenziale percepita (IQURP). Manuale d'uso di scale psicometriche per scopi di ricerca e applicativi*. FrancoAngeli. [Cit. on p. 27.]

Förster, A., Borschewski, K., Bolton, S., and Jääskeläinen, T. (2020) "The matter of meta in research data management: Introducing the CESSDA Metadata O□ce Project," *IASSIST Quarterly* 44(3). [Cit. on p. 51.]

Fowler Jr, F. J. and Mangione, T. W. (1990) *Standardized survey interviewing: Minimizing interviewer-related error*. Sage, vol. 18. [Cit. on p. 47.]

Fusco, A., Guio, A.-C., and Marlier, E. (2013) "Building a material deprivation index in a multinational context: lessons from the EU experience," *Poverty and social exclusion around the Mediterranean Sea* pp. 43-71. [Cit. on p. 29.]

Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., Bullinger, M., Kaasa, S., Leplege, A., Prieto, L., *et al.* (1998) "Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project," *Journal of clinical epidemiology* 51(11), pp. 1171-1178. [Cit. on pp. 30, 32.]

Gentzkow, M. and Shapiro, J. M. (2014) "Code and data for the social sciences: A practitioner's guide," *Chicago, IL: University of Chicago* [Cit. on p. 51.]

Gerosa, T., Gui, M., and Büchi, M. (2022) "Smartphone use and academic performance: a pervasiveness approach beyond addiction," *Social Science Computer Review* 40(6), pp. 1542-1561. [Cit. on p. 28.]

Gerosa, T. and Respi, C. (2022) *Ad-hoc module 1: ITA.LI COVID-19 brief report*. Milano: Institute for Advanced Study of Social Change. [Cit. on p. 15.]

Groves, R. M. (2006) "Nonresponse rates and nonresponse bias in household surveys," *Public opinion quarterly* 70(5), pp. 646-675. [Cit. on p. 17.]

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2011) *Survey methodology*. John Wiley & Sons. [Cit. on pp. 5, 6.]

Hahn, E., Gottschling, J., and Spinath, F. M. (2012) "Short measurements of personality–Validity and reliability of the GSOEP Big Five Inventory (BFI-S)," *Journal of Research in Personality* 46(3), pp. 355-359. [Cit. on p. 29.]

Jenkinson, C. and Layte, R. (1997) "Development and testing of the UK SF-12," *Journal of health services research & policy* 2(1), pp. 14-18. [Cit. on p. 30.]

Kalton, G., Kali, J., and Sigman, R. (2014) "Handling frame problems when address-based sampling is used for in-person household surveys," *Journal of Survey Statistics and Methodology* 2(3), pp. 283-304. [Cit. on p. 8.]

Kroh, M. (2007) "Measuring left–right political orientation: The choice of response format," *Public Opinion Quarterly* 71(2), pp. 204-220. [Cit. on p. 33.]

Lang, F. R., John, D., Lüdtke, O., Schupp, J., and Wagner, G. G. (2011) "Short assessment of the Big Five: Robust across survey methods except telephone interviewing," *Behavior research methods* 43, pp. 548-567. [Cit. on p. 29.]

Lipps, O., Herzing, J. M., Pekari, N., Ernst Stähli, M., Pollien, A., Riedo, G., and Reveilhac, M. (2019) *Incentives in surveys*, tech. rep., FORS, University of Lausanne. [Cit. on pp. 41, 48.]

Lucchini, M., Gerosa, T., Pancheva, M., Pisati, M., Respi, C., and Riva, E. (2021) "Di□erential e□ects of COVID-19 and containment measures on mental

health: Evidence from ITA. LI—Italian Lives, the Italian household panel," *Plos one* 16(11), e0259989. [Cit. on p. 15.]

Mudryk, W., Bougie, B., Xiao, P., and Yeung, A. (2001) *Statistical methods in quality control at Statistics Canada*. [Cit. on p. 19.]

Patel, M. D., Rose, K. M., Owens, C. R., Bang, H., and Kaufman, J. S. (2012) *Performance of automated and manual coding systems for occupational data: a case study of historical records*. [Cit. on p. 59.]

Pisati, M. (2023) *The Italian Lives Survey: Sample Design, Weighting, Variance Estimation, and Data Analysis*. Milano: Institute for Advanced Study of Social Change. [Cit. on pp. 6, 7, 9, 16, 18, 37, 49, 55, 60.]

Respi, C. and Gerosa, T. (2021) "Survey participation and non-response error in a pandemic scenario. Results from the ITA. LI Covid-19 study," *Rassegna Italiana di Sociologia* 62(1), pp. 39-65. [Cit. on p. 15.]

Rose, D., Buck, N., and Johnston, R. (1994) "The British Household Panel Study: a valuable new resource for geographical research," *Area* pp. 368-376. [Cit. on p. 2.]

Sala, E., Terraneo, M., Lucchini, M., and Knies, G. (2013) "Exploring the impact of male and female facial attractiveness on occupational prestige," *Research in Social Stratification and Mobility* 31, pp. 69-81. [Cit. on p. 34.]

Saris, W. (1988) "Variation in response functions: A source of measurement error in attitude research (Vol. 3)," *Amsterdam: Sociometric Research Foundation* [Cit. on p. 33.]

Sjödin, H., Wilder-Smith, A., Osman, S., Farooq, Z., and Rocklöv, J. (2020) "Only strict quarantine measures can curb the coronavirus disease (COVID-19) outbreak in Italy, 2020," *Eurosurveillance* 25(13), p. 2000280. [Cit. on p. 13.]

Smith, M. G., Witte, M., Rocha, S., and Basner, M. (2019) "E□ectiveness of incentives and follow-up on increasing survey response rates and participation in field studies," *BMC medical research methodology* 19(1), pp. 1-13. [Cit. on p. 41.]

Ware, J. E., Keller, S. D., and Kosinski, M. (1995) *SF-12: How to score the SF-12 physical and mental health summary scales the Health Institute*. The Health Institute, Boston, MA, New England Medical Center, Second Edition. [Cit. on pp. 31, 32.]

Ware, J. E., Kosinski, M., and Keller, S. D. (1996) "A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity," *Medical care* pp. 220-233. [Cit. on pp. 30, 32.]

Ware, J. E. and Sherbourne, C. D. (1992) "The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection," *Medical care* pp. 473-483. [Cit. on p. 30.]

Watson, M. (2015) “When will ‘open science’become simply ‘science’?” *Genome biology* 16(1), pp. 1-3. [Cit. on p. 51.]

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J., da Silva Santos, L., Bourne, P., *et al.* (2016) *The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018*. [Cit. on pp. 51, 64.]

Willis, G. (2018) “Cognitive interviewing in survey design: State of the science and future directions,” *The Palgrave handbook of survey research* pp. 103-107. [Cit. on p. 35.]

Züll, C. (2016) “Open-Ended Questions. GESIS Survey Guidelines. Mannheim, Germany: GESIS–,” *Leibniz Institute for the Social Sciences* [Cit. on p. 57.]