Journal of Official Statistics, Vol. 39, No. 1, 2023, pp. 1–26, http://dx.doi.org/10.2478/JOS-2023-0001

Characteristics of Respondents to Web-Based or Traditional Interviews in Mixed-Mode Surveys. Evidence from the Italian Permanent Population Census

Elena Grimaccia[1], *Alessia Naccarato*[2], *Gerardo Gallo*[1], *Novella Cecconi*[1], *and Alessandro Fratoni*[1]

In order to provide useful tools for researchers in the design of actions to promote participation in web surveys, it is key to study the characteristics that define the profile of a "web respondent", so that specific interventions can be planned. In this contribution, which draws on data collected during the 2019 housing population census in Italy, we define the set of familial and geographical characteristics that correspond to a greater probability that the interviewed household will choose to respond online, by estimating a multilevel model. The profile of a "computer-assisted web interview household" (CAWI-H) is then defined, on the basis of the structural characteristics of this population. Moreover, the geographical distribution of households is studied according to their distance from the CAWI-H profile. The results show that households that are more distant from the CAWI-H profile have characteristics that correspond to segments of the population generally affected by economic and social fragility; they are mainly elderly, foreigners, residents in small towns, and people with a low level of education. It is to these households in particular that survey designers can address specific actions that can enhance their willingness to participate in web surveys.

Key words: Mixed-mode surveys; respondent profiling; multilevel models; computer assisted web interview (CAWI).

1. Introduction

In recent decades, developments in information technology and the ever-increasing availability of administrative data have led several European countries to develop innovative methods for their population censuses (Eurostat 2020).

Mixed-mode surveys have been adopted by necessity in survey practice (Biemer 2010; De Leeuw 2005) and are, nowadays, also essential for National Statistical Institutes (NSIs). According to Tourangeau (2017) their usage is only expected to increase over time.

Online surveys have become part of mixed-mode data collection strategies, since web data collection presents several advantages, like a reduction in cost, a general improvement in timeliness (since the higher the share of web respondents, the shorter the time devoted to data collection as in Dillman et al. 2014 and in De Leeuw 2005), and a potential to use complex questionnaires (De Leeuw and Berzelak 2016), and to reduce

[1] Istat, Italian National Institute for Statistics, via Cesare Balbo 16, 00184 Rome, Italy Emails: elgrimac@istat.it, gegallo@istat.it, ncecconi@istat.it, and fratoni@istat.it
[2] Department of Economics, Roma Tre University, via Silvio D'Amico 77, 00145 Rome, Italy. Email: alessia.naccarato@uniroma3.it
Acknowledgments: We would like to thank the anonymous reviewers whose suggestions and comments helped improve and clarify this manuscript.

coverage and nonresponse error (Bianchi et al. 2017; Brick and Tourangeau 2017; Cobben and Bethlem 2013; De Leeuw 2018; De Leeuw et al. 2019; Luiten et al. 2020).

In particular, sequential mixed-mode strategies (which offer one mode at a time, starting with the cheapest and quickest) improve response rates when different modes acquire different types of respondents: typically, younger respondents respond through the web, and respondents who are older or first-generation immigrants respond in interviewer administered modes (Benzeval et al. 2021; De Leeuw and Berzelak 2016; Kappelhof 2015). Indeed, including a mail survey for those who do not have internet access and as a follow-up for web nonrespondents also improves representativity on demographic and attitudinal variables (Bandilla et al. 2014; Messer and Dillman 2011).

However, there are some disadvantages because of the comparability of data collected via different modes, which could lead to different measurement errors. For instance, past research has shown that self-administered forms (e.g., both paper mail surveys and web surveys) perform better when more sensitive questions are asked allowing less social desirable answers than interview surveys (De Leeuw and Berzelak 2016; De Leeuw et al. 2008, 299).

In this framework, no one could have predicted that the coronavirus pandemic emergency and the impossibility of carrying out field operations would have boosted the need for web-based interviews with such strength and urgency.

However, even if the use of the internet is quite widespread in Europe (Eurostat 2020), respondents' attitudes toward web surveys cannot be taken for granted. Therefore, planning interventions targeting specific groups of individuals or households can support wider participation in web surveys.

The analysis of auxiliary variables in mixed-mode surveys has certainly become much more important (De Leeuw 2018; De Leeuw et al. 2019), both because of the digital divide in the population, with highly educated and younger people more often having an internet connection (Biffignandi and Bethlehem 2012; Couper 2017; Mohorko et al. 2013), and because the interaction between the characteristics of the respondents and the decisions of the survey designer influences the response rate and the success of the survey (Antoun et al. 2017; Biffignandi and Pratesi 2002; Durrant and Steele 2009; Scherpenzeel and Bethlehem 2011, 105).

In this article, we study the determinants that influence the cooperation of respondents with a sequential mixed-mode survey, such as the Italian Permanent Population and Housing Census (PPHC), in order to point out the specific characteristics of the population. This could enhance the efficacy of the actions of survey designers to improve participation in web surveys.

By estimating a multilevel logit model (Durrant and Steele 2009; Mohorko et al. 2013), we distinguish between the features of online respondents to a Computer Assisted Web Interview (CAWI) and the features of respondents in traditional modes (in the case of the PPHC, a Computer Assisted Personal Interview, CAPI), given that respondents with different characteristics tend to have different propensities for being interviewed in a given mode.

We consider the set of the socio-economic characteristics of a respondent as their "profile". The profiling of respondents, in the literature, has traditionally been achieved through a logistic model. Among others, Maslovskaya et al. (2019) studied the effect of different response tools in six social surveys in the UK. The results of their bivariate analysis suggest that, across the surveys, age, gender, marital status, employment, religion, household size, children in household, household income, number of cars and frequency of internet use are significantly associated with the device used. Bianchi et al. (2017) studied the effects of mixed-mode design on participation rate, sample composition, and costs in a longitudinal survey in the UK. To investigate whether the mixed-mode design had different effects for different characteristics of the respondents (such as gender, age, race, working status, household type, urbanicity, web-user), the authors estimated a logit model employing individual characteristics and interactions, but this did not produce significant results. Pratesi et al. (2004) focused on the timeliness and quality of web surveys in Slovene households. Their findings suggest that nonresponse rates and quick reactions in web surveys are sensitive to individuals' characteristics, but that demographics are not significant in this. Methodological issues and the research perspectives of web surveys have been studied in depth and in a thorough and comprehensive way by Biffignandi and Bethlehem in 2012 and, more recently, in 2021.

The analysis presented here, then, explores the significant differences, in terms of individual and familial characteristics, between households who respond by completing a CAWI and those who respond by the face-to-face method. In this way, it is possible to define the profile of CAWI respondents (which we call the CAWI-H profile) to identify which household characteristics contribute the most to increasing the probability of responding via the web. As in previous research (Durrant and Steel 2009; Mohorko et al. 2013), this study employs multilevel modelling alongside logistic regression, and it provides also some evidence that a multilevel model is better than a logit model at explaining the probability of answering via the web.

Once the profiles of web respondents have been defined, it is possible to measure how far the features of CAPI respondents deviate from the CAWI-H profile, establishing a measure of distance that allows us to classify respondents according to their "resistance" to web completion. Compared to other studies employing propensity scores, the computation of distances from the profiles identified by the multilevel model has been preferred since it could be successfully employed for other surveys and used as a tool for planning specific survey interventions.

This study identifies the territorial features related to web responses providing specific tools for promoting web surveys and, in particular, addressing the logistical issues. Compared to previous studies concerned with Internet usage in general, this study provides insight on web survey.

The article is organised as follows: Section 2 presents the data collection design for the PPHC; Section 3 describes the variables employed to define the respondents' profiles; Section 4 presents the statistical methodology to obtain the respondents' profiles and to measure the distances between them; Section 5 presents the main outcomes of the study and the territorial distribution of the different types of respondents; and, finally, Section 6 presents some concluding remarks with reference to particular population groups that could be made the subject of actions aimed at improving the CAWI response.

2. Italian Population and Housing Census Data Collection Design

This article draws on data from the 2019 PPHC. In Italy, the PPHC, which was started in 2018 by the Italian National Institute of Statistics (Istat), currently provides a mixed-mode survey, since respondents may choose to fill in the questionnaire via the web or by the traditional way, according to the data collection design described in Figure 1.

In Figure 1, the flowchart for the data collection process is presented. The current progression of the actions for data collection are represented in white, while our proposal to increase the share of web respondents is shown in grey.

The population census provides the official estimate of the resident population at the municipality level (Righi et al. 2021). In Italy, as in most European countries, an assessment of the coverage results of the last population census in 2011 and the high costs of data collection led to a change in the strategy for the population census from the traditional door-to-door enumeration (every ten years for the whole population) to a yearly register-based sample survey, combining the use of administrative sources with annual surveys (Chieppa et al. 2018; Citro 2014; Crescenzi 2015; Righi et al. 2021). A further advantage of these data collection methodologies is the greater containment of the census participation burden: as noted by UNECE (2018), the response burden on the population is lower with a combined census, both because the number of questions is reduced (since some information is available from the population register), and because the physical presence of a stranger in people's homes is avoided (which could be a reason for a refusal to participate, especially in a pandemic situation).

The PPHC foresees two different yearly sample surveys: an Areal sample and a List sample. The first is only used to update the population register and it is only conducted face-to-face; it is not of interest for a comparison of different data collection modes. Therefore, we analyse the List sample, which presents a most interesting survey design and is carried out in order to collect information on socio-economic issues that is not available in the population registers. The List sample is based on a yearly sample size of about 950,000 households (Table 1). Around 2,400 municipalities (of the 7,904 municipalities in Italy) are involved in the survey every year.

The List survey design is "sequential", and respondents may choose the most suitable mode to respond to the questionnaire. The first survey mode proposed by the survey researcher is the CAWI mode, since it is the cheapest and quickest. Helplines are provided to the respondents in case they need more information or specific support in answering to the survey. If the CAWI is not answered, respondents are contacted, in order to recall the interview and to offer help in the filling of the questionnaire. However, as a second option, respondents can go in person to the many offices that are present in each municipality and that are devoted to help respondents, and to conduct a "face to face" interview. Finally, as a last option, it is foreseen also the traditional interviewer administered survey at home, if required by the respondent for any reason (Istat 2018). The share of CAWI responses is therefore unknown "a priori".

Data are collected at the household level, as a unique questionnaire is provided to collect the data. Households included in the List sample receive an official letter from Istat, giving information about the importance of the survey and the fact that answering is mandatory by law, and providing the credentials for answering the census questionnaire online. The letter is addressed to the oldest member of the household, who is assumed to be the one who will answer the questionnaire, providing information on the whole household. The questionnaire is written in Italian, but a complete guide is provided in 14 foreign languages.

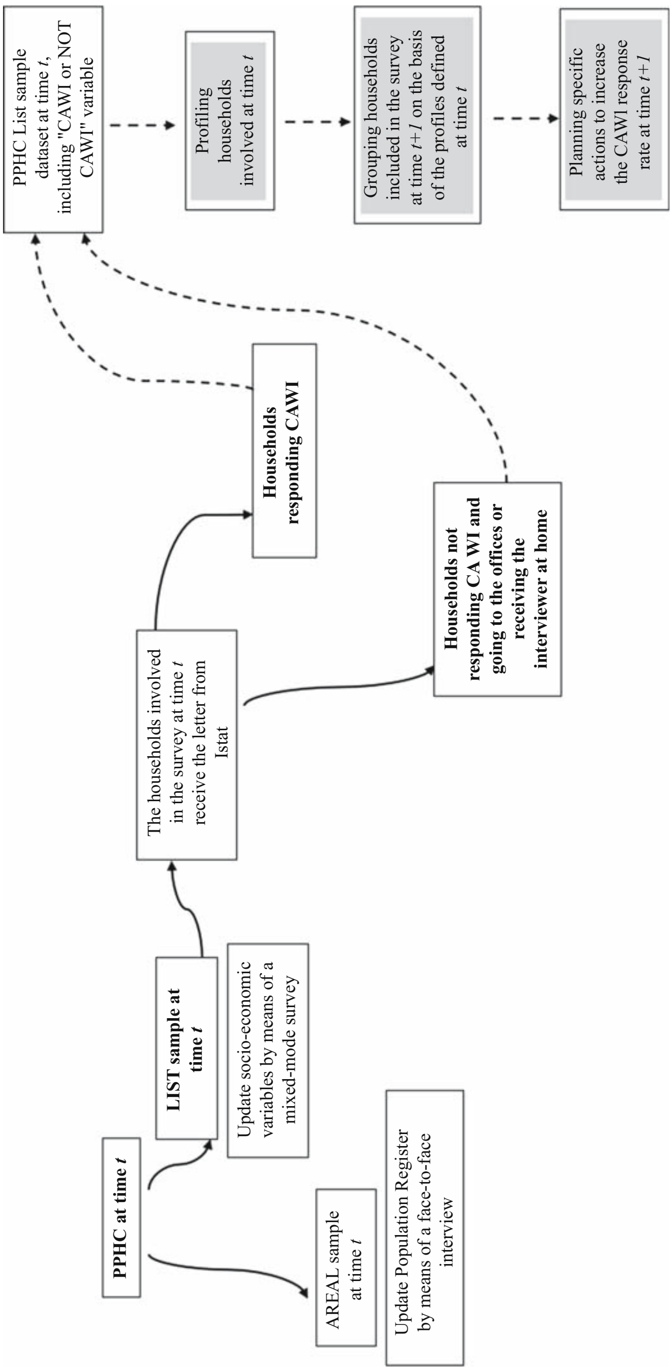

Fig. 1. Italian population and housing census data collection design.

Table 1. Sample size – municipalities and households by population size of municipalities – 2019.

Demographic size of Italian municipalities (number of inhabitants)	Municipalities	Households
Up to 5,000	1,028	190,458
From 5,001 to 20,000	815	288,360
From 20,001 to 50,000	376	184,593
From 50,001 to 100,000	98	120,149
From 100,001 to 250,000	34	94,226
More than 250,000	12	73,253
Total	**2,363**	**951,039**

Source: 2019 PPHC-Istat

It is only if the household cannot fill in the CAWI that they can answer in a face-to-face interview that must take place in a public office responsible for data collection.

Under the auspices of NSIs, this share should be close to 100%. In Italy, the use of the internet is quite widespread, with 76% of households having internet access and 75% having a broadband connection in 2019. However, there is still a large digital divide between households, mainly due to generational and cultural factors. Almost all households with at least one member younger than 18 years old have a broadband connection (95%), while among households composed exclusively of people over the age of 65, this share falls to 34% (Istat 2019). Since the share of CAWI respondents is stuck at 50% of the population, the diffusion of the internet would suggest the possibility of a much higher share of CAWI responses, provided that targeted actions such as those identified in this study are conducted on the profiles of respondents.

Although the PPHC is not a panel, since households are included in the sample only once, the profiling of web respondents could be of great support in the successive data collection occasions, because the survey is repeated over time with the same design.

3. Data

The PPHC List sample, for each household, provides information on the survey mode, the socio-demographic characteristics of the household, and the geographical features of the municipality where the household resides.

The socio-demographic information available on households is household size, household citizenship, age of youngest household member, and highest educational level in the household. The household size indicates the number of people residing in the same dwelling. The citizenship variable classifies the households based on the citizenship of the members: all foreigners, all Italians, or a mixed citizenship household. The youngest household-member age variable indicates the age of the youngest member of the household; this variable was considered in the analysis on the basis of the hypothesis that the presence of young individuals in a household facilitates the use of information technology and therefore has an effect on the decision to fill in the questionnaire online. Minors are not included in the definition of this variable since the census questionnaire can only be filled in for members who are over 18 years of age. The highest educational qualification held by at least one member of the household is classified according to the primary, secondary and tertiary levels that group together the grades (0, 1, 2), (3, 4), and (5, 6, 7, 8) of the International Standard Classification of Education (ISCED).

All these features are coherent with the predictors accounted for in previous studies. Besides the literature cited in the Introduction, it is worth mentioning the work of Hargittai (2002), which is one of the first studies on the profiling of online skills, employing personal and familial features, such age and education. Similar predictors were chosen by Durrant and Steele (2009).

Furthermore, in order to take into account the strong territorial disparities, both in socio-economic issues and in digitalization (Benassi and Naccarato 2017; Cellini and Torrisi 2014; Cracolici et al. 2007; De Clercq et al. 2020; Santarelli and Cottone 2009), several characteristics of municipalities are considered in the analysis: local capital or metropolitan city, municipality population size, municipality degree of urbanisation, and altitudinal zone. Local capitals or metropolitan cities are those municipalities which are the seats of their boroughs. The variable 'municipality population size' classifies municipalities on the basis of the number of their inhabitants, from very small municipalities (up to 5,000 inhabitants) to very large ones (more than 250,000 inhabitants), while the degree of urbanisation is a classification of municipalities introduced by Eurostat and based on a criterion of geographical contiguity and on minimum population thresholds (European Union 2017). The territory is classified as one of three types of area: (1) a densely populated area (city or large urban area), defined as clusters of contiguous cells of 1 km^2, with a density of not less than 1,500 inhabitants per km^2 and a population of not less than 50,000 inhabitants; (2) an area with an intermediate density level (or small urban area), defined as clusters of contiguous cells with a density of not less than 300 inhabitants per km^2 and a population of not less than 5,000 inhabitants; and (3) a sparsely populated area (or rural area), defined as single cells (rural) not classified in the previous groups (Eurostat 2021). In order to attribute this classification to single municipalities, the areas identified according to the degree of urbanisation are compared with the municipal boundaries. The altitudinal zone defines homogeneous areas constituted by aggregating contiguous municipalities on the basis of threshold values for the altitude (elevation above sea level). According to this classification, mountain, hill, and plain areas can be identified. The mountain and hill areas are divided into inland mountain and inland hill areas and coastal mountain and coastal hill areas, respectively, in order to take account of the moderating action of the sea on climate. The altitudinal zone is taken into account in the analysis since it could affect the choice to answer online due to the different internet access capacity of mountainous areas, which presents some specificities regardless of the population density and the composition of population (Reynaud et al. 2020).

Municipalities can be grouped into regions, which are very important territorial units that are responsible for various economic and social policies, or into macro-regions: the North, the Centre, and the southern areas of Italy plus the two major islands, together traditionally referred to as the "Mezzogiorno".

Table 2 shows descriptive statistics on the frequency distribution of the households according to the CAWI or not CAWI survey mode and the explanatory variables included in the analysis (Section 5). The share of CAWI respondents varies according to the household citizenship, size, age of the youngest member, and highest education level in the household. The size and sign of the relationships are presented (when significant) in Subsection 5.1. At this point of the study, however, it is worth mentioning that for instance

Table 2. Households by survey mode, and familial and social characteristics[(a)].

Household characteristics		Survey mode: Not CAWI (%)	CAWI (%)	Composition of the CAWI sample (%)	Number of cases
Household size	1	54.01	45.99	27.4	245155
	2	50.98	49.02	28.6	239632
	3	45.67	54.33	21.6	163624
	4	43.93	56.07	17.6	128977
	5 or more	56.16	43.84	4.8	25294
	Pearson Chi2(4)=5.50E+03; Pr=0.000				
Household citizenship	All foreigners	75.27	24.73	1.7	28296
	All Italians	48.86	51.14	95.9	770700
	Mixed citizenship	57.29	42.71	2.4	23431
	Pearson Chi2(2)=8.10E+03; Pr=0.000				
Youngest household-member age	18-34	46.98	53.02	31.4	243631
	35-64	47.23	52.77	45.8	356701
	65+	57.78	42.22	22.8	222095
	Pearson Chi2(2) = 7.40 + E03; Pr = 0.000				
Household highest educational level	Primary	65.03	34.97	24.9	293213
	Secondary	46.59	53.41	44.1	339277
	Tertiary	32.90	67.10	31.0	189937
	Pearson Chi2(2)=5.00E+04; Pr=0.000				
Total		50.01	49.99	100.0	822427

Source: 2019 PPHC-Istat

[(a)]There are no missing cases in the data set for the variables used here.

the share of CAWI respondents increases with higher level of education. The Chi-square tests referring to these distributions preliminarily suggest that there are relationships between the type of response and the variables used to define profiles.

4. Methods

The first step of the research is to identify the profile of the online respondent households (CAWI-H) and, more specifically, to identify the variables that are associated with the households with the highest probability of responding via the web.

To compute the probability that a family responds in CAWI mode, we estimate a multilevel logit model that also takes into account territorial specifications (Goldstein 2010). The endogenous variable is the dichotomous variable that assumes the value 1 if the family responded in CAWI mode, and 0 otherwise (Durrant and Steele 2009; Keusch et al. 2019; Mohorko et al. 2013), the binary response Y_i defined as follows:

$$Y_i = \begin{cases} 0 & Not\ CAWI\ response \\ 1 & CAWI\ response \end{cases} \tag{1}$$

where i $(i = 1, \ldots, n)$ denotes the household.

The auxiliary variables are the social and demographic characteristics of the households illustrated in Section 3.

The multilevel model can be written as:

$$log\left(\frac{\pi_{ij}^{(s)}}{\pi_{ij}^{(0)}}\right) = \boldsymbol{\beta}^{(s)}\boldsymbol{x}_{ij}^{(s)} + u_j^{(s)} \tag{2}$$

where j ($j = 1, \ldots k$) indicates the territorial level, $\pi_{ij}^{(s)} = \Pr(y_{ij} = s)$, $s = \begin{cases} Not\ CAWI \\ CAWI \end{cases}$, $\boldsymbol{x}_{ij}^{(s)}$ is a vector of household level covariates, $\boldsymbol{\beta}^{(s)}$ is a vector of coefficients, and $u_j^{(s)}$ is a random effect representing unobserved regional characteristics.

By means of the $u_j^{(s)}$ component, the multilevel model (2) takes into account the relationship between a CAWI response and the region where the household resides. In this way, the model acknowledges the unobserved regional influences on the different types of response, and the estimates of the coefficients referring to the explanatory variables are more accurate. The use of random effects is justified by the fact that the respondent living in a region is a feature that does not change over the period accounted for in the model.

As a further check on the robustness of the results and to illustrate their stability to various model specifications, we test whether a model including region fixed effects fits the data better than the multilevel model. The estimated logistic regression with binary response Y_i in Equation (1) (Wooldridge 2012) is:

$$logit(g_i) = \log\left(\frac{g_i}{1 - g_i}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots \tag{3}$$

in which the probabilities $g_i = \Pr(Y_i = 1|x_i)$ are related to a linear predictor $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots$ through the logit function.

In Appendix (Section 7), we present the results of the logit model with different specifications, including joint effects of the explanatory variables.

In this first phase of the research, the profile of the households with the highest probability of answering via the web is identified. All in all, the number of possible combinations of categories (i.e., profiles) is 135, since there are four exogenous variables used in the model (2): household size, with five categories; household citizenship, with three categories; youngest household-member age, with three categories; and household highest educational level, with three categories (Table 2). Among the 135 profiles, we give the name 'CAWI-H' to the profile that presents the highest probability of answering via CAWI, that is to say, the combination of the categories that presents the highest values of the multilevel model coefficients. Each household, then, presents a certain profile because it differs from the CAWI-H profile for one or more categories. Therefore, it is possible to establish a measure of the distance between the profile of a household and the CAWI-H profile, and to compute this distance for each household.

The maximum possible value for this distance is 4 (if all the explanatory variables lie in a category that is different from the CAWI-H profile), while the minimum value is equal to 0 if the household presents the CAWI-H profile. The higher the value of the distance, the lower the probability that the specific family type will respond to on the web.

Since the variables that define the profiles are all categorical, the measure employed is the Jaccard distance (Jaccard 1908; McCormick et al. 1992), defined as follows:

$$D_i = \frac{M_{c,i}}{M_1 + M_2 + M_3 + M_4} \tag{4}$$

where i corresponds to a single household, $M_k, k = 1, \ldots, 4$ is the number of categories for each categorical variable, and $M_{c,i}$ is the number of categories for which the CAWI-H profile and the profile of household i are different. For instance, if household i has the same size and age of the youngest member as CAWI-H, but different citizenship and level of education, then $M_{c,i}$ is equal to 2.

To obtain a measure that varies between 0 and 1, the relative distance D_{Ri} is calculated by dividing the value D_i in equation [4] by the maximum value.

Households with the same value of the distance D_{Ri} have similar characteristics and, thus, the same probability of answering by a CAWI.

One of the most common methods for controlling for differential nonresponse in mixed-mode surveys is to use the estimated propensity of a respondent to participate in each mode. These propensity scores are typically estimated from a generalised linear model (e.g., probit, logit), where a given mode is treated as a possible outcome, conditioned on available covariates information, often limited to participant demographics, as is done by Hox et al. (2015), and more recently by Maslovskaya et al. (2019) and Rivero et al. (2019). However, in this article, we propose a procedure that would allow survey design researchers to control the distance from the CAWI-H profile in the actual data collection and also before proceeding to a new phase of data collection, while propensity scores can only be estimated after collecting the data.

In order to plan interventions for the promotion of web surveys in different territories, it is useful to study where households present the strongest 'resistance' to CAWI and the characteristics of the areas of the country in which it is more important to intervene. The third and final step of the article is then to create and analyse a measure of 'Municipality Resistance' to CAWI data collection. The Municipality Resistance (MR) is computed as the average of the distances (D_{Ri}) for all the households residing in a given municipality. In this way, this average can be considered as the score assigned to the particular municipality. Using the MR as an endogenous variable in a regression model in which the covariates are the geographical variables (Section 3), it is possible to identify the characteristics of municipalities that are predictors of the MR. The characteristics of the territories employed in the analysis are all variables associated with the penetration of the internet (Ciapanna and Roma 2020; Rados 2021).

5. Results

The results provided by the multilevel model are presented in Subsection 5.1. In Subsection 5.2, the distribution of households, based on their distance from the CAWI-H profile, is illustrated. Finally, in Subsection 5.3, the estimation of a model that identifies the geographical predictors of municipalities' ability to collect responses via the web is presented.

5.1. Determinants of a CAWI Response

As illustrated in Section 4, a multilevel binomial model is estimated to explore the effects of household characteristics on the probability of CAWI or not CAWI responses. More specifically, the CAWI-H profile is a household whose members are all Italian, with at least one member with a tertiary level of education, in which the youngest member is between 35 and 64 years old, and, to a lesser extent, having four members (Table 3).

The highest coefficient for a CAWI response refers to households in which the level of education is highest. For people with a secondary or tertiary education, the coefficients referring to the probability of answering via a CAWI method are significantly higher than for those who have a primary level (respectively 0.734 and 1.350).

Moreover, households exclusively composed of foreigners present a lower probability of using the CAWI option than those with at least one Italian member. In particular, households whose members are all Italian present a coefficient equal to 1.379.

The empirical evidence also suggests that the CAWI response rate of households where the age of the youngest member is between 35 and 64 years old is significantly higher than the rate of households where the age of the youngest member is between 18 and 24 years old. Households composed only of elderly people present the lowest probability of a CAWI response.

Table 3. Estimate of parameters of multilevel logit model.

CAWI	Coefficient	Std. Err.	P > |z|
Household size (Base = 1)			
2	0.012	0.006	0.05
3	0.018*	0.008	0.02
4	0.112***	0.008	0.00
5 or more	-0.233***	0.012	0.00
Household citizenship (Base = All foreigners)			
All Italians	1.379***	0.015	0.00
Mixed	0.711***	0.020	0.00
Youngest household-member age (Base = 18–34 years old)			
35–64	0.050***	0.006	0.00
65+	-0.045***	0.008	0.00
Household highest educational level (Base = Primary)			
Secondary	0.734***	0.006	0.00
Tertiary	1.350***	0.007	0.00
Constant	-2.090***	0.016	0.00
Region			
var(_cons)	3.50E + 08	3548290	
LR test vs logistic model: Chibar2(01) = 40684.27, Prob > = Chibar2 = 0.0000			

Legend: * p < .05; ** p < .01; *** p < .001
Source: 2019 PPHC-Istat.

The analysis of the size of the household offers more equivocal results: the coefficient referring to a CAWI response is the same in households with up to three members, it increases for households with four members, while it significantly decreases in households with five or more members.

All in all, the survey designer should pay particular attention to households with foreigners, those without any members with at least a secondary-level education, and households whose members are all elderly.

Finally, the estimate indicates that the effect of education in choosing a CAWI prevails over the other determinants. The results are coherent with those of Keusch et al. (2019), Maslovskaya et al. (2019), and Mohorko et al. (2013). In fact, education in general been found to be among the determinants of response in surveys, with a higher level linked to a higher propensity to answer through the CAWI option (Hox et al. 2015).

It is worth noting that the results are coherent with those obtained by estimation of the logit model in Appendix.

We found that the likelihood-ratio test (LR), which evaluates the goodness of fit of two comparable models, justifies the use of a multilevel model (Table 3). Indeed, the result confirms the validity of the multilevel model, compared to the logistic regression, providing a significant coefficient (Maddala and Lahiri 2010).

5.2. Households' Distance from the CAWI-H Profile

Tables 4 and 5 show the distribution of households by survey mode and their relative distance D_{Ri}.

The large majority (69%) of households that present the CAWI-H profile ($D_{Ri} = 0$) do indeed answer through a web interview. By contrast, the households that are most distant from the identified CAWI-H profile ($D_{Ri} = 1$) typically answer via a traditional mode. The dependence between the distance from the CAWI-H profile and the survey mode is significant, according to the Chi-square test results (Table 4). Altogether, the probability of answering through a CAWI questionnaire decreases as the distance from the CAWI-H profile increases. Therefore, the results suggest that our models select useful categories that identify the characteristics of respondents with the highest probability of a CAWI response. In this way, we also recognise those profiles for which contact with the

Table 4. Household distribution by type of response and relative distance from CAWI-H profile.

Relative distance from CAWI-H	Survey mode		
	CAWI	NOT CAWI	Total
0	69.11	30.89	100
0.25	54.70	45.30	100
0.5	47.42	52.58	100
0.75	45.89	54.11	100
1	27.35	72.65	100
Total	49.99	50.01	100
Pearson Chi2(4)=1.30E+04; Pr=0.000			

Source: 2019 PPHC-Istat.

Table 5. Household distribution by survey mode, relative distance from CAWI-H, and macro-region.

		Survey mode		
Macro-region	Relative distance from CAWI-H	CAWI	NOT CAWI	Total
North	0	77.27	22.73	100
	0.25	63.25	36.75	100
	0.5	55.83	44.17	100
	0.75	58.61	41.39	100
	1	29.56	70.44	100
	Total	58.91	41.09	100
	Pearson Chi2(4) = 7.30E+03 Pr=0.000			
Centre	0	70.84	29.16	100
	0.25	58.46	41.54	100
	0.5	50.65	49.35	100
	0.75	51.40	48.60	100
	1	29.83	70.17	100
	Total	53.77	46.23	100
	Pearson Chi2(4) = 2.70E+03; Pr=0.000			
Mezzogiorno	0	56.83	43.17	100
	0.25	39.88	60.12	100
	0.5	35.06	64.94	100
	0.75	31.81	68.19	100
	1	16.86	83.14	100
	Total	36.55	63.45	100
	Pearson Chi2(4) = 3.90E+03; Pr=0.000			

Source: 2019 PPHC-Istat.

respondents (often under the control of the survey designer) like an e-mail solicitation plan, survey awareness, and so on, is most necessary and useful.

The study of the different profiles and their distance from the CAWI-H profile allows us to highlight two situations: those families for which a CAWI response is almost certain (0) or very unlikely (1), and those households that are on the border between being CAWI and being not CAWI. Considering only the households presenting the lowest relative distance from the CAWI-H profile (e.g., the first two groups in Table 4), the gain in the share of web respondents that it would be obtained if all these HHs would respond in a CAWI mode is 7.4 percentage points. Therefore, the share of CAWI responses would increase from 49.99 to 57.37%, if all the HHs with a distance from CAWI-H lower than 0.5 would actually respond via web.

The geographical effects that appear in the models to be significant for the probability of responding through a CAWI method (Table 3) become apparent in the analysis of CAWI-H distances (Table 5).

In the North and Centre of the country, the households that have a distance of 0.5 from the CAWI-H profile also have a 50% probability of answering through a web questionnaire. By contrast, in the Mezzogiorno the probability is much lower (35%). This means that even households with a CAWI-H profile have more difficulty in answering via the web. Moreover, in the Mezzogiorno area 43% of households with a CAWI-H profile do not choose to answer via the web. Therefore, the survey designer should pay more attention to the less developed areas.

This methodology can be applied to any country, since it is widely known that territorial differences, such as the urbanisation or geo-morphology of the territory, can be important predictors of digitalisation and other socio-economic features.

5.3. The Geographical Distribution of Municipalities with Web Respondents

The MR to CAWI mode increases when moving from the municipalities of the northern regions of Italy to the central areas and increases again when moving to the municipalities of the southern regions, as shown in Figure 2.

To verify the effects of the geographical characteristics of the municipalities on the MR variable, a generalised linear model (GLM) is estimated. Table 6 shows the results of the estimated model, where the endogenous variable is the MR, and the exogenous variables are the geographical variables (Section 4).

The demographic dimension of the municipality is important in explaining the MR to a CAWI response, and it is negatively related to the distance from the CAWI-H profile. Therefore, small centres with fewer than 20,000 inhabitants should be the focus of the survey designer's attention. In particular, local capitals or metropolitan cities present the highest probability of the CAWI-H profile. Indeed, internet users generally experience faster download and upload speeds in urban areas, although there is high variability in the internet coverage in particular territories (Rizzato 2020).

The results shown in Table 6 indicate that the geographical differences among the macro-regions are significant; thus, the model is also estimated separately for the North, the Centre, and the Mezzogiorno, in order to find out the specific details for each macro-region. In the Mezzogiorno, the MR is significantly higher (0. 0186).

The degree of urbanisation has no effect on the MR in the North and Centre of Italy, while in the Mezzogiorno small towns and rural areas are instead related to a smaller

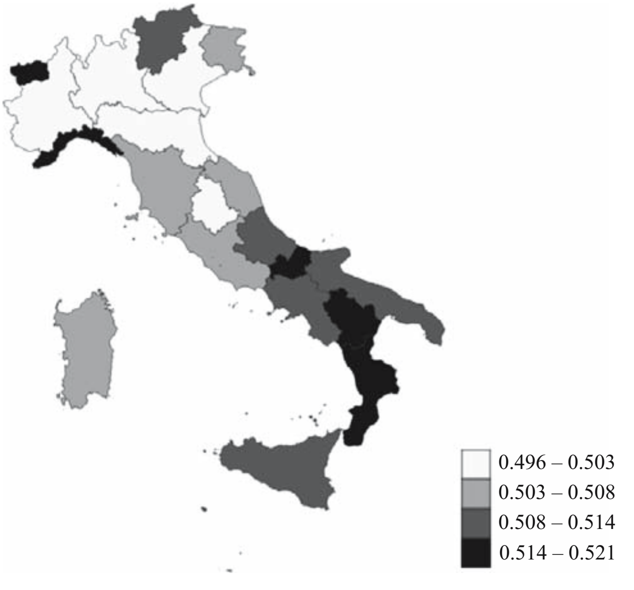

Fig. 2. Distribution of "Municipality Resistance" (MR – Regional Averages), Italy 2019.

Table 6. Estimate of parameters for generalised linear model and standard errors (in parentheses).

Municipality Resistance	Items	Italy	North	Centre	Mezzogiorno
Population size (Base = Less than 5,000 inhabitants)					
	From 5,001 to 20,000 inhabitants	0.0023	0.0013	-0.0017	.0073*
		(0.0015)	(0.0019)	(0.0038)	(0.0031)
	From 20,001 to 50,000 inhabitants	-.00437*	-.0104***	-0.0067	0.0048
		(0.0020)	(0.0028)	(0.0052)	(0.0038)
	From 50,001 to 100,000 inhabitants	-.00754*	-.0140*	-0.0128	-0.0020
		(0.0037)	(0.0064)	(0.0078)	(0.0058)
	From 100,001 to 250,000 inhabitants	-0.0115	-0.0166	0.0050	-0.0156
		(0.0064)	(0.0094)	(0.0160)	(0.0108)
	Over 250,000 inhabitants	-.01958*	-.0268*	-0.0400	0.0047
		(0.0088)	(0.0126)	(0.0211)	(0.0154)
Altimetric area (Base = Inland mountain)					
	Coastal mountain	-0.0038	-0.0186	0.0141	0.0029
		(0.0045)	(0.0128)	(0.0147)	(0.0055)
	Inland hill	.00374*	0.0029	0.0044	.0077**
		(0.0016)	(0.0022)	(0.0038)	(0.0029)
	Coastal hill	-.00422*	-.0218***	-0.0032	0.0004
		(0.0020)	(0.0043)	(0.0044)	(0.0031)
	Plain	.0114***	.0115***	0.0007	.0133***
		(0.0016)	(0.0019)	(0.0060)	(0.0033)
Municipality urbanisation degree (Base = Densely populated city or area)					
	Small town and suburb or intermediate population density zone	-.01047***	0.0015	0.0054	-.0274***
		(0.0025)	(0.0034)	(0.0125)	(0.0039)
	Rural area or sparsely populated area	-.00832**	0.0047	0.0029	-.0231***
		(0.0029)	(0.0037)	(0.0132)	(0.0050)
Local capital/Metropolitan city (Base = No)					
	Yes	-.0355***	-.0212***	-.0246**	-.0497***
		(0.0038)	(0.0061)	(0.0082)	(0.0061)
Macro-region (Base = North)					
	Centre	.0051**			
		(0.0017)			
	Mezzogiorno	.0186***			
		(0.0013)			
Constant		.4601***	.4504***	.4537***	.4869***
		(0.000)	(0.039)	(0.0138)	(0.054)
R^2		0.1114	0.168	0.1437	0.1983
Chi2		542.99	236.65	56.539	201.31
AIC		-10554	-5372.6	-1612.5	-3646

Legend: * $p < .05$; ** $p < .01$; *** $p < .001$
Source: 2019 PPHC-Istat.

distance from the CAWI-H profile. In the Mezzogiorno, the effect of being a local capital is maximal: living in one of the principal towns makes a real difference with respect to the probability of a CAWI response.

6. Final Remarks

For any statistical survey, web interviews imply considerable advantages: first of all, a reduction of costs, and then the containment of the interviewer effect (Bethlehem et al. 2011; Scherpenzeel 2011), and the timeliness of data collection. These advantages become even more significant when reference is made to official statistical surveys because of the large size of the samples and the number of variables surveyed, as well as the high standards that must be guaranteed.

The interest in identifying the profile of CAWI respondents is that web surveys still face some resistance from the population. In order to foster a positive attitude towards web interviewing, an awareness campaign is necessary, and this would be more effective if it is targeted at a specific population.

It is also necessary to adapt the data collection strategy to geographical imbalances. For instance, small centres, rural areas, and some remote and mountainous zones show the highest resistance to filling in census questionnaires online and should be made target of specific actions by the survey designer.

All over the world, NSIs employ population lists to design surveys, and to define samples. In these population or firms' registers, some characteristics are included, and the amount of information available in advance is further increasing because the huge development of registers and archives. The variables employed in this study are those available in the lists of population that are used before the survey. Therefore, these variables are known in advance and they can be employed by the survey researchers to design specific actions to enhance web responses before the actual data collection. This study thus proposes a procedure to better exploit this information that is available before the survey.

This contribution has shown how some structural characteristics of households allow us to classify them on the basis of an attitude: a preference for filling in the population census questionnaire online, and that this attitude refers to household that generally are classified as living in less fragile conditions (Benassi and Naccarato 2017; D'Ambrosio 2003).

In conclusion, families with a higher level of education, composed of Italian citizens, with at least one member younger than 65 years old, and who live in urban areas are those for whom the probability of answering via the web is highest; they are also those who historically live in better economic and social conditions. We can therefore conclude that, regardless of the advantages of a statistical survey, the probability of answering online can be considered as a further sign of the conditions of greater or lesser social disadvantage in which Italian families live.

Other than providing some new insights on the profiling of web respondents in mixed-mode surveys, this research offers a procedure for calculating distances between profiles that could be successfully employed for any other survey and used as a tool for planning specific survey interventions to enhance web survey participation.

The classification of households in profiles identifies more homogeneous population groups, and it could be useful in the design of successive survey occasions, with the aim of reducing the variability of the estimates and, eventually, the sample size. Also, the correction of partial or total nonresponse data would benefit from the profiling of households, through the analysis of more homogeneous groups.

Nevertheless, further research could be devoted to an analysis of the stability over time of the profiles and their geographical distribution. Having identified territorial differences in the profiles, also identifying regional profiles could be an insight to further develop.

7. Appendix

Tables 7 and 8 show the results of a base model (M0) that takes into account only the socio-economic characteristics of the households as described in Section 4; a model (M1) that also considers the geographical macro-region variable; and models from model (M2) to (M6) that include the variable indicating the region of residence and the joint effects of the explanatory variables.

Table 7. Estimate of parameters of logit model (odds ratios and standard errors).

	M0		M1		M2		M3	
Variable	Odds ratios	SE	Odds ratios	SE	Odds ratios	SE	Odds ratios	SE
2	0.013*	.006	0.014*	.006	0.011	.006	-0.030	.019
3	-0.022**	.007	0.023**	.008	0.017*	.008	-0.047**	.017
4	0.009	.008	0.110***	.008	0.109***	.008	-0.050**	.018
5 or more	-0.349***	.012	-0.240***	.012	-0.237***	.012	-0.382***	.020
All Italians	1.158***	.014	1.380***	.015	1.383***	.015	1.384***	.015
Mixed citizenship	0.716***	.020	0.718***	.020	0.714***	.020	0.713***	.020
35–64	0.105***	.105	0.059***	.	0.051***	.006	-0.089***	.017
65+	0.030***	.030	-0.040***		-0.044***	.008	-0.120***	.017
Secondary	0.789***	.006	0.729***		0.736***	.006	0.735***	.006
Tertiary	1.363***	.007	1.341***		1.355***	.007	1.351***	.007
Centre			-0.262***					
Mezzogiorno			-0.995***					
Regional Fixed Effects								
Trentino Alto Adige	-		-		0.024	.018	0.022	.018
Lombardia	-		-		0.281***	.010	0.281***	.010
Piemonte	-		-		0.062***	.012	0.061***	.012
Friuli Venezia Giulia	-		-		0.038*	.015	0.037*	.015
Veneto	-		-		0.116***	.012	0.115***	.012
Liguria	-		-		-0.014	.015	-0.014	.015
Valle d'Aosta	-		-		-0.167***	.031	-0.168***	.031
Toscana	-		-		-0.060***	.011	-0.060***	.011
Marche	-		-		-0.224***	.014	-0.224***	.014

Table 7. Continued

Variable	M0		M1		M2		M3	
	Odds ratios	SE	Odds ratios	SE	Odds ratios	SE	Odds ratios	SE
Lazio	-		-		-0.221***	.013	-0.221***	.013
Umbria	-		-		-0.209***	.019	-0.209***	.019
Abruzzo	-		-		-0.527***	.015	-0.527***	.015
Campania	-		-		-1.049***	.012	-1.049***	.012
Sardinia	-		-		-0.482***	.014	-0.482***	.014
Molise	-		-		-0.924***	.024	-0.924***	.0235
Puglia	-		-		-0.660***	.012	-0.660***	.0122
Basilicata	-		-		-0.951***	.020	-0.951***	.0205
Sicilia	-		-		-1.046***	.012	-1.046***	.0115
Calabria	-		-		-1.337***	.015	-1.337***	.0151
2#35–64							0.089***	.021
2#65+							-0.005	.021
3#35–64							0.080***	.020
3#65+							-0.123*	.058
4#35–64							0.300***	.021
4#65+							-0.662**	.238
5 or more#35–64							0.294***	.028
5 or more#65+							-0.407	1.162
_cons	-1.785***	.016	-1.556***	.016	-1.667***	.018	-1.569***	.022
Pseudo R^2	0.054		0.086		0.091		0.091	
Chi2	61841		97970		1.00E + 05		1.00E + 05	
AIC	1.10E + 06		1.00E + 06		1.00E + 06		1.00E + 06	

Legend: * $p < .05$; ** $p < .01$; *** $p < .001$
Source: 2019 PPHC-Istat.

Table 8. Estimate of parameters of logit model (odds ratios and standard errors).

	M4		M5		M6	
Variable	Odds ratios	Standard Errors	Odds ratios	SE	Odds ratios	SE
2	0.031***	.006	0.032***	.006	-0.024	.020
3	0.014	.008	0.014	.008	-0.046	.018
4	0.097***	.008	0.096***	.008	-0.056	.018
5 or more	-0.240***	.012	-0.241***	.012	-0.378	.020
All Italians	1.361***	.015	0.935***	.027	0.940	.027
Mixed citizenship	0.695***	.020	0.438***	.042	0.439	.042
35–64	0.105***	.015	0.137***	.015	0.022	.022
65+	0.266***	.015	0.315***	.015	0.237	.021
Secondary	0.945***	.014	0.400***	.034	0.410	.034
Tertiary	1.610***	.015	1.138***	.041	1.140	.041
Regional Fixed Effects						
Trentino Alto Adige	0.022	.018	0.022	.018	0.020	.0181917
Lombardia	0.283***	.010	0.283***	.010	0.282	.0103586
Piemonte	0.065***	.012	0.065***	.012	0.064	.0118956
Friuli Venezia Giulia	0.039*	.015	0.039*	.015	0.038	.0151733
Veneto	0.113***	.012	0.113***	.0117817	0.112	.0117847
Liguria	-0.005	.0152751	-0.005	.0152751	-0.005	.0152787
Valle d'Aosta	-0.158***	.0316612	-0.158***	.0316612	-0.158	.0316687
Toscana	-0.057***	.011311	-0.057***	.011311	-0.050	.0113138
Marche	-0.229***	.0140247	-0.229***	.0140247	-0.220	.0140282
Lazio	-0.219***	.012665	-0.219***	.012665	-0.219	.0126678
Umbria	-0.211***	.0190954	-0.211***	.0190954	-0.211	.0191004
Abruzzo	-0.530***	.0154521	-0.532***	.0154521	-0.531	.0154557
Campania	-1.047***	.0124042	-1.048***	.0124042	-1.040	.012407
Sardinia	-0.470***	.0143121	-0.468***	.0143121	-0.467	.014

Table 8. Continued

	M4		M5		M6	
Variable	Odds ratios	Standard Errors	Odds ratios	SE	Odds ratios	SE
Molise	-0.929***	.0235753	-0.931***	.0235753	-0.931	.023
Puglia	-0.656***	.0122558	-0.656***	.0122558	-0,655	.012
Basilicata	-0.957***	.0205521	-0.959***	.0205521	-0.959	.020
Sicilia	-1.040***	.0115886	-1.040***	.0115886	-1.040	.011
Calabria	-1.336***	.0151669	-1.338***	.0151669	-1.330	.015
2#35−64					0.089	.0217572
2#65+					0.030	.0216346
3#35−64					0.060	.0203841
3#65+					-0.046	.0571932
4#35−64					0.274	.0212905
4#65+					-0.612	.2352247
5 or more#35−64					0.260	.0279282
5 or more#65+					-0.317	1.167.796
Secondary#35−64	-0.002	.017	-0.036*	.0168228	-0.050	.0168708
Secondary #65+	-0.584***	.018	-0.643***	.0183936	-0.644	.0184273
Tertiary#35−64	-0.084***	.018	-0.116***	.0184085	-0.130	.0184731
Tertiary #65+	-.72548***	.022	-.77502***	.0225984	-0,776	.0226243
Secondary#Mixed citizenship			0.607***	.0340457	0.600	.0340634
Secondary#All Italians			0.356***	.05087	0.354	.0508955
Tertiary#Mixed citizenship			0.523***	.0401673	0.519	.0401839
Tertiary#All Italians			0.400***	.0573441	0.399	.0573705
_cons	-1.845***	.021	-1.469***	.029	-1.380	.031
Pseudo R^2	0.093		0.094		0.093	
Chi2	1.10E+05		1.10E+05		1.00E+05	
AIC	1.00E+06		1.00E+06		1.00E+06	

Legend: * $p < .05$; ** $p < .01$; *** $p < .001$
Source: 2019 PPHC-Istat.

8. References

Antoun, C., M.P. Couper, and F G. Conrad. 2017. "Effects of Mobile versus PC Web on Survey Response Quality: A Crossover Experiment in a Probability Web Panel." *Public Opinion Quarterly* 81: 280–306. DOI: https://doi.org/10.1093/poq/nfw088.

Bandilla, W., M.P. Couper, and L. Kaczmirek. 2014. "The Effectiveness of Mailed Invitations for Web Surveys and the Representativeness of Mixed-Mode versus Internet-only." *Survey Practice* 7(4). DOI: https://doi.org/10.29115/SP-2014-0020.

Benassi, F., and A. Naccarato. 2017. "Households in Potential Economic Distress. A Geographically Weighted Regression Model for Italy, 2001–2011." *Spatial Statistics* 21: 362–376. DOI: https://doi.org/10.1016/j.spasta.2017.03.002.

Benzeval, M., J. Burton, T.F. Crossley, P. Fisher, C. Gardiner, A. Jäckle, and J. Moore. 2021. "High Frequency Online Data Collection in an Annual Household Panel Study: Some Evidence on Bias Prevention and Bias Adjustment." *Understanding Society Working Paper Series* 03: 1–26. Available at: https://www.understandingsociety.ac.uk/sites/default/files/downloads/working-papers/2021-03.pdf (accessed September 2021).

Bethlehem, I., F. Cobben, and B. Schouten. 2011. *Handbook of Nonresponse in Household Surveys*. New York: Wiley.

Bianchi, A., S. Biffignandi, and P. Lynn. 2017. "Web-face-to-face Mixed-mode Design in a Longitudinal Survey: Effects on Participation Rate, Sample Composition, and Costs." *Journal of Official Statistics* 33(2): 385–408. DOI: http://dx.doi.org/10.1515/JOS-2017-0019.

Biemer P.P. 2010. "Total Survey Error: Design, Implementation, and Evaluation." *Public Opinion Quarterly* 74: 817–848. DOI: https://doi.org/10.1093/poq/nfq058.

Biffignandi, S. and J.G. Bethlehem. 2012. "Web Surveys: Methodological Problems and Research Perspectives." In *Advanced Statistical Methods for the Analysis of Large Data sets,* edited by A. Di Ciaccio, M. Coli, and J.M.A. Ibanez: 363–374. Berlin: Springer.

Biffignandi S., and J. Bethlehem. 2021 *Handbook on Web Surveys.* Second edition. New York: Wiley.

Biffignandi, S. and M. Pratesi. 2002. "Modelling the Respondents' Profile in a Web Survey on Firms in Italy." In *Developments in Social Science Methodology,* edited by A. Ferligoj and A. Mrvar: 171–185. Ljubljana: FDV.

Brick, J.M., and R. Tourangeau. 2017. "Responsive Survey Designs for Reducing Nonresponse Bias." *Journal of Official Statistics* 33(3): 735–752. DOI: https://doi.org/10.1515/jos-2017-0034.

Cellini R., and G. Torrisi. 2014. "Regional Resilience in Italy: A Very Long-run Analysis." *Regional Studies* 48(11): 1779–1796. DOI: https://doi.org/10.1080/00343404.2013.861058.

Chieppa, A., G. Gallo, V. Tomeo, F. Borrelli, and S. di Domenico. 2018. "Knowledge Discovery for Inferring the Usually Resident Population from Administrative Registers." *Mathematical Population Studies* 26(2): 96–102. DOI: https://doi.org/10.1080/08898480.2017.1418114.

Ciapanna, E. and G. Roma. 2020. *Connected Italy*. Rome: Bank of Italy. (Questioni di Economia e Finanza. Occasional Papers No. 573). Available at: https://www.bancaditalia.it/pubblicazioni/qef/2020-0573/QEF_573_20.pdf?language_id = 1 (accessed September 2021).

Citro, C.F. 2014. "From Multiple Modes for Surveys to Multiple Data Sources for Estimates". *Survey Methodology* 40(2): 137–161. Available at: https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2014002/article/14128-eng.pdf?st = x9gGl05C (accessed September 2021).

Cobben, F., and J.G. Bethlehem. 2013. *Web Panels for Official Statistics.* The Hague: Statistics Netherland. (Discussion Paper 201307). Available at: https://www.cbs.nl › 2013 › 2013-07-x10-pub (accessed September 2021).

Couper M.P. 2017. "New Developments in Survey Data Collection." *Annual Review of Sociology* 43: 121–145. DOI: https://doi.org/10.1146/annurev-soc-060116-053613.

Cracolici M.F., M. Cuffaro and P. Nijkamp. 2007. "Geographical Distribution of Unemployment: An Analysis of Provincial Differences in Italy." *Growth and Change* 38(4): 649–670. DOI: https://doi.org/10.1111/j.1468-2257.2007.00391.x.

Crescenzi, F. 2015. "Changing census by combining administrative sources and sample surveys". In *Proceedings of the 60th ISI World Statistics Congress*, 26–31 July 2015, Rio de Janeiro, Brazil. Available at: https://2015.isiproceedings.org/Files/IPS022-P4-S.pdf (accessed September 2021).

D'Ambrosio C. 2003. "Household Characteristics and the Distribution of Income in Italy: An Application of Social Distance Measures." *The Review of Income and Wealth* 47(1): 43–64. DOI: https://doi.org/10.1111/1475-4991.00003.

De Clercq M., J. Buysse, and M. D'Haese. 2020. The state of Rural Digitalisation in Europe. Desira. Available at: https://desira2020.eu/wp-content/uploads/2020/11/Briefing_Digitalisation-in-Europe.pdf (accessed September 2021).

De Leeuw, E.D. 2005. "To Mix or Not to Mix Data Collection Modes in Surveys". *Journal of Official Statistics* 21(2): 233–255.

De Leeuw E.D. 2018. "Mixed Mode: Past, Present, and Future." *Survey Research Methods* 12(2): 75–89. DOI: doi:10.18148/srm/2018.v12i2.7402.

De Leeuw, E.D., and N. Berzelak. 2016. "Survey Mode or Survey Modes?" In *The Sage Handbook of Survey Methodology*, edited by C. Wolf, D. Joye, T.W. Smith, and Y. Fu, 142–156. London: Sage Publications.

De Leeuw, E.D., J. Hox, and D. Dillman. 2008. *International Handbook of Survey Methodology.* New York: Routledge.

De Leeuw E.D., Z.T. Suzer-Gurtekin, and J.J. Hox. 2019. "The Design and Implementation of Mixed-mode Surveys." In *Advances in Comparative Survey Methods: Multinational, Multiregional, and Multicultural contexts (3MC)*, edited by P. Johnson, B.E. Pennell, I.A.L. Stoop, and B. Dorer: 387–408. New York: Wiley.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method.* Fourth edition. New York: Wiley.

Durrant, G.B., and F. Steele. 2009. "Multilevel Modelling of Refusal and Non-contact in Household Survey Evidence from Six UK Government Surveys." *Journal of the Royal Statistical Society: Series A* 172(2): 361–381. DOI: https://doi.org/10.1111/j.1467-985X.2008.00565.x.

European Union 2017. *Regulation of the European Parliament as Regards the Territorial Typologies 2017/2391.* Official Journal of the European Union. Available at: https://eurlex.europa.eu/legal-content/EN/TXT/?uri = celex%3A32017R2391 (accessed September 2021).

Eurostat 2020. Digital Economy and Society Statistics – Households and Individuals. Luxemburg. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/-Digital_economy_and_society_statistics_-_households_and_individuals (accessed September 2021).

Eurostat 2021. *Applying the Degree of Urbanisation. A Methodological Manual to Define Cities, Towns and Rural Areas for International Comparisons*. 2021 edition. Eurostat: Luxemburg.

Goldstein, H. 2010. *Multilevel Statistical Models*. Fourth Edition. New York: Wiley.

Hargittai, E. 2002."Second-Level Digital Divide: Differences in People's online skills." *First Monday* 7(4). DOI: https://doi.org/10.5210/fm.v7i4.942.

Hox J.J., E.D. de Leeuw, and E.A.O. Zijlmans. 2015. Measurement Equivalence in Mixed Mode Surveys. *Frontiers in Psychology* 6: 87. DOI: https://doi.org/10.3389/fpsyg.2015.00087.

Istat 2018. "Preliminary Experimental Results on the Italian Population and Housing Census Estimation Methods." In *Proceedings of the Twentieth Conference of European Statisticians Group of Experts on Population and Housing Censuses*, 26–28 September, 2018. Geneva: United Nations Economic Commission for Europe. Available at: https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2018/Meeting-Geneva-Sept/ECE_CES_GE.41_2018_20-1811953E.pdf (accessed September 2021).

Istat 2019. *Cittadini e ICT*. Statistiche report. Rome: Istat. Available at: https://www.istat.it/it/archivio/236920 (accessed September 2021).

Jaccard, p. 1908. "Nouvelles recherches sur la distribution florale." *Bulletin de la Société Vaudoise des Sciences Naturelles* 44(163): 223–270.

Kappelhof, J.W.S. 2015. "Face-to-Face or Sequential Mixed-Mode Survey Among Non-Western Minorities in the Netherlands: The effect of Differential Survey Designs on the Possibility of Nonresponse Bias." *Journal of Official Statistics*, 31(1): 1–3. DOI: https://doi.org/10.1515/jos-2015-0001.

Keusch F., B. Struminskaya, C. Antoun, M.P. Couper, F. Kreuter. 2019. "Willingness to Participate in Passive Mobile Data Collection." *Public Opinion Quarterly* 83: 210–235. DOI: https://doi.org/10.1093/poq/nfz007.

Luiten A., J. Hox and E.D. de Leeuw. 2020. Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study across Countries and Surveys. *Journal of Official Statistics* 36(3): 469–487. DOI: https://doi.org/10.2478/-jos-2020-0025.

Maddala, G.S. and K. Lahiri. 2010. Introduction to Econometrics. Fourth Edition. New York: Wiley.

Maslovskaya, O., G.B. Durrant, P.W.F. Smith and A. Villar. 2019. "What are the Characteristics of Respondents using Different Devices in Mixed-device Online Surveys? Evidence from Six UK Surveys." *International Statistical Review* 87(2): 326–346. DOI: https://doi.org/10.1111/insr.12311.

McCormick, W.P., N.I. Lyons, and K. Hutcheson. 1992. "Distributional Properties of Jaccard's Index of Similarity." *Communications in Statistics – Theory and Methods* 21(1): 51–68. DOI: https://doi.org/10.1080/03610929208830764.

Messer, B.L., and D.A. Dillman. 2011. "Surveying the general public over the internet using address-based sampling and mail contact procedures." *Public Opinion Quarterly*, 75(3), 429–457. DOI: https://doi.org/10.1093/poq/nfr021.

Mohorko A., E. de Leeuw, and J. Hox. 2013. "Internet Coverage and Coverage Bias in Europe: Developments Across Countries and Over Time." *Journal of Official Statistics* 29(4): 609–622. DOI: https://doi.org/10.2478/jos-2013-0042.

Pratesi, M., K. Lozar Manfreda, S. Biffignandi, and V. Vehovar. 2004. "List-based Web Surveys: Quality, Timeliness and Nonresponse in the Steps of the Participation Flow." *Journal of Official Statistics* 20(3): 451–465. Available at: https://www.scb.se/contentassets/ca21efb41fee47d293bbee5bf7be7fb3/list-based-web-surveys-quality-timeliness-and-nonresponse-in-the-steps-of-the-participation-flow.pdf.

Rados, J. 2021. "Broadband Access in Rural and Mountainous Areas." The European Parliament Intergroup on Rural, Mountainous and Remote Areas and Smart Villages. Available at: https://www.smart-rural-intergroup.eu/broadband-access-in-rural-and-mountainous-areas-in-the-eu/ (accessed September 2021).

Reynaud C., S. Miccoli, F. Benassi, A. Naccarato, and L. Salvati. 2020. "Unravelling a Demographic 'Mosaic': Spatial Patterns and Contextual Factors of Depopulation in Italian Municipalities, 1981–2011." *Ecological Indicators* 115: 106356. DOI: https://doi.org/10.1016/j.ecolind.2020.106356.

Righi P., P.D. Falorsi, S. Daddi, E. Fiorello, P. Massoli, and M.D. Terribili. 2021. "Optimal Sampling for the Population Coverage Survey of the New Italian Register Based Census." *Journal of Official Statistics* 37(3): 655–671. DOI: https://doi.org/10.2478/jos-2021-0029.

Rivero, M.S., M.C.R. Rangel, and J.M.S. Martín. 2019. "Geotourist Profile Identification using Binary Logit Modeling: Application to the Villuercas-Ibores-Jara Geopark (Spain)." *Geoheritage* 11: 1399–1412. DOI: https://doi.org/10.1007/s12371-019-00384-z.

Rizzato, F. 2020. "Broadband Snapshots: Italy". Opensignal. Available at https://www.opensignal.com/2020/09/04/broadband-snapshots-italy (accessed September 2021).

Santarelli, E., and F. Cottone. 2009. "Leaving Home, Family Support and Intergenerational Ties in Italy: Some Regional Differences." *Demographic Research* 21(1): 1–22. DOI: https://doi.org/10.4054/DemRes.2009.21.1.

Scherpenzeel, A.C. 2011. "Data Collection in a Probability-based Internet Panel: How the LISS Panel Was Built and How It Can Be Used." *Bulletin of Sociological Methodology* 109(1): 56–61. DOI: https://doi.org/10.1177/0759106310387713.

Scherpenzeel, A.C. and J.G. Bethlehem. 2011. "How Representative are Online Panels? Problems of Coverage and Selection and Possible Solutions." In *Social and Behavioural Research and the Internet: Advances in Applied Methods and Research Strategies*, edited by J.W.M. Das, P. Ester, and L. Kaczmirek: 105-132. New York: Routledge.

Tourangeau R. 2017. Presidential Address: Paradoxes of Nonresponse. *Public Opinion Quarterly* 81: 803–817. DOI: https://doi.org/10.1093/poq/nfx031.

UNECE (United Nations Economic Commission for Europe). 2018. *Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses.* ECE/CES/STAT/2018/4. New York: United Nations. Available at: https://unece.org/guidelines-use-registers-and-administrative-data-population-and-housing-censuses-0 (accessed September 2021).

Wooldridge J.M. 2012. *Introductory Econometrics: A Modern Approach*. Fifth Edition. Boston: Cengage Learning.

Received March 2021
Revised October 2021
Accepted July 2022