SORT 42 (2) July-December 2018, 183-206 DOI: 10.2436/20.8080.02.74

Field rules and bias in random surveys with quota samples. An assessment of CIS surveys

Jose M. Pavía[1] and Cristina Aybar[2]

Abstract

Surveys applying quota sampling in their final step are widely used in opinion and market research all over the world. This is also the case in Spain, where the surveys carried out by CIS (a public institution for sociological research supported by the government) have become a point of reference. The rules used by CIS to select individuals within quotas, however, could be improved as they lead to biases in age distributions. Analysing more than 545,000 responses collected in the 220 monthly barometers conducted between 1997 and 2016 by CIS, we compare the empirical distributions of the barometers with the expected distributions from the sample design and/or target populations. Among other results, we find, as a consequence of the rules used, significant overrepresentations in the observed proportions of respondents with ages equal to the minimum and maximum of each quota (age and gender group). Furthermore, in line with previous literature, we also note a significant overrepresentation of ages ending in zero. After offering simple solutions to avoid all these biases, we discuss some of their consequences for modelling and inference and about limitations and potentialities of CIS data.

MSC: 62D05

Keywords: Centre for Sociological Research, quota sampling, fieldwork rules, age and gender groups, inter-quota distributions, intra-quota distributions

1. Introduction

The revolution in information and communication technologies is transforming society and changing the world of business. Collecting, transmitting and storing vast amounts of data (both structured and unstructured) is more viable than ever and has made it easier to find out the individual needs and wishes of certain sectors of the population. As instruments for studying the opinions and attitudes of society as a whole, however, they appear less effective (e.g., Burnap et al., 2016; Jungherr et al., 2017; Kalampokis et al., 2017). Challenges, such as the polarization of opinions expressed on the internet, the

[1] UMICCS, GIPEyOP, Av. Tarongers s/n, Dpt. d'Economia Aplicada, Facultat d'Economia, Universitat de València, 46022 València (Spain).

[2] GIPEyOP, Av. Tarongers s/n, Dpt. d'Economia Aplicada, Facultat d'Economia, Universitat de València, 46022 València (Spain).

Received: May 2018
Accepted: November 2018

division of sociopolitical communities, the difficulties that still arise in the automatic processing of natural language and the biases of connectivity and internet use (e.g., Del Vicario et al., 2017; Ebrahimi, Yazdavar and Sheth, 2017; Mellon and Prosser, 2017) together with the enormous limitations of representativeness still apparent in Big Data (Meng, 2016), point towards a future in which random surveys will continue to be a necessary and fundamental tool for governmental and business decision-making.

Conducting an opinion poll using sampling methods is a very complex process with many interconnected issues, which must respond to well-established principles and methodologies (Groves et al., 2009). The design and planning of a probabilistic survey involves bringing together many facets to create a single tool. Specifying objectives, delineating the target population, designing the questionnaire, choosing the sample, specifying the fieldwork rules and offering guidelines on how the interviewers should act are just some of the tasks that the survey designer must ponder before starting any research, trying to anticipate biases and any problems that may arise (Cea D'Ancona, 2004). The biases from which a survey may suffer are not insignificant, could have multiple origins and can even vary depending on the context (Pavía, Badal, and García-Cárceles, 2016), so they should be avoided wherever possible.

In Spain, surveys carried out by the Centre for Sociological Research (CIS, from its acronym in Spanish: Centro de Investigaciones Sociológicas) and, in particular, its monthly barometers, are a benchmark in the sector of opinion and market studies, thanks to, among other things, the professionalization of its network of interviewers, the size of its samples and the spatial distribution procedures it implements (Pavía and García-Cárceles, 2012). However, there is still room to improve CIS surveys; without cost. In this paper, we focus on analysing the impact of the rules that CIS uses for choosing respondents within the home. In particular, we study the effects that fieldwork rules R1 and R2 have in terms of age distribution within each quota (which we will call intra-quota distributions): (R1) "when in a dwelling there is more than one person who meets the conditions demanded by the quota, the youngest one will be interviewed", and (R2) "if it is impossible to obtain a certain age quota, it can be replaced by one of the age adjacent quotas" (Díaz de Rada, 2005, 2014).

Although the controversy following the 1948 US Presidential Election definitely stated the superiority of probability sampling over (old versions of) quota sampling (Mosteller et al., 1949), market researchers and political pollsters all over the world have continued using quota sampling alongside probability-based methods due to its cost-effective relationship (Vehovar, 1999; Vavreck and Rivers, 2008). Indeed, despite its detractors (e.g., Smith, 1983; Sudman, 1976; Marsh and Scarbrough, 1990), quota sampling is currently the dominant method in online studies through the use of panels (Kennedy et al., 2016) and is gaining even more popularity in traditional telephone and face-to-face surveys due to the growing increase of nonresponse rates and the extra costs they entail (Yang and Banamah, 2014). Certainly, although criticism from the statistical academic community against any practice of quota sampling can be sometimes fierce, the reality is that complex, modern versions of quota sampling, like the one used by

CIS, are closer to probability sampling than to (traditional) quota sampling. The quota sampling design operated by CIS in door to door surveys, often referred to as quota-random sampling or quasi-random sampling (Trochim and Donnelly, 2006), is clearly more refined than the standard quota sampling employed in market research and has been positively used by the academia (Stephenson, 1978).[1]

As discussed in the second section, CIS uses quotas for age and gender to choose interviewees within the household. The number of people to be interviewed from each age and gender group (which we will call inter-quota distributions) is determined exogenously, by design, looking to adjust to a reference (or design) population. The number of people who are interviewed for each age within each quota (intra-quota distributions) is, however, endogenous, depends on chance although influenced by R1 (and R2). Our hypothesis is that within each quota the distributions by age will not be adjusted to the distributions of the target population [2] (from which the respondents are chosen) and that there will be an overrepresentation of the younger subjects as a consequence of R1. Regarding the effect of R2 in the intra-quota distributions, we do not have any hypothesis a priori although, as we will see, its effects are visible. In light of the results, we venture a hypothesis (not evaluable with our data) on how the interviewers interpret R2. Likewise, although our research focuses on the study of intra-quota distributions, we also analyse for completeness whether the empirical inter-quota distributions conform to those designed, reflecting on the possible effect on them of R2.

The analysis was carried out based on the study of the variables of sex, age and province collected from the more than 545,000 interviews completed in the 220 barometers conducted by CIS between January 1997 and December 2016. All the microdata used in this research have been obtained from the CIS Data Bank (`www.cis.es`). The supplementary material lists the numbers of the surveys analysed. The data for the comparisons have been obtained from the Spanish National Institute of Statistics, INE (`www.ine.es`). In the case of intra-quota analysis (the main objective of this research), the comparisons are made on the target populations associated with the collection dates of each barometer and in the case of inter-quota analyses on the reference populations used in each barometer to define quotas. Regarding the latter issue, it should be noted that although the CIS now uses the figures of Spanish residents to define the quotas, until 2015 it used the figures of total residents, with a non-regular update schedule (see Table 1).

The rest of the article is structured as follows. The second section briefly describes the sample design used by CIS in its monthly barometers, paying special attention to the selection of individuals within the household. The third section is devoted to delimiting comparison populations. It specifies the reference and target populations of

1. For example, a version of random-quota sampling, similar to current CIS designs, was used during the seventies (due to budgets constraints) in some waves of the prestigious General Social Survey conducted by NORC.

2. In each barometer, the target population is made up of residents over 18 years of age with Spanish nationality at the time of the survey. In the CIS barometers, the reference and target populations have not historically coincided.

each barometer, details how these have been calculated from the official statistics and outlines our reflections on the consequences of applying R1 and R2. The fourth section introduces methodological issues. In the fifth section, the main results obtained are presented. These are extended and complemented substantially in the supplementary material that accompanies this article. Finally, the sixth section summarises the results, proposes solutions and discusses limitations and strengths of CIS data.

2. The barometers of the CIS. Selection of final units

A CIS barometer consists of a personal survey, conducted monthly in homes (except in August), using a standardized questionnaire whose universe (target population) is the population of 18 years or older resident in Spain and with Spanish nationality. In general, barometers had a designed size of 2,500 individuals[3] with, *theoretically*, all units having the same probability of being selected.

The selection procedure for interviewees is carried out within households, after choosing them through a complex sampling procedure of several stages in which stratified sampling, cluster sampling, random routes and age and gender quotas are applied. First, strata are formed crossing the 17 administrative regions of Spain (CC.AA) with seven categories of habitat size (less than or equal to 2,000 inhabitants, from 2,001 to 10,000, from 10,001 to 50,000, from 50,001 to 100,000, from 100,001 to 400,000, from 400,001 to 1,000,000, and more than 1,000,000 inhabitants)[4]. Once the sample size is spread among the strata (usually by proportional allocation), the next step is to determine how many and which census sections to visit in each stratum. For this, municipalities are chosen at random and, within the municipalities, census sections. Municipalities and sections are selected with probabilities proportional to their different sizes. More details on this procedure can be found in Díaz de Rada (2005, 2008, 2014), Rodríguez Osuna (1991, 2005) and Pavía and García-Cárceles (2012).

Once the interviewers are in place, in the selected census sections, the last stage begins: that of choosing respondents. The interviewees are chosen directly by the interviewers, in the field, through a combined application of random routes and quotas for age and gender. Given the impossibility of having a detailed list of the residents in each household, the interviewers choose interviewees using a series of rules, set in advance, that aim to favour randomness and representativeness. First, households are selected and, then, subjects within the household.

For the selection of households, the interviewer receives a list with the streets (and door numbers) that make up the census section and information about the starting point

3. Since September 2018 there are 3000 individuals.

4. Among other deviations from this norm, it should be noted the positive inclusion of Ceuta and Melilla in the barometers in July 2013 or that, in the definition of the strata, there are certain peculiarities related to metropolitan and insular areas.

of her route. From here, the interviewers visit homes following very detailed criteria (which have suffered some variations over time) related to the routes they must follow, the number of dwellings to visit at each door number, the distances between dwellings and door numbers when an interview takes place, the criteria of election and substitution, etc. (Díaz de Rada, 2005, 2008, 2015). When a dwelling is selected (and its members are willing to collaborate), the last stage begins: that of choosing the person to interview. Only one person is interviewed per household.

For the selection of people within a dwelling the most commonly used methods are either chance (e.g., Kish, 1995; Lind, Link and Oldendick, 2000; O'Rourke and Blair, 1983) or quotas. In the barometers, CIS uses the quota method. Before beginning the fieldwork, the interviewers receive instructions on the number of people of each age and gender (quota) that they must interview in each census section and they only select homes with people who meet the characteristics set in the quotas, until they have completed the survey. This approach means that, as field work progresses, more and more households are discarded as a result of not pursuing quotas that have already been filled (Díaz de Rada, 2014). The CIS considers 12 possible groups crossing gender (Female, Male) with age, classified into 6 categories: 18-24; 25-34; 35-44; 45-54; 55-64; and 65 or more years.

When faced with only one person who meets the quota in the home selected, this person is interviewed. The difficulty comes when there is more than one suitable person representing the quotas still to be filled. In this case, the R1 fieldwork rule of CIS recommends interviewing the youngest person. Given that on average younger people spend more time outside the home, this rule tries to facilitate the work of the interviewers, allowing them to choose the people that are more difficult to encounter in the home first. This rule makes sense when people belong to different quotas but it does not favour a representative selection within each age bracket; which, in line with our hypothesis, we will see causes the appearance of imbalances in the distributions by age of the samples. Together with the previous rule, in case of impossibility or difficulty in encountering a certain quota, CIS allows the quota to be met by replacing it with one of the adjacent quotas.[5]

3. Comparison populations

Through the sample design, CIS aims to replicate the structure of age and gender of its target population, which is made up of the Spanish population aged 18 years or over and resident in Spain. Unfortunately, the target and design populations have not always coincided. Until 2015, CIS exclusively used the statistics of total residents to select municipalities and sections and to determine the size of the quotas. This (historical)

5. The fieldwork rules of CIS only permit alterations in the age quota, never in the gender, and only allow one substitution in each route (Díaz de Rada, 2015).

divergence between the target and the reference (or design) populations has introduced certain deviations in the variables being considered that are worth quantifying[6]. In our research, we focus on another issue: studying the impact of R1 and R2 on the distributions by age collected. Thus, the demographic structures derived from the barometers (empirical inter-quota and intra-quota distributions) are compared with the corresponding population structures associated with each barometer (theoretical distributions of design and target populations). The theoretical distributions are obtained from the official statistics of the Spanish National Institute of Statistics (INE).

On the one hand, for the inter-quota analyses, the population figures used by CIS are employed to determine the size of the quotas in the design of each barometer (reference populations). On the other hand, for the intra-quota comparisons, the population figures of resident Spaniards corresponding to the month of completion of each barometer (target populations) are considered, conditional on the sizes of the design quotas in each survey.

3.1. Inter-quota distributions[7]

Historically, in the 20 years of barometers analysed, we can identify three stages with respect to the demographic structures used for the design of the surveys. The first stage covers the period from 1997 to 2005 (except September 2005) and encompasses 98 barometers, in which CIS used the resident population and the 1995 Municipal Register for deciding the distribution of the sample size, the selection of units (census sections) and the determination of the quotas. In this first stage, the size assigned to each cross of age and gender (quota) in each autonomous community was obtained by multiplying the sample size that corresponded to the autonomous community by its corresponding population structure.

The second stage covered the period from January 2006 to June 2015 (including September 2005). In this period, which included 106 barometers, the census is updated annually (with a schedule not completely regular) taking as a reference the resident population of the latest Municipal Register available from INE: 10 different Registers were used (see Table 1). In this stage, the quota size that corresponds to each autonomous community is determined as in the previous stage, although an adjustment is introduced regarding its distribution among census sections. Previously, they were distributed using a random procedure with restrictions. In this period, they are distributed taking into account the structure by age and gender of the census sections to be visited such that when a particular section has more residents of a certain quota, more people in that census section of that quota will be assigned to be interviewed. This new

6. In Pavía and García-Cárceles (2012) a first approximation can be found in the case of intention to vote in pre-election survey.

7. We appreciate the invaluable help of Valentín C. Martínez (technical advisor of the CIS research department) in answering our questions for the preparation of this section. Any inaccuracy that exists is the sole responsibility of the authors.

Table 1: Municipal Registers (and reference variables) used in the analysed barometers.

Barometers (dates)	Register	Variable	Number of barometers
1997.01-2005.12 (except 2005.09)	1995	Total residents	98
2006.01-2006.09 (including 2005.09)	2004	Total residents	9
2006.10-2007.09	2005	Total residents	11
2007.10-2008.10	2006	Total residents	12
2008.11-2009.10	2007	Total residents	11
2008.11-2009.10	2007	Total residents	11
2009.11-2010.09	2008	Total residents	10
2010.10-2011.12	2009	Total residents	14
2012.01-2012.10	2010	Total residents	9
2012.11-2013.06	2011	Total residents	8
2013.07-2014.03	2012	Total residents	8
2014.04-2015.06	2013	Total residents	14
2015.07-2016.05	2014	Spanish residents	10
2016.06-2016.12	2015	Spanish residents	6

Source: Compiled by the authors from personal communication with V. Martínez.

strategy seeks to reduce field work time and to compensate for the increasing rates of non-response (Pavía and Larraz, 2012; Díaz de Rada, 2014).

The last stage began in July 2015 and continues today (up to the time of writing this paper). In this stage, new improvements are introduced regarding the construction of the strata and relative to the reference populations used for the design of the survey. The Register is updated annually (see Table 1), but now CIS uses the total number of residents as reference population to define the strata but the total Spanish population to distribute the sample among the strata. Likewise, the Spanish resident population aged 18 or over is used for the allocation of the quota sizes, not taking into account the stratification for its distribution but only the sections that are going to be visited. That is, quotas are now only representative at the national level.

Although the main objective of this paper is to evaluate the bias introduced by the rules R1 and R2 in the intra-quota distributions (since the adjustment to the inter-quotas distributions should be fulfilled, at least approximately, by design), for the sake of completeness, we have also made comparisons (between theoretical and empirical distributions) for inter-quota distributions in the first part of the results section 5.1. Inter-quota distributions are, in turn, necessary to make intra-quota comparisons when analysing combined age groups (see Figure 3).

Before we can use the theoretical inter-quota distributions employed in the design of each barometer, however, we have to address two questions. On the one hand, the problem posed by the unavailability of the data corresponding to the 1995 Municipal Register, used in the barometers from 1997 to 2005. On the other hand, the possible impact that the so-called rounding effect[8] may have on inter-quota distributions, which

8. The rounding effect emerges when one tries to replicate a population structure of several million people through a sample of, at most, a handful of thousands of people, and manifests itself in the differences that exist between the percentage distributions of the design and sample populations.

may cause significant differences between the inter-quota distributions of the reference populations and the inter-quota distributions actually used.

Both problems are addressed in the supplementary material. The first of the questions is solved by making a comparison between the official statistics of the Municipal Registers and Population Now-Cast estimates of INE, from which we obtain an estimate of residents corresponding to the 1995 Municipal Registers[9]. Regarding the second question, the effect of rounding, we find that, although this is significant at CC.AA level (which would affect the autonomous region inter-quota distributions in the barometers prior to July 2015), it does not appear at the aggregate level, nationally[10]. Given that our comparisons between theoretical inter-quota distributions and empirical quotas are limited to the national ambit, we infer from the analysis that we can use as comparative inter-quota distributions the demographic structures that were used to define the quotas in each group of barometers, which can be calculated directly from the official statistics of INE[11].

3.2. Intra-quota distributions

The inter-quota distributions are determined a priori, by design, with the objective of replicating in the sample the same demographic structure (by groups) that exists in the reference population, although, as we have already seen, the reference and target populations have historically shown certain divergences. In the case of intra-quota distributions, however, there is a correspondence with the target population[12]. The interviewers are instructed to select Spanish residents (in households) aged 18 years or older.

With the limits imposed by the inter-quota distribution of the sample (number of people of each group of age/gender that should be interviewed), the interviewers select individuals within each quota among the resident Spanish population. Therefore, without the effect of R1 and R2, the expected intra-quota structures of the surveyed population should mimic those of the target population; a result that will not be expected when the whole sample is considered together. For the whole sample, the expected structure of the surveyed population will not exactly match that of the resident Spanish population (unless the target and reference populations are equal), since each quota (age/gender group) can have a different relative weight in the design (reference) and target populations.

---

9. See the section "Estimation of the distribution by age and gender (inter-quota) of the 1995 Register" in the supplementary material.

10. The details that support these conclusions are offered in the section "Analysis of the impact of the rounding effect on inter-quota distributions" in the supplementary material.

11. This result is interesting because it makes the analyses that we have carried out more transparent and easily replicable.

12. It should be noted, however, that Ceuta and Melilla are incorporated into the sampling frame in July 2013. This issue has been taken into account when constructing the intra-quota comparison populations.

Once the previous points have been addressed, we are in a position to define the comparison populations for the intra-quota distributions. Specifically, if we denote by $p^t_{e,s}$ the proportion of Spanish residents with age e, sex s at time t, and by $\pi^t_{g,s}$ the proportion of people who, by design, should be selected within the age group g with sex s at time t, we would have that, without the effect of R1 or R2, the proportion (relative size) of the expected intra-quota of Spanish residents with age e, sex s at time t for the population as a whole, $Cp^t_{e,s}$, will be given by equation (1)[13], and that, on the other hand, the expected intra-quota proportion of Spanish residents within the corresponding quota, $Dp^t_{e,s}$, will be given by equation (2).

$$Cp^t_{e,s} = \frac{\sum_{g\in G} p^t_{e,s}\cdot \delta^t_{e,s}\cdot \pi^t_{e,s}}{\sum_{g\in G}\sum_{e^*\in E_g} p^t_{e^*,s}\cdot \delta_{e,g}}\cdot \frac{1}{\sum_{g\in G}\pi^t_{g,s}} \tag{1}$$

$$Dp^t_{e,s} = \frac{p^t_{e,s}}{\sum_{g\in G}\sum_{e^*\in E_g} p^t_{e^*,s}\cdot \delta_{e,g}} \tag{2}$$

where $s \in S = \{W\text{: Women}, M\text{: Men}\}$, $e \in E = \{18, 19, 20, \ldots, 84, 85+\}$[14], $g \in G = \{18-24, 25-34, 35-44, 45-54, 55-64, 65+\}$, E_g denotes the set of ages included in the group g (for example, $E_{18-24} = \{18, 19, 20, 21, 22, 23, 24\}$), and $\delta_{e,g} = 1$ if $e \in E_g$ and, otherwise, equal to zero.

The previous notation is complex but its meaning is easy to understand by looking at the following example. Supposing that (i) the percentage of Spanish women residents (target population) of 18, 19, 20, 21, 22, 23 and 24 years old at a given moment are respectively, 0.8%, 0.7%, 1.0%, 1.3%, 1.1%, 1.2% and 1.5% ($p^t_{18,W}, \cdots, p^t_{24,W}$) (ii) the total percentage of women resident in Spain between 18 and 24 years old is 7.0% ($\pi^t_{18-24,W}$) and (iii) the total of residents of 18 years or more in Spain is 85% ($\sum_{g\in G}\pi^t_{g,s}$), then it can be seen that, without the effect of R1 and R2, the expected proportion of women aged 20 in the total population interviewed, $Cp_{20,W}$, and the expected proportion of women aged 20 within the corresponding quota in the surveyed population, $Dp_{20,W}$, would be:

$$Cp_{20,W} = \frac{1.0\%\cdot 7.0\%}{0.8\%+0.7\%+1.0\%+1.3\%+1.1\%+1.2\%+1.5\%}\cdot\frac{1}{85\%} \approx 1.08\%$$

$$Dp_{20,W} = \frac{1.0\%}{0.8\%+0.7\%+1.0\%+1.3\%+1.1\%+1.2\%+1.5\%} \approx 13.16\%$$

In other words, these proportions are no more than (i) the relative weight that each age represents in the target population, re-weighted by the design weight of the group

13. Note that $p^t_{e,s}/\sum_{e\geqslant 18} p^t_{e,s} = Cp^t_{e,s}$, if the target and reference populations coincide.

14. In the intra-quota comparisons, people of 85 years or more (85+) have been added to simplify the graphical presentations and to facilitate the verification of the theoretical conditions that are required by the hypothesis tests of goodness-of-fit implemented.

to which it belongs, and (ii) the relative weight that each age represents within its group (quota). Obviously, when working with the subset of men or women, the corresponding Cp proportions would be calculated in a similar way.

In general, the comparisons are made within each quota (Dp proportions), as the results are sharper and easier to interpret. For an overview — i.e., when we compare the empirical and theoretical (expected) distributions of all ages together — we will use Cp proportions, since the distributions equivalent to those collected in the barometers are obtained by re-weighting the intra-quota distributions of the target populations with the inter-quota distributions actually used in the design of each barometer.

In fact, the demographic structure defined by the set of proportions $\{p^t_{e,s}/\sum_{e\geqslant 18} p^t_{e,s}\}$ should not be used as a population comparison to evaluate the impact of rules R1 and R2 on the set of empirical intra-quota proportions since part of the discrepancies that could be observed would not be attributable to the effect of R1 and R2 but would be a consequence of the possible differences existing between $\pi^t_{g,s}$ and $\sum_{e^*\in E_g} p^t_{e,s}\cdot\delta_{e,g}$. A different issue would be to study to what extent the distribution by age and gender of the responses collected in the barometers matches the corresponding structure of the target population, regardless of whether the discrepancies originate from the use of R1 and R2 rules or from the differences between target and design populations. In that case, it would be pertinent to use the set of proportions $\{p^t_{e,s}\}$, duly weighted by restricting to persons of 18 years of age or older, $\{\sum_s p^t_{e,s}/\sum_{e^*\geqslant 18} p^t_{e^*,s}\}$, or men or women of 18 years of age or older, $\{p^t_{e,s}/\sum_{e^*\geqslant 18} p^t_{e^*,s}\}$. For completeness, we also carry out this comparison (see bottom panel of Figure 3). In fact, we think that the differences between the discrepancies of both comparisons could be interpreted as an indicator of the practical impact of the divergences between target and design populations.[15]

To calculate the comparable intra-quota distributions, however, we need to know the distributions of the target populations in each of the months in which the barometers were carried out. The difficulty lies in the fact that Spanish resident population statistics are only published once a year (referenced, except in 1996, to January 1), so we have proceeded to estimate them in each of the months where official statistics are not available. Specifically, we estimate the number of resident Spanish people with age e, sex s in a month τ, $N^\tau_{e,s}$, located between two Municipal Registers referenced in the months t and t^*, with $t<\tau<t^*$, as a weighted average of the corresponding official figures, $N^t_{e,s}$ and $N^{t^*}_{e,s}$, with inverse weighting coefficients to the number of months away. The mathematical expression of the estimator is given by equation (3).

$$N^\tau_{e,s}=\frac{N^t_{e,s}\cdot(\tau-t)^{-1}+N^{t^*}_{e,s}\cdot(t^*-\tau)^{-1}}{(\tau-t)^{-1}+(t^*-\tau)^{-1}} \tag{3}$$

15. Obviously it would be simple to make a direct comparison between the intra-quota distributions of the target and design populations, for which no data needs to be collected. However, these discrepancies would only highlight the impact of the differences between these populations and would not take into account the possible interactions with R1 and R2.

The solution to equation (3) is equivalent to assuming a uniform distribution of demographic events between two consecutive censuses (Lledó, Pavía and Morillas, 2017).

3.3. Intra-quota and inter-quota distributions. Discussion

At this point and before making the comparisons, it is time to think over on the different nature of intra-quota discrepancies (understood as the deviations between the theoretical percentages of people that would be expected by chance to be interviewed within each group of age/gender and those that were really interviewed) and inter-quota discrepancies.

The deviations, statistically significant, in the intra-quota distributions will be an indicator of a non-representative selection of subjects within each quota and of an effect of the rule R1 (and perhaps also of R2). On the other hand, inter-quota deviations theoretically should not exist, since the sample design imposes to the sample to mimic the reference population. There are several reasons, however, why there could also be divergences in the inter-quota distributions, that is, between the sizes of the reference and collected quotas. Discounting (at national level) the possible rounding effect, the discrepancies could occur as a consequence of rule R2.[16] If this rule manifests any effect, however, it should not be very pronounced because the random substitutions between adjacent quotas should occur sometimes in one direction and sometimes in the opposite direction. In any case, given that in practice it is easier to find older people in households (Díaz de Rada, 2014), we expect a greater propensity to make substitutions towards higher age quotas and, therefore, a slight underrepresentation of the youngest quotas.

Regarding the empirical intra-quota distributions, we would initially expect that, at random, the number of people interviewed within each age group would replicate the corresponding distribution of the quota, with the relative higher probability of older people being at home not playing any significant role here, given that within each quota the age differences are relatively small (except for the group of 65 and older). However, due to rule R1, which recommends interviewing the youngest person if there is more than one person suitable for the survey in the home, our a priori is that within each quota there will be overrepresentation of younger people.

4. Methodology

Our main hypothesis is that as a consequence of R1 there will be an overrepresentation of younger people within each quota. This implies, on the one hand, that the proportion of people interviewed with the minimum age of each quota (18, 25, 35, 45, 55 and 65

16. The possible effect of non-response is negligible here. In the 220 barometers analysed, 99.42% of the 550,000 planed interviews (546,789) were carried out.

years) will tend to be greater than the corresponding proportion in the target population and, on the other hand, that the distribution of people surveyed within each quota will not adjust to the theoretical distribution of the target population. Regarding the inter-quota distributions, our hypothesis is that these will be adjusted to the designed distributions, that is, to the distributions of the reference populations used to determine the quotas during the planning of the surveys.

To evaluate these hypotheses we have used tests of hypothesis. On the one hand, unilateral parametric hypothesis tests for a proportion, with null hypotheses $p_m^{\upsilon} \leq p_m^t$ and alternative hypotheses $p_m^{\upsilon} > p_m^t$, where p_m^{υ} represents the true probability that a person of minimum age (m = 18, 25, 35, 45, 55, 65) in the corresponding quota is interviewed according to the fieldwork rules and p_m^t the theoretical proportion of people with that same age in the target population. On the other hand, classical goodness-of-fit χ^2 tests, where the null hypotheses postulate that the empirical distributions conform to the theoretical one and the alternative hypotheses state that they do not fit. Given the tendency of hypothesis tests, especially parametric ones, to accept null hypotheses, the rejection of these, mainly of $p_m^e \leq p_m^t$, will provide strong evidence in favour of our hypothesis regarding intra-quota distributions.

It is known, however, that the classic goodness-of-fit tests (χ^2, Kolmogorov-Smirnov, Kuiper, ...) have a lot of statistical power, so they tend to reject the null hypothesis when the sample size is extremely large (Badal-Valero, Alvarez-Jareño and Pavía, 2018). Hence, as an alternative to the classic χ^2 test, we have also implemented, when working with very large samples, the equivalent Monte Carlo test. This test has as a limit, when the sample size tends to infinity, the uniformly most powerful test associated with the hypothesis (Hope, 1968).

In addition to the results of the statistical tests, which we analyse through p-values, we have also used a classic indicator of dissimilarity to numerically evaluate the degree of adjustment between empirical and theoretical distributions. Specifically, we have employed the mean absolute statistical error, Δ, to measure the average percentage difference between distributions. For the inter-quota comparisons the statistical mathematical expression is given by equation (4) and for the intra-quota comparisons by equation (5)[17].

$$\Delta = \frac{1}{12} \sum_{s \in S} \sum_{g \in G} |\pi_{g,s}^{t*} - \widehat{\pi_{g,s}^{t*}}| \tag{4}$$

$$\Delta = \frac{1}{|E_g|} \sum_{e \in E_g} |Dp_{e,s}^t - \widehat{Dp_{e,s}^t}| \tag{5}$$

where $\pi_{g,s}^{t*} = \pi_{g,s}^t / (\sum_{s \in S} \sum_{g' \in G} \pi_{g',s}^t)$ represents the theoretical percentage that, by design, corresponds to the quota of group g and sex s, reweighted so that the sum of all

17. When working with Cp proportions, equation (5) is of course modified.

the quotas is 100%, $\widehat{\pi^{t*}_{g,s}}$ is the corresponding value in the sample, $\widehat{Dp^t_{e,s}}$ is the sample estimation of $Dp^t_{e,s}$ and $|E_g|$ represents the number of ages included in E_g.[18]

The presentation of results using only statistical summaries is very useful and succinct, but it can also obscure other interesting facts and make communication of the results less agile. Therefore, in line with recommendations made in Cleveland (1993), we decided to make also comparisons from a graphical perspective. This has made it possible to discover aspects present in the data that could otherwise have gone unnoticed. In our opinion, the power of the graphic representations (presented in the article and in the supplementary material) is such that they will convince the most sceptical reader of the conclusions derived from this study. All statistical calculations and analyses have been performed in version 3.3.1 of R (R Core Team, 2016).

5. Results

This section shows the main results of the comparisons made to study the effect of the fieldwork rules R1 and R2 on the age distributions collected in the 220 monthly barometers carried out by the CIS between January 1997 and December 2016. In addition, the supplementary material, which accompanies this article, considerably broadens the analyses presented in this section, especially in relation to intra-quota distributions.

In the first subsection of this section, a comparison is made between the theoretical and empirical inter-quota distributions, which are expected to show no statistically significant differences. In the second subsection, the analysis of intra-quota distributions is discussed; this shows a greater and richer range of results.

5.1. Inter-quota analysis

As can be seen in Figure 1, where the comparison between the empirical and theoretical inter-quota distributions are grouped according to the Municipal Register used in each barometer to determine the quota sizes, both sets of distributions are very similar, as expected. The greatest differences between distributions are detected for the barometers that used the 1995 Municipal Register for the sample design. This is not surprising since the data actually used in the barometers of that stage are not available and, as explained in the third section and detailed in the supplementary material, they had to be estimated from the Population Now-Cast estimates.

In fact, except for the data set associated with the surveys made with reference to the 1995 Register, the null hypothesis of adjustment of the empirical data to the corresponding theoretical distribution is not rejected for any other set of observations with the usual level of significance (α) of 5%. When the responses of all the barometers

18. The number of ages in E_g equals to 7 when $g = 18-24$, equals to 10 when $g \in G = \{25-34, 35-44, 45-54, 55-64\}$ and equals to 21 when $g = 65-85+$.

__Figure 1:__ Comparison of distributions between theoretical (design, INE) and empirical quotas (collected in the corresponding barometers, CIS). The comparisons have been grouped according to the Register used to determine the quotas for each survey. As a summary, the last two panels offer an aggregate comparison, with and without the barometers that used the 1995 Register as the reference population. As well as the graphical comparison, each panel shows the p-value associated with the χ^2 goodness-of-fit test, the number of observations used (sample size, n) and the value of the dissimilarity statistic Δ defined in equation (4).

are combined and compared with the population obtained as a weighted average of the different theoretical populations, the null hypothesis is rejected however. This occurs despite the fact that the dissimilarity statistic value Δ for the set of barometers associated with the Registers from 2004 to 2015 is very small, with an average discrepancy of only 0.08%. This rejection may be due to the known propensity that classic goodness-of-fit tests show to reject the null hypothesis when the sample size tends to infinity. Using the equivalent Monte Carlo test we obtained a p-value of 0.0112, which would not result in a rejection of, in this case, the null hypothesis for $\alpha = 1\%$. The same exercise for the set of observations associated with the 1995 Register or the set of all the Registers, however, continues producing p-values smaller than 0.0001 and indicating the need to reject the null hypothesis in these cases.

Although in terms of joint distributions the R2 rule does not show any effect that represents a statistically significant overall deviation of the theoretical inter-quota distributions, a result is visible in the data, which had already been foreseen in subsection 3.3, related to the existence of a certain tendency towards underrepresentation of the lower age quotas with respect to the designed sizes. This result is consistent with the greater difficulty that exists in practice of encountering younger people to interview and the consequent greater likelihood of substitution for the adjacent higher age quota, as suggested in the application of R2. In fact, we find in 27 of the 30 analysed quotas corresponding to groups of 18 to 24 years old (90%) that the percentage of respondents is lower than the percentage designed. In the case of women, the quota is underrepresented, on average, by 0.25%. This figure is reduced to 0.11% in the case of men, increasing to 0.18% when excluding the three cases in which the proportion of respondents belonging to the quota of men between 18 and 24 years is above the proportion designed. This trend also manifests itself in the following quotas of groups between 25 and 34 years, whose percentage of underrepresentation reaches 70%. Overrepresentation is predominant for the rest of the groups.

5.2. Intra-quota analysis

Just as expected, the sizes of the inter-quota distributions collected fit to the sizes designed in the barometers, with the possible deviations introduced by R2 not having statistically significant effects. In the same vein, the results for the intra-quota distributions are also as expected, in this case, for not adjusting to the theoretical distributions. In fact, as shown in Figure 2, where the comparison between empirical and theoretical intra-quota distributions for the set of barometers is shown graphically, the null hypothesis, which states that within each quota the empirical distributions by ages conform to the theoretical ones, is rejected for all age and gender groups.

Added to the generalized rejection of the hypothesis of equality between empirical and theoretical intra-quota distributions is the expected rejection of the hypothesis that, in each quota, the observed proportion of respondents with the minimum age of the quota is equal to or less than the corresponding theoretical proportion. This rejection, which is to be expected as a logical consequence of the use of the R1 rule, provides evidence to support our hypothesis. Looking more closely at the different representations shown in Figure 2 (and also in Figure 3, where all the information contained in the different panels of Figure 2 is presented together), the previous conclusions are not the only interesting conclusions to be made. In addition to the sample overrepresentation that is apparent for the minimum ages of each quota, two further results stand out. On the one hand, it is clear that for each age/gender group, the upper limits of each quota are also overrepresented. On the other hand, similarly, the ages ending in zero are also generally overrepresented, excluding 20 years[19].

The existence of overrepresentation of the 'whole' ages (finishing in zero or in five) in censuses and surveys has been a recognized dysfunction for many years (e.g., Myers, 1940; Bachi, 1951 and Carrier, 1959), which usually manifests itself more strongly in populations with lower levels of education, as various studies on historical populations or more recent transnational analyses have shown (Brian, Baten and Crayen, 2009; Lyons-Amos and Stones, 2017). In this literature, the numbers ending in zero and, to a lesser extent, in five are seen to act as poles of attraction (rounding) for the interviewees when declaring their age, causing relative gaps in the adjacent ages. These same patterns are also observed, in part, for the ages ending in zero in the empirical age distributions of the CIS barometers, especially after age 40. In our data, a greater gap is observed in the adjacent older ages than in the adjacent younger ones, which could suggest that the rounding responds to a psychological self-deception on the part of the interviewees who wishes to consider themselves younger or to a certain effect of social desirability in an attempt to appear younger before the interviewer.

For ages ending in five, which coincide with the lower limits of our intervals (except for the lower age quota), we also observe overrepresentation, although the same previous patterns as in ages ending in zero are not observed in Figure 2 or, if anything, to a small degree for ages 45 and 65. For these ages, the fact that the ages 46 and 66 show values below 47 and 67, respectively, would suggest that a part of the overrepresentation observed for the minimum ages of these intervals could be a consequence of the attraction effect towards 'whole' ages. Fortunately, for our hypothesis regarding the impact of R1, two results clearly point out that the distortion effect of R1 in the intraquota distributions is genuine and of appreciably greater intensity. On the one hand, the overrepresentation shown by the age(s) 18 (and 19) cannot be attributed to the effect of attraction towards 'whole' ages. On the other hand, our data show an overrepresentation of ages ending in five consistently greater than the overrepresentation of ages ending in zero (see also Figures S3 to S137 in the supplementary material), whilst the literature states the overrepresentation of ages ending in five as being systematically of less intensity.

---

19. Although in Figure 2 for the 30 years age group there is no overrepresentation in the strict sense, it is notable that the percentage of people who claim to be 30 years old is noticeably greater than those who claim to be 29 or 31 years, which is a clear indicator of relative overrepresentation.

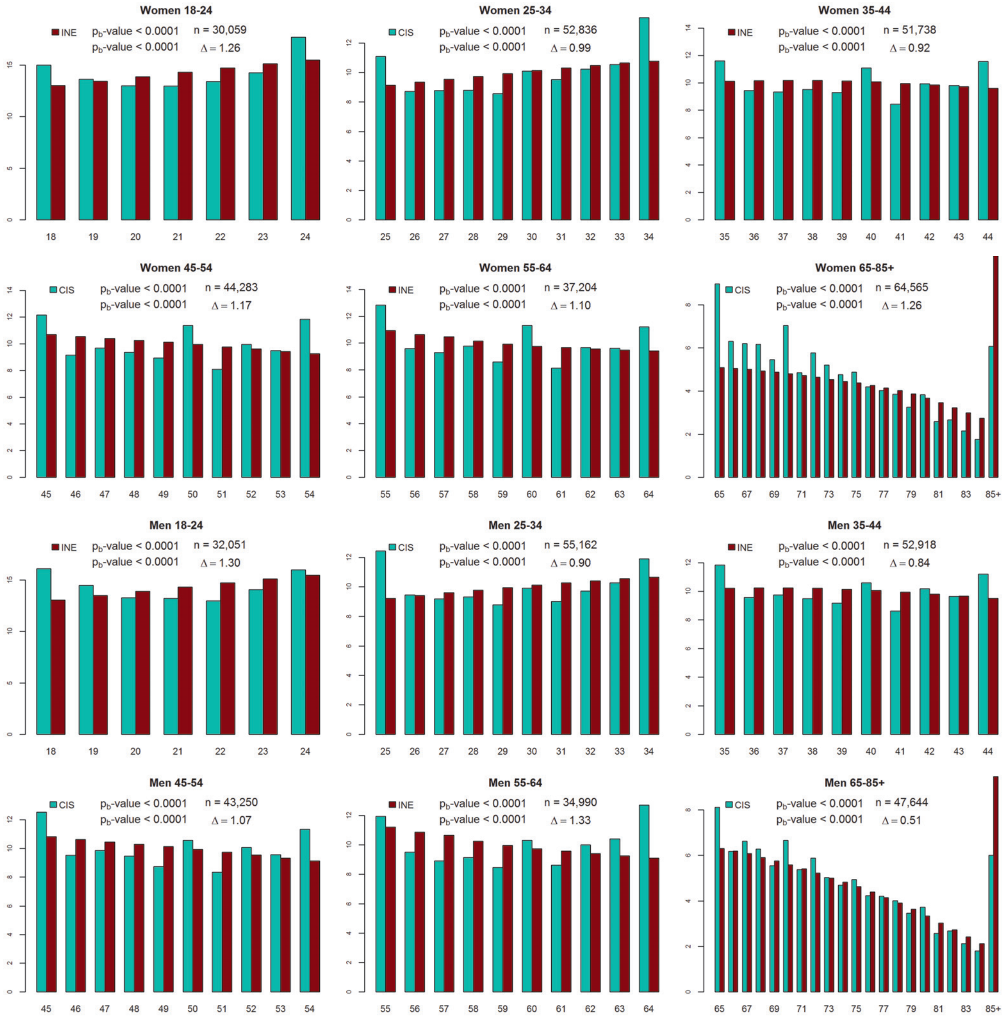

Figure 2: *Comparison between the theoretical intra-quota distributions (in the target population, INE) and the empirical distributions (set of responses collected in the 220 barometers analysed, CIS). The theoretical distributions have been calculated as the sum of the theoretical distributions associated with each barometer. In addition to the graphical comparison, each panel shows the p_b-value associated with the χ^2 goodness-of-fit test, the p_m-value associated with the unilateral test for the minimum proportion of each quota, the number of observations used (size of the sample, n) and the value of the dissimilarity statistic Δ defined in equation (5).*

Regarding the other surprising result, that of the huge overrepresentation shown by the upper limits of all quotas of age and gender (except obviously in the groups of 65 and over), our assumption (impossible to contrast with our data) is that this is due to the way the interviewers apply the R2 rule. Our assumption would be that, in order to minimize their work time, the interviewers are more willing to use the R2 rule as the field work progresses. Specifically, we conjecture that when interviewers have fewer and fewer

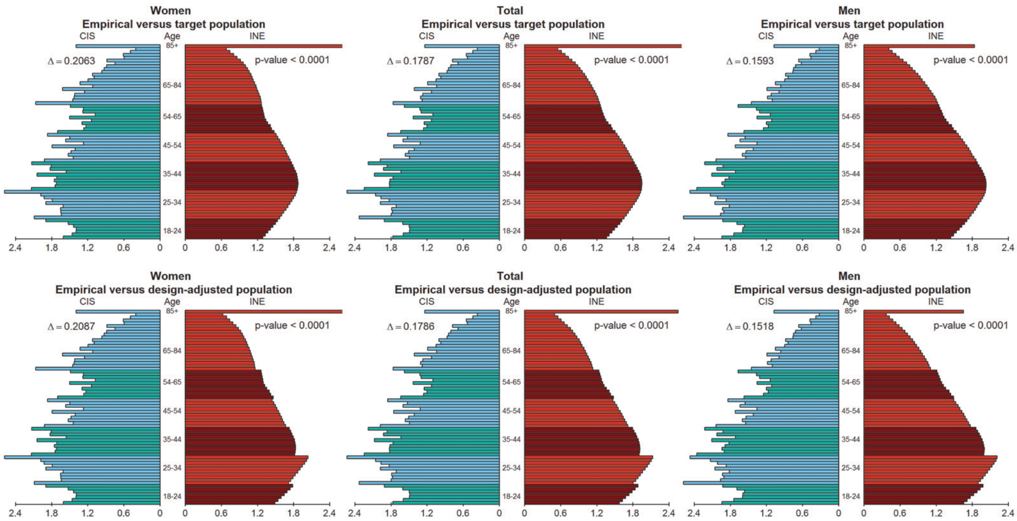

Figure 3: *Comparison between the empirical percentages of interviewees of each age group (for women, men and total) for the set of barometers (CIS) and the theoretical percentages expected according to the target and design population (INE) of the survey (upper panel) and the target population (bottom panel). In the panels corresponding to women and the total, the right axis is truncated (the percentage of people with 85 years and more is higher than shown) in order to make the results clearer. The theoretical percentages are calculated as a weighted sum of the theoretical percentages corresponding to each stage and each barometer. In addition to the graphical comparison, each panel shows the p-value associated with the* χ^2 *goodness-of-fit test and the value of the dissimilarity statistic* Δ *defined by equation (5) adapted to the* Cp *proportions.*

free quotas available they tend to take advantage of the R2 rule to fill a missing quota with a person whose age is one year either side of that quota as soon as s/he is available. This assumption would explain the observed overrepresentation of the upper limits and would help to explain part of the overrepresentation of the lower limits of the different quotas, except, for obvious reasons, in quotas of ages from 18 to 24.

The alternative explanation according to which the overrepresentation of the upper limits within each quota could be mainly due to the positive correlation that exists between age and time spent in the home does not hold. On the one hand, a big difference is seen between ages ending in three and those ending in four and, on the other hand, the figures for ages ending in two and in three are practically the same.

The observed patterns in the data set as a whole do not seem to be a consequence of isolated operations or trends, occurring only at certain moments of time or in a part of the territory, but to a generalized global trend that extends across all provinces, autonomous communities, years and stages, as can be seen in Figures S3 to S137 and in Tables S3 to S6 of the supplementary material. For example, the average overrepresentations at provincial level in the intra-quota percentages of men aged 35, 40 and 44 are, respectively, 1.6%, 0.5% and 1.7%, affecting 80.8%, 59.6% and 76.9% of the provinces, respectively (see Table S4). This result, along with many others visible in the data (see Figures S3 to S104 and Tables S3 to S6), would reinforce the likelihood of our conjecture about how interviewers apply R2.

Finally, by performing a combined analysis of the intra-quota distributions, we see, as expected, that the empirical distributions are more adjusted to the populations defined by the Cp proportions (that is, incorporating the differences between the design and target inter-quota distributions) than those of target populations. Although this is not entirely evident by comparing the values of the dissimilarity coefficients Δ of the upper and lower panels of Figure 3, this result becomes unquestionable when analysing Figures S105 to S137 of the supplementary material. For example, considering exclusively the values of Δ for the years 2006 to 2016,[20] we see that, on average, the dissimilarity value Δ is 11.8% higher when the combined empirical distributions are compared with the target populations (upper rows of the panels) than when compared with the adjusted-target populations (lower rows of the panels).

6. Summary and concluding remarks

Due to budget and time constraints, quota samples are extensively used by opinion pollsters and consumer researchers all over the world. Quota sampling, however, is strongly contested by the statistical community due to its break from randomness. Hence, to achieve representative samples, the more prestigious organizations relying on this approach try to follow wherever possible random steps and rules in their sampling designs and only use quota sampling in the final step. CIS is one of these organizations. Their surveys are far removed from traditional quota sampling, being closer to probability sampling. Indeed, surveys carried out by CIS, in particular its monthly barometers, are a point of reference for Spanish market research and public opinion polls.

Whatever the approach adopted, however, we should not overlook the detail. This paper shows that even seemingly innocuous fieldwork rules can have significant consequences, biasing the collected samples. Our analysis tries to warn practitioners about the need of thinking carefully about all the components that define a sampling design. We show that, for selecting individuals within the household, the CIS fieldwork rules R1 (of selection, in case of doubt, between and within quotas) and R2 (of substitution, in case of difficulty, between quotas) should be improved, since they lead to biases in the distribution by age within each quota. In this article we study, using the more than 545,000 responses collected in the 220 barometers conducted between 1997 and 2016, the effect of these rules on age distribution and analyse the existence of bias comparing the empirical distributions with the theoretical ones (those expected according to the sample design and/or the target populations).

20. An assessment excluding the barometers designed using the 1995 Register is pertinent as it avoids the possible distorting effect that could introduce in the analysis the use of estimated data for inter-quota distributions associated with the barometers referenced to the 1995 Register.

The study implemented reveal a set of highly interesting results that support the existence of biases in the intra-quota distributions of the CIS barometers. Regarding the inter-quota distributions, it is observed that rule R2 induces a certain tendency for underrepresentation of younger quotas, consistent with the lower probability of encountering younger people in the home, although this causes no statistically significant deviation. The most interesting results are seen for the intra-quota distributions.

Specifically, within all quotas the hypothesis that the empirical distribution by ages fits the theoretical distribution is rejected, illustrated by the three overrepresentations highlighted. Firstly, as a logical consequence of the use of the R1 rule, the results show that there is a significant overrepresentation of the proportion of interviewees in the survey having the minimum age of each quota. Secondly, as has been previously observed in dozens of studies, we detect the existence of a significant overrepresentation of ages ending in zero. Thirdly and unexpectedly, it is notable that the maximum ages of each quota (excluding older quotas) are clearly overrepresented.

Regarding the last overrepresentation, our speculation is that this is a consequence of R2. We surmise that, shielded by rule R2, interviewers tend to choose a person from an adjacent quota when the person to be interviewed is only one year of age away from a quota not covered. If this conjecture were true, this would also help to explain part of the overrepresentation observed in the minimum ages of each quota. Obviously, our previous supposition is questionable. We could ask ourselves why, if R2 allows interviewers to substitute respondents with whatever age of an adjacent quota, do they decide, as a rule, to restrict themselves to just one year either side of the quota. We venture that maybe, just to stop what they can anticipate it will be a hard search, they are reluctant to discard an encounter when they find a willing respondent who is only a year away from the preferred quota, believing that this substitution would not represent a significant deviation. In order to confirm (or reject) our supposition, we would need to know in each census section the order followed to complete all the interviews. This will allow us to build the ordered distribution of ages of the interviewees and to assess the likelihood of our assumption. Furthermore, having access to the "sampling sheet" would also allow us to check if interviewers follow the rule of performing a quota-age substitution up to a maximum per census section.

Regarding the other sources of overrepresentation, these could be easily remedied. On the one hand, to avoid the effect of R1, if the eligible younger persons belong to the same age group (including the quotas of women and men) the choice could be made at random, for example, using the Kish tables (Kish, 1995). On the other hand, to eliminate the attraction effect of 'whole' ages, the solution would be to ask the date of birth instead of the age at last birthday, although later, to reinforce the anonymity of the respondents, only the age at last birthday should be recorded in the corresponding microdata file.

The biases encountered do not invalidate the samples collected by CIS; they simply represent new challenges in research. The same way that analysts try to amend the collected data from the deviations induced by (total or partial) non-response or introduced as a consequence of using quota sampling in the last step (which, as it is well-known, tends to over-represent the persons that spend more time at home), these new biases should also be taken into account to carry out proper inferences and model analyses. When making inferences, like estimating the voting intention or the climate economic mood of the whole population, the detected biases should be accounted for to correctly weigh the probability that each individual age being included in the sample. Otherwise, given the known inverse relationship between age and ideology in the left-right axis, we might obtain, for example, estimates slightly biased to the left side when making either assessments about global attitudes and values in Spanish society or estimates of trends in public opinion. In our view, however, when modelling, these biases would not introduce additional deviations to the ones already presented in the data[21] once the individual age (not the age group) of each observation was included in the model as an explanatory (predictor) variable. Conditional on age, rules R1 and R2 would have no effect on the collected data, other biases apart. In any case, despite CIS data limitations, it should be noted that the CIS databank comprises the largest and most reliable public database available in Spain for the study of social and political issues. The datasets of CIS can be free downloaded from its website and, as we have already mentioned, its sampling designs has been previously tested with acceptable results by the academia (Stephenson, 1978).

Acknowledgments

We are grateful to Valentín C. Martínez Martín (technical advisor of the CIS research department) for his first-rate assistance kindly answering all our doubts about the CIS sampling designs and to the anonymous referees and an editor for their valuable comments and suggestions. Any inaccuracy is the sole responsibility of the authors. We wish also to thank M. Hodkinson for translation of this paper into English and the support of the Spanish Ministerio de Economía y Competitividad under grants CSO2013-43054-R and ECO2017-87245-R.

References

Bachi, R. (1951). The tendency to round off age returns: measurement and correction. *Bulletin of the International Statistical Institute*, 33, 195–222.

Badal-Valero, E., Alvarez-Jareño, J. A. and Pavía, J. M. (2018). Combining Benford's law and machine learning to detect money laundering. An actual Spanish court case. *Forensic science international*, 282, 24–34. https://doi.org/10.1016/j.forsciint.2017.11.008.

21. As many other survey data, CIS datasets present weaknesses. Their main flaws, however, are not a result of the sampling design followed by CIS but coming from sources of bias (non-response, measurement error, coverage, ...) inherent to the survey approach. Other limitations presented in the data (in particular in the series of data) are, however, imputable to the institution. The historical changes to the board of directives of CIS have often been accompanied by swings in topics of interest, scientific orientation or sense of political opportunity. New issues appear only to disappear soon after or the same issue, just worded differently, is revisited time and time again over the years. This breaks the historical series and makes intertemporal comparisons, at best, difficult. Only a handful of topics (vote intention, government assessment, or political climate, among them) have proven to be resilient.

Brian, A., Baten, J. and Crayen, D. (2009). Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, 69, 783–808. https://doi.org/10.1017/S0022050709001120.

Burnap, P., Gibson, R., Sloan, L., Southern, R. and Williams, M. (2016). 140 characters to victory?: Using twitter to predict the UK 2015 general election. *Electoral Studies*, 41, 230–233. https://doi.org/10.1016/j.electstud.2015.11.017.

Carrier, N. (1959). A note on the measurement of digital preference in age recordings. *Journal of the Institute of Actuaries*, 85, 71–85.

Cea D'Ancona, M. A. (2004). *Métodos de encuesta: teoría y práctica, errores y mejora*. Síntesis, Madrid.

Cleveland, W. (1993). *Visualizing Data*. New Jersey: Hobart Press.

Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. and Quattrociocchi, W. (2017). Modeling confirmation bias and polarization. *Scientific Reports*, 7, 40391. https://doi.org/10.1038/srep40391.

Díaz de Rada, V. (2005). *Manual de trabajo de campo en la encuesta*. Cuadernos Metodológicos del CIS, 36. Centro de Investigaciones Sociológicas, Madrid.

Díaz de Rada, V. (2008). La selección de los entrevistados últimos en encuestas presenciales: un análisis de la utilización conjunta del método de rutas y el método de cuotas. *Revista Española de Investigaciones Sociológicas (REIS)*, 123, 209–247. https://doi.org/10.2307/40184898.

Díaz de Rada, V. (2014). Analysis of incidents in face-to-face surveys: Improvements in fieldwork. *Revista Española de Investigaciones Sociológicas (REIS)*, 145, 43–72.

Díaz de Rada, V. (2015). *Manual de Trabajo de Campo en la Encuesta*. Cuadernos Metodológicos del CIS, 36. Centro de Investigaciones Sociológicas, Madrid.

Ebrahimi, M., Yazdavar, H. and Sheth, A. (2017). Challenges of sentiment analysis for dynamic events. *IEEE Intelligent Systems*, 32, 70–75. https://doi.org/10.1109/MIS.2017.3711649.

Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E. and Tourangeau, R. (2009). *Survey Methodology*. John Wiley & Sons, New Jersey.

Hope, A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B*, 30, 582–598.

Jungherr, A., Schoen, H. Posegga, O. and Jürgens, P. (2017). Digital trace data in the study of public opinion: An indicator of attention toward politics rather than political support. *Social Science Computer Review*, 35, 336–356. https://doi.org/10.1177/0894439316631043.

Kalampokis, E., Karamanou, A., Tambouris, E. and Tarabanis, K. (2017). On predicting election results using twitter and linked open data: The case of the UK 2010 election. *Journal of Universal Computer Science*, 23, 280–303.

Kennedy, C., Mercer, A., Keeter, S., Hatley, N., McGeeney, K. and Gimenez, A. (2016). *Evaluating Online Nonprobability Surveys*. Pew Research Center.

Kish, L. (1995). *Survey Sampling*. John Wiley and Sons, Inc, New York.

Lind, K., Link, M. and Oldendick, R. (2000). A comparison of the accuracy of the last birthday versus the next birthday methods for random selection of household respondents. *Proceedings of the American Statistical Association, Survey Research Methods Section*, 23, 887–889.

Lledó, J., Pavía, J. M. and Morillas, F. G. (2017). Assessing implicit hypothesis in life table construction. *Scandinavian Acturial Journal*, 6, 495–518. https://doi.org/10.1080/03461238.2016.1177585.

Lyons-Amos, M. and Stones, T. (2017). Trends in demographic and health survey data quality: An analysis of age heaping over time in 34 countries in sub saharan africa between 1987 and 2015. *BMC Research Notes*, 10, 1–7. https://doi.org/10.1186/s13104-017-3091-x.

Marsh, C. and Scarbrough, E. (1990). Testing nine hypotheses about quota sampling. *Journal of the Market Research Society*, 32, 485–506.

Mellon, J. and Prosser, C. (2017). Twitter and facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research and Politics, July-September*, 1–9. https://doi.org/10.1177/2053168017720008.

Meng, X.-L. (2016). Statistical paradises and paradoxes in big data. In *Royal Statistical Society Annual Conference 2016*. https://www.youtube.com/watch?v=8YLdIDOMEZs.

Mosteller, F., Hyman, H., McCarthy, P., Marks, E. and Truman, D. (1949). *The Pre-Election Polls of 1948*. Social Science Research Council, New York.

Myers, R. (1940). Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America*, 41-2, 395–415.

O'Rourke, D. and Blair, J. (1983). Improving random respondent selection in telephone surveys. *Journal of Marketing Research*, 20, 428–432. https://doi.org/10.2307/3151446.

Pavía, J. M., Badal, E. and García-Cárceles, B. (2016). Spanish exit polls: Sampling error or nonresponse bias? *Revista Internacional de Sociología*, 74, e043. https://doi.org/10.3989/ris.2016.74.3.043.

Pavía, J. M. and García-Cárceles, B. (2012). Una aproximación empírica al error de diseño muestral en las encuestas electorales del CIS. *Metodología de Encuestas*, 14, 45–62.

Pavía, J. M. and Larraz, B. (2012). Nonresponse bias and superpopulation models in electoral polls. *Revista Española de Investigaciones Sociológicas*, 137, 237–264. https://doi.org/10.5477/cis/reis.137.237.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. htpps://www.R-project.org/.

Rodríguez Osuna, J. (1991). *Métodos de Muestreo*. Cuadernos Metodológicos del CIS, 1. Centro de Investigaciones Sociológicas, Madrid.

Rodríguez Osuna, J. (2005). *Métodos de Muestreo. Casos Prácticos*. Cuadernos Metodológicos del CIS, 1. Centro de Investigaciones Sociológicas, Madrid.

Smith, T. (1983). On the validity of inference from non-random samples. *Journal of the Royal Statistical Society, Series A*, 146, 394–403. https://doi.org/10.2307/2981454.

Stephenson, C. B. (1978). A comparison of full-probability and probability-with-quotas sampling techniques in the general social survey. *GSS Technical Report, no. 5. University of Chicago: NORC, Chicago*.

Sudman, S. (1976). *Applied Sampling*. Academic Press, New York.

Trochim, W. and Donnelly, J. (2006). *The Research Methods Knowledge Base*. Boston, Massachusetts: Cengage Learning, Boston, Massachusetts.

Vavreck, L. and Rivers, D. (2008). The 2006 cooperative congressional election study. *Journal of Elections, Public Opinion and Parties*, 18, 355–366. https://doi.org/10.1080/17457280802305177.

Vehovar, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics*, 15, 335–350.

Yang, K. and Banamah, A. (2014). Quota sampling as an alternative to probability sampling? an experimental study. *Sociological Research Online*, 19, 1–11. htpps://doi.org/10.5153/sro.3199.