

Article

# Treating Nonresponse in Probability-Based Online Panels through Calibration: Empirical Evidence from a Survey of Political Decision-Making Procedures

Antonio Arcos <sup>1,\*</sup>, Maria del Mar Rueda <sup>1</sup> and Sara Pasadas-del-Amo <sup>2</sup>

<sup>1</sup> Department of Statistics and Operational Research, University of Granada, 18071 Granada, Spain; mrueda@ugr.es

<sup>2</sup> Institute for Advanced Social Studies, 14004 Cordoba, Spain; spasad@iesa.csic.es

\* Correspondence: arcos@ugr.es

Received: 14 February 2020; Accepted: 11 March 2020; Published: 15 March 2020



**Abstract:** The use of probability-based panels that collect data via online or mixed-mode surveys has increased in the last few years as an answer to the growing concern with the quality of the data obtained with traditional survey modes. However, in order to adequately represent the general population, these tools must address the same sources of bias that affect other survey-based designs: namely under coverage and non-response. In this work, we test several approaches to produce calibration estimators that are suitable for survey data affected by non response where auxiliary information exists at both the panel level and the population level. The first approach adjusts the results obtained in the cross-sectional survey to the population totals, while, in the second, the weights are the result of two-step process where different adjusts on the sample, panel, and population are done. A simulation on the properties of these estimators is performed. In light of theory and simulation results, we conclude that weighting by calibration is an effective technique for the treatment of non-response bias when the response mechanism is missing at random. These techniques have also been applied to real data from the survey Andalusian Citizen Preferences for Political Decision-Making Procedures.

**Keywords:** calibration estimation; complex surveys; non response bias; probability panel

**MSC:** 62D05

## 1. Introduction

In the last few years, we are witnessing a strong development of online research methods in general and web surveys specifically. According to the World Association of Opinion and Marketing Research professionals (ESOMAR) data referring to 2016, 56% of the global revenues in the market research field was made using different kinds of quantitative and qualitative online methods. Online is also the preferred mode for survey-based research such as web or smartphone surveys accounts for two-thirds of the revenues on this chapter [1].

The growth of online research methods goes in parallel with the increasing difficulties that traditional modes, such as face to face or telephone interviews, are facing to obtain samples that achieve the highest quality standards [2–4]. Changes in the patterns of the distribution of working and leisure time, the incorporation of women to the workforce, and the increase in residential and geographic mobility have hampered the location of survey respondents. On the other hand, respondents are less willing to participate in surveys as they are more concerned with privacy and confidentiality issues and tired of the growing number of unsolicited requests from the survey and the marketing

industry [5–7]. As a result of these changes, the production of data that is fully representative of the general population is more and more costly both in terms of money and time [8].

However, at the same time, we have never been more connected and willing to share our thoughts and opinions in the comments section of online newspapers or in our social networks. In this context, there are more and more examples of research that extract public opinion information from the net and social media on issues such as hospitality and tourism management [9], urban mobility [10], health [11], or the social perception of energy [12], among others. In addition, new methods and techniques are being developed in order to take advantage of the enormous amount of information available from a wide range of online sources [13].

From the start of the modern Internet, in the 1990s, survey research practitioners decidedly embraced this technology as a way to contact and interview respondents and web surveys have become the predominant mode of survey data collection nowadays [14]. Web surveys are a very attractive option because of its cost-effectiveness, compared with other modes such as mail, telephone, and face to face, the speed of data collection, and the advantages associated with the computerization of the questionnaire (prevention of branching-errors, data validation, adaptive design, etc.) and self-administration (less intrusive and convenient for respondents) [15].

However, currently the web survey mode has some limitations to adequately represent the general population. In spite of the fast adoption of the Internet in the last few decades, the number of non-users is still important in most countries and their profiles differ significantly from internet users. Moreover, new divides are appearing among internet users related to the knowledge and the type of usage that they do on the Internet [16]. As a result, web surveys that fail to include non-internet or the less intensive internet users are at a high risk of incurring coverage bias [5,17]. A second problem that hinders the use of probability sampling in web surveys of the general population is the lack of a proper sampling frame of email addresses [14].

The main strategy developed in order to overcome those limitations is to use probability-based online panels, where the participants are recruited via survey modes that allow probabilistic selection (CATI, Computer-Assisted Telephone Interviewing, or CAPI, Computer-Assisted Personal Interviewing) and non-internet users are either offered a device and internet connection in order to be able to respond to the surveys, or interviewed with an alternative mode (mail, CATI or CAPI) [18,19]. Some examples of these tools are the Longitudinal Internet Studies for the Social Sciences (LISS) in the Netherlands, the Étude Longitudinal par Internet pour les Sciences Sociales (ELIPSS) in France or the German Internet Panel (GIP) and the GESIS Panel in Germany, among others [20].

Coverage or self-selection are not a problem for probability-based online panels as compared to access panels, but, as with any other survey, they suffer from non-response even if the use of responsive or adaptive design features (incentives, switch in modes, refusals conversion, etc.) account for participation rates that are significantly higher than the average web or CATI survey [19]. Nonresponse may be an important source of bias if respondents and non-respondents differ in key characteristics [21]. To minimize non-response bias and improve panel representativeness, panel data must be corrected by using different weighting adjustment techniques [17].

The two most common methods to weight online panels are post-stratification and propensity score weighting (see [19]). However, other more advanced weighting techniques, such as calibration or pseudo-empirical likelihood, can be used. The application of calibration for the treatment of non-response has been treated by many authors (e.g., [22–28], ...). These papers consider samples with a single sample design and a single non-response mechanism and a double calibration is used to treat the two types of auxiliary information that are available (at the sample level and at the population level). The extension of the calibration method to the case of probability panels is not straightforward. In the case of panels, non-response can occur at two different moments of the process: during recruitment and during each of the subsequent surveys [29], and their behavior patterns may be different. However, in the probability panels, there is an additional problem as there are two sources of randomness derived

from the two sample designs (the first by which the panel is obtained, and the second by which the sample is selected from the panel) that must be taken into consideration.

In this paper, we will adapt the calibration methodology to the case of probabilistic panels for the treatment of non-response. To the best of our knowledge, this is the first time that this methodology is studied in this context. We will propose different calibration methods in one and two steps and compare the efficiency of these estimators by means of a simulation study. Finally, we will apply the proposed techniques to a real survey conducted using the Citizen Panel for Social Research (PACIS), a probability-based mixed-modes panel of the Andalusian population recruited and maintained by the Institute for Advanced Social Studies (IESA-CSIC).

## 2. Proposed Calibration Strategies in Panel Surveys

Let  $y$  be the interest variable related to a finite population  $U = \{1, \dots, N\}$  consisting of  $N$  units. For this population  $U$ , a panel  $P$  of size  $N_P$  is drawn with probability  $p_{d1}(P)$  according to  $d_1$ . For this sampling design, the first order probabilities are given by  $\pi_i^{(1)} = \sum_{P \ni i} p_{d1}(P)$  for  $i \in U$ , where  $\sum_{P \ni i}$  denotes the sum of the panels containing the unit  $i$ . We denote by  $d_i^{(1)} = 1/\pi_i^{(1)}$  the panel weight for unit  $i$ .

In a second-phase, a probability sample  $s$  of size  $n$  is selected from the panel  $P$  with conditional probability  $p(s | P)$ . In this sampling design  $d_2$ , the first-order inclusion probabilities are denoted by  $\pi_{i|P}$ . For each unit in the panel  $P$ , we denote by  $d_i^{(2)} = 1/\pi_{i|P}$  the design weight from the panel.

The final sampling design  $p(\cdot)$  defined in  $U$  has first-order inclusion probabilities  $\pi_i = \sum_{P \ni i} p_{d1}(P)\pi_{i|P}$ ,  $i \in U$ . Let  $d_i = 1/\pi_i$  denote the sampling design-basic weight for unit  $i \in U$ .

We denote by  $y_i$  the value of the variable  $y$  associated with the  $i$ -th individual. We consider the linear parameter  $Y = \sum_{i \in U} y_i$ , the population total. We assume missing data on the sample  $s$  for the variable  $y$ , which can be divided into the disjoint sets

$$\begin{aligned} s_r &= \{i \in s / i \text{ responds to main variable}\} \\ s_m &= \{i \in s / i \text{ do not responds to main variable}\}, \end{aligned}$$

where  $s_r$  is the respondent sample is of size  $r$ , and  $s_m$  is the size  $n - r$ .

The natural candidate to estimate the population total is the Horvitz–Thompson estimator  $\hat{Y}_{HT} = \sum_{i \in s_r} d_i y_i$ ; however, it may lead to biased estimates because certain specific groups can be substantially under-represented. These errors can be overcome by the use of reweighting techniques. A standard weighting procedure is to rake or poststratify weights to external population control totals. Raking and post-stratification are special cases of calibration adjustment.

Calibration adjustments [30,31] can be used to extrapolate the estimations of a survey. Calibration estimates reduce non-coverage, non-response, and other biases [23,32]. We will adapt this methodology in our context.

To do so, we assume the existence of auxiliary information relative to several variables related to the main variable  $y$ . Auxiliary information can exist simultaneously at the population and the panel level, thus this auxiliary information is given as two vectors:  $\mathbf{x}_k^*$  and  $\mathbf{x}_k^o$ . Auxiliary information at population level is denoted by  $\mathbf{x}_k^*$ , and the population vector total  $\sum_U \mathbf{x}_k^*$  is known. Auxiliary information at panel level is noted by  $\mathbf{x}_k^o$ , and these values are known for  $k \in P$ . The total panel  $\mathbf{X}_P^o = \sum_P \mathbf{x}_k^o$  is thus known and the total population of these variables can be estimated by  $\sum_P d_k^{(1)} \mathbf{x}_k^o$ . We denote by  $\mathbf{x}_k$  the auxiliary vector constructed by combining the two vectors:

$$\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_k^* \\ \mathbf{x}_k^o \end{pmatrix} \tag{1}$$

and

$$\mathbf{X} = \begin{pmatrix} \sum_{\mathcal{U}} \mathbf{x}_k^* \\ \sum_P d_k^{(1)} \mathbf{x}_k^o \end{pmatrix} \tag{2}$$

We will use these auxiliary variables in different ways in the calibration process.

### 2.1. Calibration in One Step

Now, we use all auxiliary information to calibrate at population level.

We define a one step calibration estimator as:

$$\hat{Y}_{CAL1} = \sum_{s_r} w_k y_k, \quad w_k = d_k \cdot g_k \tag{3}$$

where the weights  $w_k$ , for  $k \in s_r$ , modify as little as possible, in an average sense, the original sampling weights  $d_k$ , while respecting the calibration equations

$$\sum_{s_r} w_k \mathbf{x}_k^* = \sum_{\mathcal{U}} \mathbf{x}_k^* \text{ and } \sum_{s_r} w_k \mathbf{x}_k^o = \sum_P d_k^{(1)} \mathbf{x}_k^o \tag{4}$$

or equivalently

$$\sum_{k \in s_r} w_k \mathbf{x}_k = \mathbf{X} \tag{5}$$

Then, given a distance measure  $G(w_k, d_k)$ , the calibration process consists of finding the solution of the following minimization problem:

$$\min_{w_k} \sum_{s_r} G(w_k, d_k) \tag{6}$$

while respecting the calibration Equation (5).

This calibration problem results in final calibrated weights  $w_k = d_k F(q_k \mathbf{x}_k' \lambda)$ , where  $F(\cdot)$  is the inverse of  $\partial G(w_k, d_k) / \partial w_k$ , where  $\lambda$  is the Lagrange multiplier and  $q_k$  are known positive constants used for scaling the calibrated weights (in many situations the value of  $q_k = 1, k \in s$ ).

Since Equation (6) depends on the chosen distance measure  $G(w_k, d_k)$ , each different distance measure leads to a specific weighting system and thereby to a new estimator. Many distance measures have been proposed for calibration. The authors in [33] consider several distance measures with desirable properties.

Under mild regularity assumptions, the authors in [33] also state the important result that all the calibration estimators given by (6) are asymptotically equivalent to the regression estimator. Thus, the choice of the distance measure  $G(w_k, d_k)$  has only a modest impact on important properties of estimators as the variance.

The raking ratio method (also known as the multiplicative method) uses the Kullback–Leibler pseudo-distance

$$G(w_k, d_k) = \frac{1}{q_k} (w_k \log(\frac{w_k}{d_k}) + d_k - w_k).$$

The calibrated weights are obtained as  $w_k = d_k \exp(q_k \mathbf{x}_k' \lambda)$ . The existence of solutions for this problem is not guaranteed (see [33,34]).

In linear calibration (the case of the Euclidean distance), the weights can be positive or negative, but raking calibration guarantees positive weights, with the same unbiasedness properties as generalized regression estimator; however, the model used for the adjustment of non-response has a slightly more reasonable form than the generalized regression estimator [28]. Finally, the raking method is an available procedure for calculating calibration weights in the best known statistical packages.

Note 1. The authors in [35] introduced an alternative calibration approach called the functional form calibration that gives the same estimators. In this paper, we consider the usual calibration approach because it is simpler to formulate.

Note 2. In some practical situations, the  $\pi_{i|P}$  for all possible panels  $P$  are not known. For these cases, one can use (following the idea of [36]) the  $\pi^*$ -estimator changing  $\pi_i$  by  $\pi^* = \pi_i^{(1)} \pi_{i|P}$  as the base estimator and thus calculate a similar  $\pi^*$ -calibrated estimator.

### 2.2. Calibration in Two Steps

Calibration is a very flexible technique to incorporate varied information at various stages. Next, we will explore several ways of doing two-stage calibration:

1. A two-step calibration method can be defined in this manner:

Step 1: Adjusting the representativeness of the panel in the population.

The calibration on population auxiliary variables whose population totals  $\sum_{\mathcal{U}} \mathbf{x}_k^*$  are known yields the calibrations weights  $w_k^{(1)} = g_k^{(1)} \cdot d_k^{(1)}$ ,  $k = 1, \dots, P$ , where  $d_k^{(1)}$  are the panel weights under the restrictions

$$\sum_{\mathcal{P}} w_k^{(1)} \mathbf{x}_k^* = \sum_{\mathcal{U}} \mathbf{x}_k^*.$$

Then, each unit in the panel has a weight that summarizes the auxiliary information obtained from the population.

Step 2: Adjusting the non-response of the sample in the panel.

The panel auxiliary information is incorporated through a calibration estimator whose calibrated weights  $w_k^{(2)} = g_k^{(2)} \cdot d_k^{(2)}$  (here the starting weights for calibration are the design weights of the units in the panel) verified

$$\sum_{s_r} w_k^{(2)} \mathbf{x}_k^o = \sum_{\mathcal{P}} \mathbf{x}_k^o.$$

The final weights are obtained by multiplying these weights and the two-step calibration estimator proposed is

$$\hat{Y}_{CAL2} = \sum_{s_r} w_k y_k = \sum_{s_r} w_k^{(1)} \cdot w_k^{(2)} \cdot y_k = \sum_{s_r} d_k^{(1)} \cdot g_k^{(1)} \cdot d_k^{(2)} \cdot g_k^{(2)} \cdot y_k.$$

2. By using the same idea as in [22], we can consider a second procedure:

Step 1. Computing first intermediate weights by calibrating the survey response set to the panel.

Thus, we calculate the weights by  $\min_{w_k} \sum_{s_r} G(w_k, w_k^{(2)})$  subject to

$$\sum_{s_r} w_k \mathbf{x}_k^o = \sum_{\mathcal{P}} \mathbf{x}_k^o.$$

The calibrated weights obtained in this first step are noted by  $w_{Pck}$ .

Step 2. Then, using these intermediate weights  $w_{Pck}$  for calibration to the population. The problem is now  $\min_{w_k} \sum_{s_r} G(w_k, w_{Pck})$  subject to

$$\sum_{s_r} w_k \mathbf{x}_k^* = \sum_{\mathcal{U}} \mathbf{x}_k^*$$

We denote the resulting calibration estimator by  $\hat{Y}_{CAL3}$ .

3. We consider other alternative by calibrating the survey response set to the sample that is:

Step 1. Compare the response set with the sample by  $\min_{w_k} \sum_{s_r} G(w_k, w_k^{(2)})$  subject to

$$\sum_{s_r} w_k \mathbf{x}_k^0 = \sum_s \mathbf{x}_k^0.$$

The calibrated weights obtained calibrating the survey response set are noted by  $w_{Sck}$ .

Step 2. Compare, using these intermediate weights, the sample with the population by  $\min_{w_k} \sum_{s_r} G(w_k, w_{Sck})$  subject to

$$\sum_{s_r} w_k \mathbf{x}_k^* = \sum_U \mathbf{x}_k^*.$$

We denote the resulting calibration estimator by  $\hat{Y}_{CAL4}$ .

Note that one particular auxiliary variable can be used in both steps (see the next Section 3).

Previously discussed calibration methods are defendable procedures. They may be comparable in terms of computational burden. In addition, the effectiveness of each method can be assessed by the theoretical variance of the estimator that cannot be expressed simply. We can apply automated linearization to each of the calibration estimators to find the approximate variance (see [37]).

### 3. Simulation Study

We select for this simulation study the adults of eusilcP data set (population size 47,123), a population available from the R-packages simFrame [38] and simPop [39]. As a panel, we draw a sample of 1500 households stratified by region and select all members in each household. This process is repeated 1000 times (average panel size over simulations 2857.607). For each panel, we select a simple random sample of size 1000 (this process is repeated 1000 times).

As target variable, we use `eqIncome`, a numeric variable about household income. As  $x^*$ -variables (for which are known the population total), we use `age`, `gender`, and `ecoStat`, a factor showing the person’s economic status. As  $x^0$ -variable (for which only are known the panel total), we select `hy040n`, a numeric income from rental of a property or land ( $\text{cor}(\text{eqIncome}, \text{hy040n}) = 0.387$ ).

For each panel, we generated missing values due to non-response in the sample using two mechanisms: Missing at random (MAR) with `age` as auxiliary variable and missing not at random (MNAR), with an auxiliary variable namely `benefits`, obtained as a sum of the variables referred to different types of subsidies ( $\text{cor}(\text{benefits}, \text{eqIncome}) = 0.530737$ ). Missing completely at random (MCAR) is not included in this simulation study because it is not a practical situation.

Uncalibrated estimator, HT, and alternative calibration estimators are computed in each simulation runs: CAL1 (one step calibration) and the three versions of calibration using two steps: CAL2, CAL3, and CAL4.

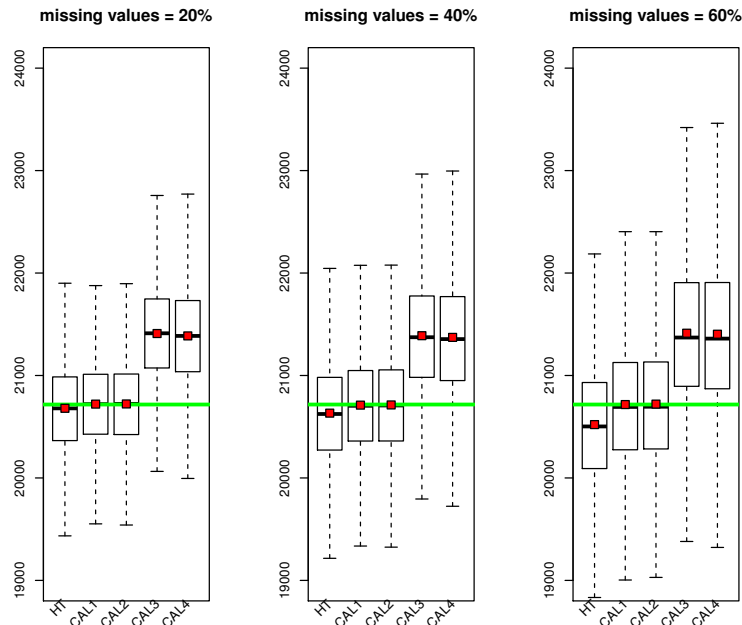
Figures 1 and 2 show the boxplots of our results. Green line is the true population mean, and blue square is the average over  $1000 \times 1000$  simulation runs for each method compared.

From Figures 1 and 2, some empirical evidence can be highlighted:

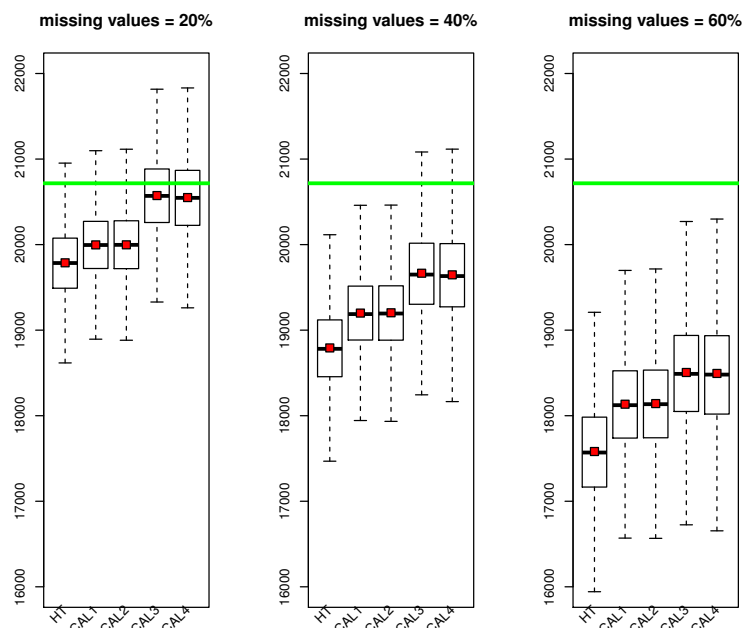
- The Horvitz–Thompson estimator is a biased estimator and the non-response bias increases with the non-response rate. The best behavior is taken by estimators CAL1 and CAL2: they correct bias very well even for the largest non-response rates and have the least variability. In this situation of MAR non-response, the calibration estimators CAL3 and CAL4 are not able to correct the non-response bias: they are even more biased than the HT estimator. Note that the bias of the Horvitz–Thompson estimator is negative, while the CAL3 and CAL4 estimators have a positive bias.
- The behavior of the estimators in the second scenario is very different. None of the estimators can correct the non-response bias (note that the green line is off the graph when the non-response rate

is 60%) especially when the non-response rate is large. The estimators that are best profiled to reduce them are the CAL3 and CAL4 estimators although the bias is very large for high response rates. This result was expected since the MNAR non-response is very complex to deal with.

- It is noteworthy that the estimators CAL1 and CAL2 have a similar behavior to each other as well as CAL3 and CAL4. It is also observed that the estimators CAL1 and CAL2 have less variability in general than CAL3 and CAL4.



**Figure 1.** Missing at Random case. Boxplots of compared estimators for missing rates 20%, 40% and 60%. Green line is the true population mean. Blue square is the average over simulations runs.



**Figure 2.** Missing Not at Random case. Boxplots of compared estimators for missing rates 20%, 40%, and 60%. Green line is the true population mean. Blue square is the average over simulations runs.

## 4. Application to the Pacis Survey on Andalusian Citizen Preferences for Political Decision-Making Processes

### 4.1. The Dataset

The survey Andalusian Citizen Preferences for Political Decision-Making Processes is part of a wider project titled *Why do we hate politics?* (PRY079/14). This project was led by two IESA/CSIC researchers and funded by the public foundation *Centro de Estudios Andaluces* [40] with the aim of analyzing the perception that lay citizens have on democracy and its limits. Fieldwork was conducted from 30 November to 19 December 2015, the day before the Spanish general election took place. This survey was the second survey conducted using the PACIS.

The PACIS is a probability-based, mixed-modes panel of the Andalusian general population. It was recruited offline following full probability sampling methods in the first semester of 2015. The target population was the general population with 16 years and older residing in independent private households in Andalusia. As obtaining a random sample of specifically named persons from the Spanish Continuous Population Register is restricted to official statistics and international surveys, we conducted address-based sampling using as a sampling frame the Spanish Cadaster, the administrative inventory of real estate that includes all the properties segmented by use (residence, rural, industrial, etc.). The cadaster includes the address of each property and geospatial data that was used to ascribe the addresses to a census section and to design the maps that helped the recruiters to locate the households selected to be part of the initial sample. Census sections are operational partitions of the territory that are defined by easily identifiable boundaries and group 1000 to 2500 residents each. From this sampling frame, households were selected using two-stage stratified cluster sampling where the primary sampling unit was the census section and the secondary sampling unit the household address. Census sections were stratified by province using proportional allocation. Once the household was located and contacted, the aim was to try to register all the persons residing in the household. Recruitment was made in two sequential phases using postal invitations first and face to face visits to households in the second place.

PACIS is a mixed-mode panel because, although web surveys are the main interviewing mode, CATI is used to interview offline population and online panel members that did not answer to online invitation to participate in the survey. In order to maximize response rates and lower attrition, participation is rewarded with 5 euros that the respondent may receive or donate to a charity. It must be noted that PACIS was not conceived to study changes over time, as it is the case with longitudinal panels, but as a valid pool of respondents for extracting cross-sectional samples.

The method used for selecting those survey samples is stratified random sampling with proportional allocation to the population distribution by sex and age.

### 4.2. Nonresponse in PACIS Surveys

As it has been explained, non-response occurs at two different moments in the case of panels: the recruitment stage and the subsequent surveys processes. Tables 1 and 2 show the final disposition of case codes and the outcome rates for the recruitment phase of the panel and the PACIS survey on Andalusian Citizen Preferences for Political Decision-Making Processes.

The overall response rate at panel registration was 21.6% (American Association for Public Opinion Research AAPOR RR3), a figure that is in line with the outcome rates obtained by the German Internet Panel (18.1%, RR3) and somewhat lower than more resourced panels like ELIPSS (27.3% AAPOR RR3) or the GESIS panel (25.1% AAPOR RR5) [20].



**Table 1.** Final disposition and outcome rates for the PACIS recruitment phase (face to face mostly).

		n
Interview		2803
	Valid registration after postal contact (online or hotline phone)	186
	Valid registration after F2F contact (visits to households)	2617
Eligible, non-interview		5886
	Refusals	5474
	Contacted, visit postponed	378
	Language barriers	34
Unknown eligibility		6123
	No answer	5690
	Unable to reach / Unsafe area	308
	Unable to locate address	125
Not eligible		2345
	Non residence	866
	Vacant housing unit	1479
Total of units in the sample		17,517
	Response Rate RR1	18.9%
	Response Rate RR3	21.6%
	Refusal Rate RF1	39.7%
	Cooperation Rate COOP2	32.3%
	Contact Rate CON2	64.3%

**Table 2.** Final disposition and Outcome rates for PACIS 2nd Wave Survey (Web + CATI).

		n
Interview		1081
	Conducted online	623
	Conducted by CATI	458
Eligible, non-interview		911
	Refusals	812
	Break off/ Implicit refusal (internet surveys)	64
	Logged on to survey, did not complete any item	33
	Telephone answering device (confirming HH)	2
Unknown eligibility		254
	No answer	254
Not eligible		152
	Fax/Data line	2
	Phone number doesn't match with panelist	150
Total of units in the sample		2398
	Response Rate RR1	48.1%
	Response Rate RR3	48.5%
	Refusal Rate RF1	40.5%
	Cooperation Rate COOP2	54.3%
	Contact Rate CON2	89.3%

Refusals at the respondent level are also the most important reason for not participating in the second wave PACIS survey. In addition, 2398 panelists were invited to participate in the survey and 1081 completed the questionnaire, which accounts for a response rate of 48.1%.

Table 3 compares the distribution of the panel members and the respondents to the second wave with the general Andalusian population according to continuous population register statistics. Given that, in both cases, sampling selection has followed a probabilistic procedure and that the recruitment and the interviewing modes are not subjected to undercoverage, we assume that the biggest part of the observed differences is caused by non-response.

**Table 3.** Sample percentages for population, panel members, and respondents to PACIS Second Wave Survey.

Sample Distribution	Panel Members	Diff. with Population	2nd Wave Respondents	Diff. with Population	Population
Women	57.1%	+6.0 %	49.2%	− 1.9 %	51.1%
Over 60 years old	16.9%	−9.0 %	19.1%	−6.8 %	25.9%
Residence					
Malaga	14.7%	−4.7 %	15.6%	−3.8 %	19.4%
Granada	14.5%	+3.6 %	12.3%	+1.4 %	10.9%
Education level					
Up to primary	19.9%	−8.1 %	13.3%	−14.7 %	28.0%
Secondary	58.0%	+8.4 %	59.5%	+9.9 %	49.6%
Tertiary	22.2%	−0.2 %	27.2%	+4.8 %	22.4%
Labor status					
Employed	37.4%	−1.6 %	38.2%	−0.8 %	39.0%
Unemployed	30.6%	+10.3 %	33.7%	+13.4 %	20.3%
Economically inactive	32.0%	−8.7 %	28.1%	−12.6 %	40.7%
Foreign nationality	3.8%	−2.9 %	2.6%	−4.1 %	7.9%

We find that the PACIS underrepresents men, the oldest age group (over 60 y.o.), inactive population, people with a basic education level, and non-nationals. Residents in Malaga, a coastal area with a large proportion of secondary residences and holiday rentals, are also less likely to have registered as members in the panel.

#### 4.3. Main and Auxiliary Variables in the PACIS

The data at the population level were obtained from the Spanish National Institute of Statistics selecting 2015 as the reference year. In our case, the recruitment of the panel and the survey were conducted the same year. However, we would like to note that an important aim for computing the weighting adjustments is to correct for the increasing mismatches between the panel and the population associated as the panel ages (panel attrition, changes in the population, etc.). To fulfill this objective in the case of cross-sectional panels, we must take the time of the survey as the reference for selecting the auxiliary data at the population level [41].

Some demographics variables are described in Table 4.

**Table 4.** Education level by sex and age at population and panel levels.

	Primary	Secondary	Higher	Primary	Secondary	Higher
<b>Men</b>						
16–29	71,968	537,814	59,433	35	217	50
30–44	91,673	702,591	194,410	40	310	143
45–59	197,901	586,726	126,083	58	239	77
>60	471,729	231,988	97,686	119	94	52
<b>Women</b>						
16–29	44,609	487,020	112,122	44	211	88
30–44	78,176	619,471	272,774	45	466	210
45–59	194,971	572,750	149,620	121	322	94
>60	681,508	224,651	72,158	204	82	28

Those estimates at population level can be found on the INE website: <http://www.ine.es/>.

Three main variables in this study are analyzed.

The first of them is related to the “interest in politics”. The variable is the answer to the following question: *How interested would you say you are in politics?: Very interested, Quite interested, Somewhat interested, Hardly interested, Not at all interested.*

The second is about the preferences that the interviewees have about the way in which political decisions should be made. The variable is the answer to the question: *How would you like to be made political decisions in Spain, Andalusia, your municipality?* The answer to this question was offered in the form of a average score of a numerical scale from 0 to 10, where 0 means that citizens should make all decisions on their own, and 10 that politicians should make all decisions on their own.

In addition, the last one is related to the concrete procedures of decision-making preferred by the interviewees. The question is: *We would like you to rate the following ways of making political decisions?* The answer to this question was offered in the form of scores in a numerical scale from 0 to 10, where 0 means that the procedure in question does not help anything in decision-making and the 10 that is the best way to make decisions. Six procedures are analyzed: *Elections are the way for citizens to intervene in decisions (Elections), Take policy decisions through consultation to experts (Consult to experts), Organize assemblies and meetings so that people can take decisions by themselves (Assemblies), Allow experts to take important policy decisions (Allow experts), Organize referenda frequently (Referenda), and Let the government take the decisions (Government).*

There are several auxiliary variables that can be used for calibration at the panel level. We have followed the criterion proposed by [25] based on the  $H_3$  bias indicator, to select the variables that, included in the auxiliary vector, were the most effective for bias reduction. The final auxiliary variables selected are described in Table 4 and 5.

**Table 5.** Auxiliary variables at panel level (percentages).

								Laboral Status			
		Working	Retired	Unemployed	Student	Housework					
		37.41	16.12	30.58	7.46	8.42					
								Province			
Almería	Cádiz	Córdoba	Granada	Huelva	Jaén	Málaga	Sevilla				
7.67	14.72	9.67	14.51	8.33	8.33	14.66	22.10				
		PC	Net	Email	Land phone	Cellular phone		% of owners			
		78.05	79.1	57.96	37.12	78.59					

#### 4.4. Results

Tables 6–8 show the estimated proportions (in percent, %) for the three variables by using the two proposed calibration methods. Root mean squared errors (RMSE) are estimated using jackknife techniques and are also shown in Tables 6–8.

By examining the estimators and the RMSE in the above tables, we are led to the following observations:

- Survey results after applying calibration weights are different from unweighted results.
- Differences in survey results after applying one-step or two-step calibration estimators are small in estimates and in errors.
- Calibration estimator in one step CAL1 come close in RMSE to CAL2. The two-step calibration estimators CAL3 and CAL4 estimators also give similar results to each other.
- There is no estimator that always has the lowest RMSE errors.

**Table 6.** Point estimation of proportions (in percent, %) using several methods: unweighted (HT), calibration in one (CAL1) or two stages (CAL2, CAL3 and CAL4). Main variable: “Interest in politics”. Estimated Root Mean Squared Error using jackknife (in percent, %).

	How Interested Would You Say You Are in Politics?									
	Est.					RMSE(%)				
	HT	CAL1	CAL2	CAL3	CAL4	HT	CAL1	CAL2	CAL3	CAL4
<i>Extremely</i>	17.02	14.63	15.05	14.88	14.71	1.53	1.28	1.39	1.28	1.27
<i>Very</i>	33.30	30.04	30.03	31.39	31.05	2.54	2.16	2.18	2.05	2.03
<i>Moderately</i>	27.34	27.02	26.85	27.04	27.07	1.95	1.85	1.84	1.77	1.78
<i>Slightly</i>	14.79	17.50	17.20	16.55	16.84	1.96	1.91	1.76	1.59	1.61
<i>Not at all</i>	7.55	10.82	10.86	10.13	10.33	1.86	1.55	1.50	1.36	1.39

**Table 7.** Point estimation of proportions (in percent, %) using several methods: unweighted (HT), calibration in one (CAL1) or two stages (CAL2, CAL3 and CAL4). Main variable: “Preferences for political decisions”. Estimated Root Mean Squared Error using jackknife.

	How Would You Like to Be Made Political Decisions in ...?					
	Spain		Andalusia		Municipality	
	Est.	RMSE	Est.	RMSE	Est.	RMSE
HT	4.59	0.19	4.35	0.19	4.28	0.19
CAL1	4.53	0.17	4.41	0.17	4.27	0.17
CAL2	4.53	0.17	4.45	0.17	4.30	0.17
CAL3	4.54	0.15	4.45	0.15	4.32	0.16
CAL4	4.55	0.15	4.46	0.15	4.32	0.16

**Table 8.** Point estimation of proportions (in percent, %) using several methods: Calibration in one (CAL1) or two stages (CAL2, CAL3 and CAL4). Main variable: “Procedures of decision-making”. Estimated Root Mean Squared Error using jackknife

	We Would Like You to Rate the Following Ways of Making Political Decisions in Andalusia?						
		Elections	Consult	Assemblies	Allow	Referendums	Political
		to Experts	to Experts		Experts		Leaders
HT	<i>Est.</i>	7.34	7.10	6.40	6.44	5.88	4.30
	<i>RMSE</i>	0.30	0.29	0.25	0.29	0.23	0.19
CAL1	<i>Est.</i>	7.33	7.03	6.70	6.51	5.90	4.28
	<i>RMSE</i>	0.24	0.22	0.24	0.22	0.22	0.18
CAL2	<i>Est.</i>	7.27	7.05	6.70	6.57	5.87	4.29
	<i>RMSE</i>	0.23	0.22	0.23	0.23	0.21	0.18
CAL3	<i>Est.</i>	7.27	7.07	6.63	6.50	5.84	4.25
	<i>RMSE</i>	0.21	0.20	0.20	0.19	0.19	0.16
CAL4	<i>Est.</i>	7.27	7.07	6.64	6.51	5.84	4.26
	<i>RMSE</i>	0.21	0.20	0.20	0.20	0.19	0.16

We would like to note that the results obtained by applying these methods to the survey indicate that 45% of Andalusians are interested in politics, a figure that is somewhat higher than that of the Spanish Citizens (40% according to 2016 European Social Survey data) but still far from the 47% of EU citizens that are interested in politics [42]. Regarding their preferences on who should make the political decisions, most Andalusians choose central positions, with a few more preferring the

participatory to the representative pole and that for the three governance levels analyzed (state, region, and municipality). This preference is coherent with the way Andalusian people evaluate different decision-making procedures, with the elections obtaining the highest score (7.27) just before consultation with experts (7.05) and citizen assemblies (6.7). The same pattern is observed in a nationally representative survey conducted in Spain in 2011 (Centro de Investigaciones Sociológicas, CIS, Study number 2860) [43].

## 5. Discussion

In this work, we examined probability-based panel survey settings where there are missing values.

We use raking calibration weighting to remove selection biases resulting from unit non-response. Although this technique was developed to reduce standard errors, many authors have shown that calibration can also be used to handle non response [22–24].... In Section 2, we consider four methods to use the different auxiliary information in the calibration although other alternatives are possible. In each method, a set of calibrated weights is computed by modifying the design weights through the available auxiliary information. This single set of weights will then be used to calculate the calibrated estimates of all the population parameters of interest since the values of the weights do not depend on  $y$  variables. This is a great advantage of the calibration technique in multipurpose surveys.

An important problem is the selection of auxiliary vector for the calibration. The authors in [44,45] point out that the selection of the variables is more important than the method used to calibrate. The role of the explanatory variables has been demonstrated to be crucial at the post-adjustment stage. The advantage of having a probability-based panel is that more auxiliary information is available for the calibration. Many different vectors can be composed. The authors in [25] provide computational tools for obtaining a preference ordering of potential  $x$ -vectors, with the objective to reduce as much as possible the bias remaining in the calibration estimator. The selection of variables for the calibration is a very interesting problem but quite complex to deal with in this situation, which is different for each method used. This is a problem to deal with in future research.

We conduct a simulation study in which it is observed that the estimators CAL1 and CAL2 have a similar behavior to each other as well as CAL3 and CAL4. The simulation study also shows that calibration can significantly reduce bias for MAR mechanisms while it is not very effective for NMAR mechanisms.

We use data from one survey conducted using the Citizen Panel for Social Research to test the behavior of proposed estimators. Although the same auxiliary information goes into the final weights, the resulting calibration estimators are not identical. In the same way, the differences for the estimators of the variances can sometimes be significant depending on the relationship between the variables.

## 6. Conclusions

The main conclusion of this work is that weighting by calibration is an effective technique for reducing non-response bias in probability-based panels when data are missing at random (MAR). Regarding the calibration method, in one or two steps, there are not theoretical reasons to prefer one over the other. However, the two-steps method has some practical advantages. Knowing the weights of each unit in the sampling frame allows for incorporating this information in the selection of the subsequent cross-sectional samples and maximizing the number of auxiliary variables when computing the weights for each specific survey. Another benefit of this approach is that it allows for isolating the sources of error produced at each step (i.e., coverage error in the recruitment of the panel and attrition bias in the subsequent surveys). It must be taken into account that panel weights have to be updated taking the time of the survey as the reference for selecting the auxiliary data at the population level when using this approach to calibrate cross-sectional panels. In this case, the calibration in one step may be more straightforward, particularly if auxiliary variables at the survey level do not add much value to the quality of the estimates.

In all, the choice of the calibration method is a complex decision where different aspects come into play: the technical features of the panel and the survey, the mechanism of non-response affecting our data, and the availability of auxiliary information.

**Author Contributions:** The authors contributed equally to this work in conceptualization, methodology, software, and original draft preparation. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work was supported by the Ministerio de Economía, Industria y Competitividad, Spain, under Grant MTM2015-63609-R.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. ESOMAR; Global Market Research: Amsterdam, The Netherlands, 2016.
2. Peytchev, A.; Carley-Baxter, L.R.; Black, M.C. Multiple Sources of Nonobservation Error in Telephone Surveys: Coverage and Nonresponse. *Sociol. Methods Res.* **2011**, *40*, 138–168. [CrossRef]
3. Eckman, S.; Kreuter, F. Undercoverage Rates and Undercoverage Bias in Traditional Housing Unit Listing. *Sociol. Methods Res.* **2013**, *42*, 264–293. [CrossRef]
4. Gummer, T. Assessing Trends and Decomposing Change in Nonresponse Bias: The Case of Bias in Cohort Distributions. *Sociol. Methods Res.* **2017**. [CrossRef]
5. De Leeuw, E. Counting and Measuring Online: The Quality of Internet Surveys. *Bull. Sociol. Methodol. Methodol. Sociol.* **2012**, *114*, 68–78. [CrossRef]
6. Dixon, J.; Tucker, C. Survey Nonresponse. In *Handbook of Survey Research*; Marsden, P., Wright, J., Eds.; Emerald Group Publishing: Bingley, UK, 2010; pp. 593–630.
7. Tucker, C.; Lepkowski, J. Telephone Survey Methods: Adapting to Change. In *Advances in Telephone Survey Methodology*; Lepkowski, J., Tucker, C., Brick, J., de Leeuw, E., Japec, L., Lavrakas, P., Sangster, R., Eds.; Wiley-Interscience: Hoboken, NJ, USA, 2008; pp. 3–26.
8. Vehovar, V.; Berzelak, N.; Lozar Manfreda, K. Mobile Phones in an Environment of Competing Survey Modes: Applying Metric for Evaluation of Costs and Errors. *Soc. Sci. Comput. Rev.* **2010**, *28*, 303–318. [CrossRef]
9. Lu, Y.; Chen, Z.; Law, R. Mapping the progress of social media research in hospitality and tourism management from 2004 to 2014. *J. Travel Tour. Mark.* **2018**, *35*, 102–118. [CrossRef]
10. Manca, M.; Boratto, L.; Roman, V.M.; i Gallissà, O.M.; Kaltenbrunner, A. Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Soc. Netw. Media* **2017**, *1*, 56–69. [CrossRef]
11. Sinnenberg, L.; Buttenheim, A.M.; Padrez, K.; Mancheno, C.; Ungar, L.; Merchant, R.M. Twitter as a tool for health research: A systematic review. *Am. J. Public Health* **2017**, *107*, e1–e8. [CrossRef]
12. Li, R.; Crowe, J.; Leifer, D.; Zou, L.; Schoof, J. Beyond big data: Social media challenges and opportunities for understanding social perception of energy. *Energy Res. Soc. Sci.* **2019**, *56*, 101217. [CrossRef]
13. Japec, L.; Kreuter, F.; Berg, M.; Biemer, P.; Decker, P.; Lampe, C.; Usher, A. Big data in survey research: AAPOR task force report. *Public Opin. Q.* **2015**, *79*, 839–880. [CrossRef]
14. Callegaro, M.; Manfreda, K.L.; Vehovar, V. *Web Survey Methodology*; SAGE: Thousand Oaks, CA, USA, 2015.
15. Fricker, R.D.; Schonlau, M. Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods* **2002**, *14*, 347–367. [CrossRef]
16. Tsetsi, E.; Rains, S.A. Smartphone Internet access and use: Extending the digital divide and usage gap. *Mob. Media Commun.* **2017**, *5*, 239–255. [CrossRef]
17. Lee, M.H. Statistical Methods for Reducing Bias in Web Surveys. Ph.D. Thesis, Simon Fraser University, Burnaby, BC, Canada 2011. Available online: <http://summit.sfu.ca/item/11783> (accessed on 15 November 2019).
18. DiSogra, C.; Callegaro, M. Metrics and Design Tool for Building and Evaluating Probability-Based Online Panels. *Soc. Sci. Comput. Rev.* **2016**, *34*, 26–40. [CrossRef]
19. Callegaro, M.; Baker, R.; Bethlehem, J.; Göritz, A.S.; Krosnick, J.A.; Lavrakas, P.J. Online panel research: History, concepts, applications and a look at the future, Chapter 1. In *Online Panel Research: A Data Quality Perspective*; Callegaro, M., Baker, R., Bethlehem, J., Göritz, A.S., Krosnick, J.A., Lavrakas, P.J., Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2014; pp. 1–22.

20. Blom, A.G.; Bosnjak, M.; Cornilleau, A.; Cousteaux, A.-S.; Das, M.; Douhou, S.; Krieger, U. A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe. *Soc. Sci. Comput. Rev.* **2016**, *34*, 8–25. [[CrossRef](#)]
21. Cheng, A.; Zamorro, G.; Orriens, B. Personality as a predictor of unit non-response in an internet panel. *Sociol. Methods Res.* **2018**. [[CrossRef](#)]
22. Andersson, P.; Särndal, C.E. Calibration for non-response treatment: In one or two stepsf. *Stat. J. IAOS* **2014**, *32*, 375–381. [[CrossRef](#)]
23. Särndal, C.E.; Lundström, S. *Estimation in Surveys with Nonresponse*; John Wiley and Sons: New York, NY, USA, 2005.
24. Särndal, C.E.; Lundström, S. Assessing auxiliary vectors for control of non-response bias in the calibration estimator. *J. Off. Stat.* **2008**, *4*, 251–260.
25. Särndal, C.E.; Lundström, S. Design for estimation: Identifying auxiliary vectors to reduce non-response bias. *Surv. Methodol.* **2010**, *36*, 131–144.
26. Kott, P.S. Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors. *Surv. Methodol.* **2006**, *32*, 133–142.
27. Kott, P.S.; Liao, D. Providing double protection for unit non-response with a nonlinear calibration-weighting routine. *Surv. Res. Methods* **2012**, *6*, 105–111.
28. Kott, P.S.; Liao, D. One step or two? Calibration weighting form a complete list frame with non-response. *Surv. Methodolgy* **2015**, *41*, 165–181.
29. Bethlehem, J.; Callegaro, M. Introduction to Part IV. In *Online Panel Research: A Data Quality Perspective*; Callegaro, M., Baker, R., Bethlehem, J., Göritz, A.S., Krosnick J.A., Lavrakas, P.J. Eds.; John Wiley & Sons, Ltd: Chichester, UK, 2014; pp. 263–272.
30. Särndal, C.E. The calibration approach in survey theory and practice. *Surv. Methol.* **2007**, *33*, 99–119.
31. Arcos, A.; Rueda, M.; Trujillo, M.; Molina, D. Review of estimation methods for landline and cell phone surveys. *Sociol. Methods Res.* **2015**, *44*, 458–485. [[CrossRef](#)]
32. Arcos, A.; Contreras, J.M.; Rueda, M. A novel calibration estimator in social surveys. *Sociol. Methods Res.* **2014**, *43*, 465–489. [[CrossRef](#)]
33. Deville, J.C.; Särndal, C.E. Calibration estimators in survey sampling. *J. Am. Stat. Assoc.* **1992**, *87*, 376–382.
34. Devaud, D.; Tillé, Y. Deville and Särndal’s calibration: Revisiting a 25-years-old successful optimization problem. *TEST* **2019**. [[CrossRef](#)]
35. Estevao, V.M.; Särndal, C.E. A functional form approach to calibration. *J. Off. Stat.* **2000**, *16*, 379–399.
36. Särndal, C.E. Methods for estimating the precision of survey estimates when imputation has been used. *Surv. Methodol.* **1992**, *18*, 241–252.
37. Estevao, V.M.; Särndal, C.E. Survey Estimates by Calibration on Complex Auxiliary Information. *Internat. Statist. Rev.* **2006**, *74*, 127–147. [[CrossRef](#)]
38. Alfons, A.; Templ, M.; Filzmoser, P. An Object-Oriented Framework for Statistical Simulation: The R Package *simFrame*. *J. Stat. Softw.* **2010**, *37*, 1–36. [[CrossRef](#)]
39. Templ, M.; Meindl, B.; Kowarik, A.; Dupriez, O. Simulation of Synthetic Complex Data: The R Package *simPop*. *J. Stat. Softw.* **2017**, *79*, 1–38. [[CrossRef](#)]
40. Alarcón, P.; Font, J.; Fernéz, J. Tell me what you trust and what you think about political actors and I will tell you what democracy you prefer. In *XIII AECPA Conference. “La fortaleza de Europa: Vallas y puentes”*; Universidad de Santiago de Compostela: Santiago de Compostela, Galicia, 2017.
41. Verma, V.; Betti, G.; Ghellini, G. Cross-sectional and longitudinal weighting in a rotational household panel: Applications to EU-SILC. *Stat. Transit.* **2007**, *8*, 5–50.
42. *ESS Round 8: European Social Survey Round 8 Data*; Data file edition 2.1; Data- Archive and distributor of ESS data for ESS ERIC; Norwegian Centre for Research Data for ESS ERIC: Bergen, Norway, 2016.
43. Font, J.; Wojcieszak, M.; Navarro, C.J. Participation, Representation and Expertise: Citizen Preferences for Political Decision-Making Processes. *Political Stud.* **2015**, *63*, 153–172. [[CrossRef](#)]

44. Rizzo, L.; Kalton, G.; Brick, J.M. A comparison of some weighting adjustment methods for panel non-response. *Surv. Methodol.* **1996**, *22*, 43–53.
45. Mercer, A.; Lau, A.; Kennedy, C. For Weighting Online Opt-In Samples, What Matters Most? *Pew Res. Center.* **2018**, *2018*, 28.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).