

# Comparison of Survey Data Collection Methods

*Comparación de métodos de campo en la encuesta*

Vidal Díaz de Rada and Juan Antonio Domínguez Álvarez

## Key words

- Mail Surveys
- Telephone Surveys
  - Internet
  - Research
- Methodology
- Survey Responses
  - Fieldwork

## Palabras clave

- Encuestas por correo
- Encuestas por teléfono
  - Internet
  - Metodología de la investigación
  - Respuestas de encuestas
  - Trabajo de campo

## Abstract

This paper presents the results of a mixed-mode survey of the general population using a combination of postal, telephone, and Internet survey protocols. Potential respondents were invited to participate by ordinary mail and were allowed to choose their preferred response mode. The paper focuses on response quality (taking non-responses into consideration), fieldwork time and data collection cost. The results reveal that the Internet survey produces the lowest rate of non-responses and requires significantly less fieldwork time, although it is slightly more costly than the postal survey. However, when differences in cost structure are taken into account, we find that the same number of completed questionnaires could have been obtained through the Internet alone at a cost that is 18.2% lower than the mixed-mode survey.

## Resumen

En este trabajo se presentan los resultados de una investigación a población general que utiliza conjuntamente una encuesta por correo postal, una encuesta telefónica y una encuesta a través de Internet. La muestra objeto de estudio recibe la *invitación* a cooperar mediante una comunicación por correo, y es el propio seleccionado el que decide cómo responder. La nota analiza la calidad de la respuesta (considerando el número de no respuestas), el tiempo necesario para llevar a cabo el trabajo de campo y el coste de la recogida de información. Los resultados desvelan que la realización de encuestas vía web reducen el tiempo del trabajo de campo, consiguen menor respuesta parcial y presentan un coste ligeramente superior a la encuesta realizada por correo. Ahora bien, la diferente estructura de costes de cada una permite concluir que de haber obtenido el total de los cuestionarios *online*, la reducción de costes hubiera supuesto un 18,2%.

## Citation

Díaz de Rada, Vidal and Domínguez Álvarez, Juan Antonio (2017). "Comparison of Survey Data Collection Methods". *Revista Española de Investigaciones Sociológicas*, 158: 137-148. (<http://dx.doi.org/10.5477/cis/reis.158.137>)

**Vidal Díaz de Rada:** Universidad Pública de Navarra | [vidal@unavarra.es](mailto:vidal@unavarra.es)

**Juan Antonio Domínguez Álvarez:** IESA-CSIC | [jadominguez@iesa.csic.es](mailto:jadominguez@iesa.csic.es)

## INTRODUCTION<sup>1</sup>

A retrospective view of the evolution of data collection in surveys reveals that Computer Assisted Interviewing (CAI) is gradually replacing the traditional in-person survey. These questionnaires, in addition to using all types of available questions (single response, multiple response, textual or numeric response, etc.) are easier to administer given that they are conducted automatically—“jumps” caused by filter questions, complicated routes, etc. And of equal importance, they allow for the establishment of *consistency rules* which are verified prior to recording the information (Couper and Bosnjak, 2010).

The purpose of this work is to consider to what point these characteristics, highlighted by most of the literature, apply to a group of Andalusian residents living abroad. Unlike other studies in which the researcher selects the type of questionnaire to be used, in this study, the interviewee decides how to respond, thereby implying a preference for one type or another.

A sample Andalusian population living abroad was contacted via ordinary mail and were presented with three potential response forms: paper questionnaire, questionnaire available on a website or telephone survey. This work focuses on the study of the differences between questionnaires, based on the number of questions that were not answered, the response speed and the cost of the fieldwork. It includes a general population and the use of the same questionnaire in three distinct formats, allowing for the comparison of up to what point the advantages—evidenced in studies carried out with specific samples—may be generalized to a general population sample.

<sup>1</sup> A previous version was presented at the *XII Congreso Español de Sociología* (12th Spanish Congress on Sociology). The authors would like to thank the Editorial Board of the journal and the three anonymous reviewers.

## SOME BRIEF THEORETICAL CONSIDERATIONS REGARDING CONTACT METHOD

Using mail as a contact and response method requires proper justification at a time when Internet-based *communications* are reducing traditional postal communications to a minimum. Various explanations may justify this course of action:

1. Lower *collaboration rate* for the Internet survey as compared to the mailed survey, which has been verified in numerous studies (including those by Lozar Manfreda *et al.*, 2008; Shih and Fan, 2008; Messer and Dillman, 2011; Tourangeau *et al.*, 2013, Callegaro *et al.*, 2015).
2. Greater *access difficulties* to the interviewee through the Internet, causing the majority of the specific literature to recommend making an initial contact via a distinct method; for example, web surveys *supported* by traditional mail surveys (Hansen *et al.*, 2012; Messer and Dillman, 2011; Messer *et al.*, 2012).

Smyth *et al.* (2010) revealed that sending a survey via mail one day after contact over the Internet increases cooperation by 33%, achieving a response rate of 55%. Furthermore, individuals having distinct sociodemographic features are included in the sample, notably improving its representativeness. Along these lines, Messer and Dillman (2011) warned that surveys via regular mail are more representative of the sociodemographic features of the population than mixed surveys.

3. As a *third* argument, there is the considerable transcendence of the *tool* used to establish the first contact with the selected party. Crawford *et al.* (2004), for example, consider that the contact via ordinary mail attains a response rate that is significantly higher than that received when using email (response rates of 52.5

and 44.9% respectively), while Harmon *et al.* (2005) compared email with an attached letter from the research sponsor, email with an attached letter from the company to carry out the field work, and letter via regular mail from the research sponsor, achieving response rates of 64.4, 63.6 and 69.9% respectively.

Bosnjak (2008) revealed that the sending of SMS messages is more effective than notifying via regular mail, with rates of 84 and 71%. Other influential aspects include the length of the communication (long invitations are more effective than short ones) and the tone (formal/informal) that is used in the same (Tourangeau *et al.*, 2013).

## AN APPLICATION TO THE STUDY “LIVING CONDITIONS AND RETURN EXPECTATIONS OF ANDALUSIAN EMIGRANTS”

In 2009, the Regional Council of the Junta of Andalusia commissioned a study with the objective of determining the situation of Andalusians living abroad, using a questionnaire with 57 questions (136 variables) and seven topics: a) sociodemographic features; b) coexistence and housing conditions; c) concerns, work and income; d) use of assistance institutions; e) assessment of diverse institutions and life satisfaction; f) closeness/distance from Andalusia; and g) social situation in Andalusia and reasons for emigrating (Moscoso *et al.*, 2010).

The majority of the questions deal with *actions* and situations, although there are also twelve questions regarding *subjective aspects*, based on the classification by García Ferrando and Llopis (2015). The closed questions have a quite varied structure:

- Single response questions that use “option buttons” in the web questionnaire, with predominantly five and two response

categories. This is the most frequent type of question, distinguishing between nominal questions (17 questions and other variables) –that mainly consider acts– and 16 ordinal questions, of which nine consider assessments and opinions (subjective aspects)<sup>2</sup>.

- Multiple choice questions, all regarding behaviors and situations, having two formats: questions with a list of responses that include information on situations of coexistence, type of medical-healthcare protection, payment for medications, individuals with whom they emigrated, and family members who remained in Andalusia.

Responses with a dichotic format in each item were also used. They considered household equipping, changes in habits due to economic issues, demand for services by institutions and disabilities.

There were 5 and 4 questions for these, respectively, offering 30 and 31 variables<sup>3</sup>.

- Battery of questions; three questions with 6, 6 and 8 items respectively, 20 variables, responded to on a scale from 1 to 10. These are opinion-based questions that include the level of confidence in public and private institutions, as well as the satisfaction with a series of aspects of their life.

<sup>2</sup> Detail of the number of questions according to the number of response categories:

|                           | Nominal | Ordinal | Total |
|---------------------------|---------|---------|-------|
| Dichotomous:              | 9       |         | 9     |
| Three response categories | 2       | 5       | 7     |
| Four categories           | 1       |         | 1     |
| Five categories           | 4       | 8       | 12    |
| Six and more categories   | 1       | 3       | 4     |
| Total                     | 17      | 16      | 33    |

<sup>3</sup> The differences between both and between these and the battery of questions are presented on page 15 of the text by Azofra (1999).

The questionnaire also includes twelve open-ended questions regarding behaviors. Six require numeric responses<sup>4</sup>, one requires a very short response (city of birth), and another five are answered with various words or a sentence<sup>5</sup>. Thus, the simpler and more complex open-ended question types have been combined (Dillman, Smyth and Christian, 2014).

The questionnaire –in its Internet version<sup>6</sup>– was completed in an average of 20.7 minutes.

The universal object of study is the Andalusian population residing abroad, calculated at 182,928 personas (Moscoso, 2010). Using the *Censo Electoral de Residentes en el Extranjero* (Electoral Census of Residents Living Abroad) as the sampling framework, a sample of 2,493 individuals was used. The sample, stratified by geographic areas of residence (Europe, Maghreb, Latin America, North America and Oceania) was contacted by postal mail and had three potential response channels in order to respond to the questionnaire: questionnaire via normal mail (with postage paid at destination), electronic questionnaire (clicking on an internet link), and CATI telephone interview (paid for by the organism conducting the study).

The implementation process was similar to that used in other such studies (Dillman, Smyth and Christian, 2014), although monetary incentives were not used, nor change in contact type. Specifically, the *first wave* received a questionnaire via mail along with

a letter or presentation detailing the purposes of the study, encouraging participation and explaining the distinct response alternatives. It was signed by the Deputy Minister of the Junta of Andalusia and throughout the entire questionnaire, the governmental logo-type appears. Those who did not respond were sent a new letter along with the questionnaire 20 weeks later, reminding them that they had yet to respond. Upon failing to attain a sufficient sample size, a second delivery was made two months later to a new sample, who received no reminder (*second wave*).

The objective of this article is to analyze how the modality produces differences between the questionnaires in number of responses, speed of response and cost. As a preliminary hypothesis, it was proposed that the internet questionnaire would obtain a lower number of non-responses, would permit a reduction in fieldwork and would decrease costs.

## RESULTS

The 15,657 letters sent received 2,493 responses, suggesting a cooperation rate of 18.9% (AAPOR, 2015). Participation was much higher via regular mail, with 2,083 questionnaires collected as opposed to the 359 received via the Internet and the 51 obtained via CATI telephone surveys. These latter have not been considered in this work given their limited sample size. As we can see in Table 1, the sample responding via Internet is younger and has a higher education level.

These results may not be attributed to a lack of infrastructure or skills in Internet use, given that two of the questions appearing in the questionnaire reveal that 60% of the survey participants have an Internet connection in their home. Furthermore, they are regular users, with 36% connecting to the Internet on a daily basis and 18% connecting on a

<sup>4</sup> Number of individuals in the household, number of rooms in the home, incomes, number of times travelling to Andalusia, year of emigration and year when began to live in this country.

<sup>5</sup> Most important occupations, most important problems/needs, reasons why they remain in this country, reasons why they emigrated and reasons for choosing this country. Each question may be responded to with up to three responses.

<sup>6</sup> The Internet version, as it precisely includes the time of starting and completing the survey, allows for knowledge of the time necessary to complete questionnaire.

weekly basis, implying that over half of those responding by mail use the Internet regularly. On the other hand, it is surprising to note that 8% of those responding to the survey via Internet do not have an Internet connection in their home. However, both the response rate as well as the sample characterization is similar to the results from similar studies (such as Dillman, Smyth and Christian, 2014).

### Response quality

The mean number of variables that were not responded by each interviewee was 11.2, a relatively low rate when taking into account that the questionnaire contained information from 137 variables. An initial explanation of this number of no responses is related to the type of questions used, expecting fewer responses from open-ended questions and, within these, from the “text type” open ended questions. In fact, these latter constitute over half of the no responses (5.5 no response variables), being significantly greater in the mailed survey (5.6) than in the Internet survey (4.4).

Fewer differences are found in the open numeric responses with a mean rate of 0.42

for no responses as compared to 0.47 for the mailed survey and 0.17 for the Internet survey.

More surprising is the high number of no responses in the closed responses, having a mean of 5.3 and making up 47% of the total no responses (5.3/11.2). The analysis by modalities of information collection reveals a low rate in the Internet survey at 2.3 as compared to the high rate of 5.9 found in the survey by mail. When analyzing the difference between questions on actions and questions on opinions, the former have a greater response, although the difference between the two is not significant.

Many researchers have found a relationship to exist between response quality and the time of responding (including Olson, 2013; Kreuter, Müller and Trappmann, 2013). They suggest that the interviewees who respond immediately after being contacted respond better than those who do so following various invitations; something that was also found in our study. A comparison between those who respond immediately and those who respond after an initial reminder (columns “without reminder” and “reminder” in

**TABLE 1.** *Sample features according to selected modality (vertical responses)*

|                   | Mail  | Internet |                    | Mail  | Internet |
|-------------------|-------|----------|--------------------|-------|----------|
| Chi-square 0.701  |       |          | Chi-square 129.9** |       |          |
| Men               | 49    | 51       | Sin estudios       | 21    | 5        |
| Women             | 51    | 49       | Primary studies    | 20    | 10       |
| N° cases          | 2,083 | 359      | Secondary studies  | 31    | 31       |
|                   |       |          | Higher studies     | 27    | 53       |
|                   |       |          | N° cases           | 2,083 | 359      |
| Chi-square 94.9** |       |          |                    |       |          |
| 19-29             | 8     | 15       |                    |       |          |
| 30-44             | 18    | 32       |                    |       |          |
| 45-59             | 25    | 30       |                    |       |          |
| 60 years and over | 48    | 23       |                    |       |          |
| N° cases          | 2,083 | 359      |                    |       |          |

Table 2) reveals an increase of over 2 no response variables, a difference which rises to three in the case of the mailed survey. In percentage terms, those who responded immediately had a mean of 8.8% of no response variables, which increased to 10.3% for those receiving the reminder in which they were informed that they had yet to respond to the questionnaire.

The *second wave*, sent to a new sample a few weeks after the reminder, reveals the stability of the findings: 11.1 variables without response, meaning some 8.8% of the questionnaire (right hand column of Table 2).

Up until now, we have focused on the number of no responses, but we have not considered that there may have been interviewees who did not provide any responses as well as others who almost completely responded to the questionnaire. This “new” line of argument considers the localized interviewees and examines the number of interviewees who did not provide responses in each of the administration modalities. Of those

interviewed via the Internet, 8% responded to ALL of the interview questions and somewhat more than half (52.6%) left less than 5% of the questions without responses; these percentages were reduced to 3 and 30% respectively in the mailed survey (Table 3). When collectively analyzing the figures, we find that 60% of those responding via the Internet responded to almost all of the survey (9.5 questions out of 10); as compared to 33% who did so in the mailed survey.

Considering the differences in administration modalities, without considering other variables, may lead to incorrect impressions, given that an in-depth analysis reveals that the mailed survey was the modality selected mainly by individuals of a higher age and lower education level, while just the opposite occurred with the Internet survey, which were responded to mainly by younger participants having a high education level (Table 1).

To analyze how sociodemographic features and selected survey modality are of influence, a multiple regression (least squares)

**TABLE 2.** Number of no responses according to modality and moment of response of the questionnaire

|                        | Average        |          |       |             |       |
|------------------------|----------------|----------|-------|-------------|-------|
|                        | No reminder    | Reminder | Total | Second wave | Total |
| Mail                   | 11.2           | 14.1     | 12.0  | 11.9        | 12.0  |
| Internet               | 6.7            | 7.3      | 6.9   | 6.8         | 6.9   |
| Total (mean variables) | 10.6           | 12.9     | 11.2  | 11.2        | 11.2  |
|                        | In percentages |          |       |             |       |
|                        | No reminder    | Reminder | Total | Second wave | Total |
| Mail                   | 9.2%           | 11.2%    | 9.8%  | 9.5%        | 9.7%  |
| Internet               | 5.8%           | 6.1%     | 5.9%  | 5.7%        | 5.9%  |
| Total (% variables)    | 8.8%           | 10.3%    | 9.0%  | 8.8%        | 9.1%  |
| N° cases               | 1,562          | 595      | 2,198 | 285         | 2,442 |

**TABLE 3.** Response quality, considering the % of interviewees who left questions without responses (vertical responses)

| No response questions            | Number of no responses |          |
|----------------------------------|------------------------|----------|
|                                  | Mail                   | Internet |
| None                             | 3.0                    | 8.1      |
| Less than 5% of the questions    | 30.0                   | 52.6     |
| Between 5.1 and 10%              | 34.9                   | 31.5     |
| Between 11 and 20%               | 22.0                   | 4.2      |
| Between 21 and 30%               | 5.8                    | 0.6      |
| Between 31 and el 50%            | 3.4                    | 1.4      |
| More than half the questionnaire | 0.9                    | 1.7      |
| Sample size:                     | 2,083                  | 359      |

was carried out. The number of no responses was selected as the dependent variable and –in the first model– response modality and –later– sociodemographic features of the interviewees was the independent variable. Their influence was measured using the same methodology as in other studies (such as Messer and Dillman, 2011; Messer *et al.*, 2012). The results, shown in Table 4, reveal the great influence of questionnaire modality on the number of no response variables (model 1).

While the first model introduces sociodemographic features and area of residence (model 2), modality continues to have a significant influence on the number of no responses, even when controlling for the influence of sociodemographic variables. Age is the most influential term, followed by the selected modality, education level and gender. The sign of the coefficients indicates that as age increases, the number of no responses also rises, just as the Internet survey generates a decrease in the number of no responses. Interviewees with a lower education level and females have a lower number of responses completed on the questionnaire. The explanatory power of the model ( $R^2$  0.153), slightly higher than that attained in similar studies

(including Messer *et al.*, 2012), indicates the low explanatory power of the sociodemographic variables on number of no response questions.

In a third analysis, four additional interactions are added: gender-studies, Internet use-age, Internet-studies and Internet-gender; making no contribution to the model (therefore, they are not included here).

From this section we can make three conclusions; on the one hand, that the paper survey offers worse results, that half of the “no responses” were made in closed questions, and finally, that only the fully completed questionnaires were considered (with 90-95% completed). Thus, the sample size of the telephone and mail surveys was cut in half.

### Response speed

The length of information collection has been, without a doubt, another of the great inconveniences of the mail surveys using paper questionnaires. In fact, in this study, questionnaires were received by mail for a period of over three months (103 days), to which it is necessary to add another 43 days after the sending of the reminder, for a total of 146 days until

**TABLE 4.** Regression on the factors influencing the number of no responses according to questionnaire modality, sociodemographic features and area of residence

|   | Dependent variable: Number of no responses |            |                           |            |
|---|--|------------|---------------------------|------------|
|   | MODEL 1                                    |            | MODEL 2                   |            |
|   | Standardized coefficients                  | t. Student | Standardized coefficients | t. Student |
| Questionnaire modality<br>(Reference: mail) | -0.219                                     | -10.914**  | -0.132                    | -6.675**   |
| Gender (Reference: female)                  |  |            | -0.064                    | -3.360**   |
| Age (continues)                             |  |            | 0.273                     | 11.742**   |
| Educational level (Ordinal <sup>a</sup> )   |  |            | -0.087                    | -3,745**   |
| Geographical area of residence              |  |            |                           |            |
| Europe (Ref.: other areas)                  |  |            | 0.009                     | -0.222     |
| Latin America (Ref.: other areas)           |  |            | -0.085                    | -2.117*    |
| North America (Ref.: other areas)           |  |            | -0.036                    | -1.504     |
| Others (Ref.: other areas)                  |  |            | 0.008                     | 0.382      |
| R2  | 0.048                                      |            | 0.153                     |            |
| Number of cases                             | 2,420                                      |            | 2,420                     |            |

Codification of the variables:

(\*) Significant at 0.05.

(\*\*) Significant at 0.01.

<sup>a</sup> Does not know how to read or write (1), incomplete (2), primary (3), secondary (4) and university (5) studies. Continues here and in the geographic area, the codification model proposed by Cea D'Ancona (2002).

having the information from the first wave. However, the Internet survey required 78 days after its first sending (25 days less), and another 35 after the reminder (8 less), thus a duration of 113 days, 33 days less than those required for the mail survey (Table 5).

In the *second wave*, reminders were not used and the mailed surveys required 48

days as compared to the 28 days required for the Internet survey. When collectively considering all of the time necessary for the field-work, the 194 days for the mailed survey are reduced to 141 for the Internet survey, meaning that the results are available 53 days earlier. This figure implies a reduction of 27% in time used.

**TABLE 5.** Response time (in days)

|                  | Difference (in days) from time of sending and date of recording |          |                       |             |                      |
|------------------|---|----------|-----------------------|-------------|----------------------|
|                  | First wave  |          | Total<br>(total days) | Second wave | Total<br>(Two waves) |
|                  | No reminder   | Reminder |                       |             |                      |
| Mail             | 103   | 43       | 146                   | 48          | 194                  |
| Internet         | 78  | 35       | 113                   | 28          | 141                  |
| Total (n° cases) | 1,562   | 594      | 2,197                 | 285         | 2,492                |



**TABLE 6.** *Fieldwork costs*

| Area                   | INTERNET     |                 | Area               | MAIL              |                 |
|------------------------|--------------|-----------------|--------------------|-------------------|-----------------|
|                        | Time (hours) | Cost (15€/hour) |                    | Nº questionnaires | Cost (15€/hour) |
| Programming            | 50           | 750             | Postage            |                   |                 |
| Server hosting         |              | 150             | Europe (1.33€)     | 1,151             | 1,530.83        |
| Fieldwork              | 40           | 600             | Others (1.8€)      | 932               | 1,677.60        |
| Total (Euros)          |              | 1,500           | Recording (1.75*€) | 2,083             | 3,645.25        |
| Cost per questionnaire |              | 4.18            |                    |                   | 6,853.68        |
|                        |              |                 |                    |                   | 3.29            |

\* Note: each questionnaire requires a mean recording time of 7 minutes; 7.57 questionnaires per hour. Based on a salary of 15 €/hour, 1.75 € per questionnaire.

### Research costs

The third aspect under study, referring to the cost of the research, shall only consider the amount of receipt of the valid questionnaires, excluding the costs of printing, handling and sending, which did not vary between modalities. Data analysis was not considered given that it did not vary based on the modality used.

Focusing our attention on the self-administered modality, the different cost structures are considered. The majority of the cost of the Internet survey is fixed, implying that the unit cost decreases as the number of questionnaires received increases. The postal survey, however, has a variable cost, which increases according to the number of questionnaires.

In the Internet questionnaire, we must consider the hours used in the programming and questionnaire test (50 hours), hosting costs for the site where the questionnaire is installed (150 euros annually), and hours of supervision and incident resolution (approximately 40 hours). Costs for the mailed questionnaire are limited to the receipt and recording of the information. The postage-paid stamps (40 grams) have a unit cost of 1.33 € if coming from Europe and 1.80 € when coming from other countries. In the recording of

the questionnaire, a mean of 7 minutes per questionnaire was required.

Considering the human costs of human resources of 15€/hour, Internet fieldwork has costs of 1,500€ as compared to 6,853.70€ for postal surveys. Thus, the 359 web questionnaires have a unit cost of 4.18 euros; whereas the 1,151 questionnaires by mail coming from Europe and the 932 coming from the rest of the world, implying a unit cost of 3.29€ which includes the 2437 hours used for its recording (Table 6).

Considering that the cost structure of the Internet survey is fixed, regardless of the number of responses, attaining another 97 Internet questionnaires (up to 456) would have equaled the costs between modalities, decreasing the cost per questionnaire (Internet) as of this moment. In fact, if the Internet survey option had been the main modality and all of the self-administered questionnaires had been attained through this means (2,493 questionnaires), a unit cost of 0.60<sup>8</sup> euros would have been reached, 18.2% of the unit cost of the postal questionnaire.

<sup>7</sup> Number of questionnaires by average time divided by minutes of the hour:  $(2,083 * 7 / 60)$ .

<sup>8</sup> Fixed costs between all questionnaires,  $1,500 / 2,493$ .

## CONCLUSIONS

The previous results reveal, firstly, a greater preference in responding to the questionnaire via postal mail, which was anticipated as suggested in the first paragraph of section 2. It is surprising that 85% of the sample responded to the paper questionnaire, even more so when considering that this is a group is quite familiarized with the Internet (55% of those interviewed use the Internet on a daily or weekly basis).

A notable difference is found between both questionnaires in terms of the response number and response quality, with the Internet questionnaire having a lower number of no response questions –in all types of questions used– and a “shortening” of the fieldwork by 53 days. At the same time, we should note that the Internet questionnaire had a higher cost than the mailed survey (a difference of 0.89 cents per questionnaire). Changing the contact mean with the interviewee –Internet instead of mail– may increase the number of questionnaires attained through the Internet, significantly reducing research costs.

The limited response rate obtained in the telephone survey (51 surveyed) may be due to the method selected to carry it out. Given the impossibility of using a 900 line, the interviewee was asked to call an automatic answering machine or to fill out and send a part of the letter or presentation indicating their telephone number and the best time to contact them. The high cost of an international call to a Spanish telephone number and the temporary separation between the intent to respond (when receiving the questionnaire) and the positive contact with the interviewer may explain this poor collaboration.

We shall end by noting that offering a variety of response modes and permitting the individual to decide which mode to use may imply that those who do not initially wish to cooperate may be encouraged to do so through a mode that they consider to be

comfortable and accessible (Olson *et al.*, 2012), thereby improving the representativeness of the study.

## BIBLIOGRAPHY

- AAPOR (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Available at: <http://www.aapor.org>, consulted on May 25, 2015.
- Azofra, María J. (1999). *Cuestionarios*. Madrid: CIS.
- Bosnjak, Michael (2008). “Prenotification in Web-based Access Panel Survey: The Influence of Mobile Text Messaging Versus e-mail on response rates and sample composition”. *Social Survey Computer Review*, 26: 313-233.
- Callegaro, Mario; Lozar Manfreda, Katja and Vehovar, Vasja (2015). *Web survey methodology*. London: Sage.
- Cea D’Ancona, M.<sup>a</sup> Ángeles (2002). *Análisis Multi-variante*. Madrid: Síntesis.
- Couper, Mick P. and Bosnjak, Michael (2010). “Internet Surveys”. In: Marsden, P. V. and Wright, J. D. (eds.). *Handbook of Survey Research*. Bingley: Emerald.
- Crawford, Scott D. *et al.* (2004). *Gaining Respondent Cooperation in College Web-based Alcohol Surveys*. 59<sup>th</sup> Annual Conference of the AAPOR, Phoenix.
- Dillman, Don; Smyth, Jolene and Christian, Leah (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys*. New York: Wiley.
- García Ferrando, Manuel and Llopis, Ramón (2015). “La encuesta”. In: García Ferrando, M. *et al.* (comps.). *El Análisis de la Realidad Social*. Madrid: Alianza.
- Hansen, Kasper M. and Tue Pedersen, Rasmus (2012). “Efficiency of Different Recruitment Strategies for Web Panels”. *International Journal of Public Opinion Research*, 24: 238-249.
- Harmon, M. A. *et al.* (2005). *Does Type of Pre-notification Affect Web-surveys Response Rates?* 60<sup>th</sup> Annual Conference of the AAPOR, Miami.
- Kreuter, Frauke; Müller, Gerrit and Trappmann, Mark (2013). “A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents”. *Sociological Methods and Research*, 43(3): 452-464.

- Lozar Manfreda, K. *et al.* (2008). "Web Surveys versus other Survey Modes: A Meta-analysis Comparing Response Rates". *International Journal of Market Research*, 50: 79-104.
- Messer, Benjamin L., Edwards, Michelle L. and Dillman, Don A. (2012). "Determinants of Item Non-response to Web and Mail Respondents in Three Address-Bases Mixed-Mode Surveys of the General Public". *Survey Practice*, 5(2). Available at: <http://www.surveypractice.org/index.php/Survey-Practice/article/view/45/pdf>, consulted on May 25, 2015.
- Messer, B. L. and Dillman, D. A. (2011). "Surveying the General Public over the Internet Using Address-based Sampling and Mail Contact Procedures". *Public Opinion Quarterly*, 75: 429-457.
- Moscoso, David *et al.* (2010). *Condiciones de Vida y Expectativas de Retorno de los Andaluces en el Exterior*. IESA-CSIC/Junta de Andalucía.
- Olson, Kristen (2013). "Do Nonresponse Follow-ups Improve or Reduce Data Quality?". *Journal of the Royal Statistical Society*, 176(1): 129-145.
- Olson, Kristen; Smyth, Jolene D. and Wood, Heather (2012). "Does Providing Respondents with their Preferred Survey Mode Really Increase Participation Rates?". *Public Opinion Quarterly*, 76(4): 611-635.
- Shih, T. H. A. and Fan, X. (2008). "Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis". *Field Methods*, 20: 249-271.
- Smyth, J. D. *et al.* (2010). "Using the Internet to Survey Small Towns and Communities". *American Behavioral Scientist*, 53: 1423-1448.
- Tourangeau, Roger; Conrad, Frederic G. and Couper, Mick P. (2013). *The Science of Web Surveys*. Oxford: Oxford University.

**RECEPTION:** October 30, 2015

**REVIEW:** February 26, 2016

**ACCEPTANCE:** July 20, 2016

# Comparación de métodos de campo en la encuesta

## *Comparison of Survey Data Collection Methods*

Vidal Díaz de Rada y Juan Antonio Domínguez Álvarez

### Palabras clave

- Encuestas por correo
- Encuestas por teléfono
- Internet
- Metodología de la investigación
- Respuestas de encuestas
- Trabajo de campo

### Key words

- Mail Surveys
- Telephone Surveys
- Internet
- Research
- Methodology
- Survey Responses
- Fieldwork

### Resumen

En este trabajo se presentan los resultados de una investigación a población general que utiliza conjuntamente una encuesta por correo postal, una encuesta telefónica y una encuesta a través de Internet. La muestra objeto de estudio recibe la *invitación* a cooperar mediante una comunicación por correo, y es el propio seleccionado el que decide cómo responder. La nota analiza la calidad de la respuesta (considerando el número de no respuestas), el tiempo necesario para llevar a cabo el trabajo de campo y el coste de la recogida de información. Los resultados desvelan que la realización de encuestas vía web reducen el tiempo del trabajo de campo, consiguen menor respuesta parcial y presentan un coste ligeramente superior a la encuesta realizada por correo. Ahora bien, la diferente estructura de costes de cada una permite concluir que de haber obtenido el total de los cuestionarios *online*, la reducción de costes hubiera supuesto un 18,2%.

### Abstract

This paper presents the results of a mixed-mode survey of the general population using a combination of postal, telephone, and Internet survey protocols. Potential respondents were invited to participate by ordinary mail and were allowed to choose their preferred response mode. The paper focuses on response quality (taking non-responses into consideration), fieldwork time and data collection cost. The results reveal that the Internet survey produces the lowest rate of non-responses and requires significantly less fieldwork time, although it is slightly more costly than the postal survey. However, when differences in cost structure are taken into account, we find that the same number of completed questionnaires could have been obtained through the Internet alone at a cost that is 18.2% lower than the mixed-mode survey.

### Cómo citar

Díaz de Rada, Vidal y Domínguez Álvarez, Juan Antonio (2017). «Comparación de métodos de campo en la encuesta». *Revista Española de Investigaciones Sociológicas*, 158: 137-148. (<http://dx.doi.org/10.5477/cis/reis.158.137>)

La versión en inglés de esta nota de investigación puede consultarse en <http://reis.cis.es>

**Vidal Díaz de Rada:** Universidad Pública de Navarra | [vidal@unavarra.es](mailto:vidal@unavarra.es)

**Juan Antonio Domínguez Álvarez:** IESA-CSIC | [jadominguez@iesa.csic.es](mailto:jadominguez@iesa.csic.es)

## INTRODUCCIÓN<sup>1</sup>

Una visión retrospectiva de la evolución en la recogida de información mediante encuesta desvela que el cuestionario con ordenador (CAI) está sustituyendo paulatinamente a la encuesta *tradicional* presencial. Este cuestionario, además de utilizar todos los tipos de preguntas disponibles (respuesta única, respuesta múltiple, respuesta textual o numérica, etc.), es más sencillo de administrar al llevar a cabo —de forma automática— «saltos» provocados por preguntas filtro, rutas complicadas, etc. No menos importante es que permite establecer *reglas de consistencia* que son comprobadas antes de grabar la información (Couper y Bosnjak, 2010).

El fin de este trabajo es considerar hasta qué punto estas características señaladas por la mayor parte de la literatura son aplicables a una parte de los andaluces residentes en el extranjero. A diferencia de otros trabajos donde es el investigador quien *elige* el tipo de cuestionario a utilizar, en este el entrevistado decide cómo responderlo, implicando así una preferencia manifiesta hacia uno u otro tipo.

Una muestra de la población andaluza en el extranjero fue contactada mediante correo ordinario y se le propusieron tres formas de respuesta: cuestionario de papel, cuestionario disponible en un sitio web o encuesta telefónica. El presente trabajo se centra en el estudio de las diferencias entre cuestionarios considerando el número de preguntas no respondidas, la velocidad de respuesta y el coste del trabajo de campo. Contar con una población general, y el empleo de un mismo cuestionario en tres formatos, permitirá comprobar hasta qué punto las ventajas —evidenciadas en investigaciones realizadas con

muestras específicas— son generalizables a una muestra de población general.

## UNAS BREVES CONSIDERACIONES TEÓRICAS SOBRE EL MEDIO DE CONTACTO

Utilizar el correo como medio de contacto y de respuesta precisa una adecuada justificación en un momento en el que las *comunicaciones* a través de la red están reduciendo al mínimo la tradicional comunicación postal. Varias explicaciones justifican esta forma de proceder:

1. Menor *tasa de colaboración* de la encuesta a través de Internet en comparación con la encuesta postal, algo que ha sido constatado en numerosas investigaciones (entre otras, Lozar Manfreda *et al.*, 2008; Shih y Fan, 2008; Messer y Dillman, 2011; Tourangeau *et al.*, 2013, Callegaro *et al.*, 2015).
2. Mayores *dificultades de acceso* al entrevistado a través de la web, por lo que la mayor parte de la literatura específica recomienda realizar un primer contacto a través de otro modo; por ejemplo, emplear encuestas web *apoyadas* en encuestas (tradicionales) por correo (Hansen *et al.*, 2012; Messer y Dillman, 2011; Messer *et al.*, 2012).

Smyth *et al.* (2010) demostraron que enviar una encuesta por correo un día después del contacto mediante web aumenta la cooperación un 33%, logrando una tasa de respuesta del 55%. Además, se incorporan a la muestra personas con distintos rasgos sociodemográficos, lo que mejora notablemente la representatividad. En esta línea se manifiestan Messer y Dillman (2011) cuando advierten que las encuestas por correo representan mejor que las encuestas mixtas los rasgos sociodemográficos de la población.

<sup>1</sup> Una versión previa fue presentada en el XII Congreso Español de Sociología. Los autores agradecen sus recomendaciones al Consejo Editorial de la revista y a los tres evaluadores anónimos.

3. Como *tercer* argumento, la gran trascendencia de la *herramienta* empleada para establecer el primer contacto con el seleccionado. Crawford *et al.* (2004), por ejemplo, consideran que el contacto mediante correo ordinario consigue una tasa de respuesta significativamente mayor que el correo electrónico (tasas de respuesta del 52,5 y del 44,9% respectivamente), mientras que Harmon *et al.* (2005) comparan el correo electrónico con carta adjunta del patrocinador de la investigación, el email con carta adjunta de la empresa que realiza el trabajo de campo y la carta por correo del patrocinador de la investigación, logrando tasas de respuesta del 64,4, 63,6 y 69,9% respectivamente.

Bosnjak (2008) demuestra que el envío de mensajes sms es más efectivo que el aviso por correo, con tasas del 84% y del 71%. Otros aspectos influyentes son la extensión de la comunicación (las invitaciones largas son más efectivas que las cortas) y el tono (formal/informal) empleado en ella (Tourangeau *et al.*, 2013).

## UNA APLICACIÓN AL ESTUDIO «CONDICIONES DE VIDA Y EXPECTATIVAS DE RETORNO DE LOS EMIGRANTES DE ANDALUCÍA»

En el año 2009 la Consejería de Gobernación de la Junta de Andalucía encargó un estudio cuyo objetivo era conocer la situación de los andaluces en el extranjero, utilizando un cuestionario de 57 preguntas (136 variables) con siete temáticas: a) rasgos sociodemográficos; b) convivencia y condiciones de la vivienda; c) preocupaciones, trabajo e ingresos; d) utilización de instituciones asistenciales; e) evaluación de diversas instituciones y satisfacción con su vida; f) cercanía/lejanía a Andalucía; y g) situación social en Andalucía y motivos de la emigración (Moscoso *et al.*, 2010).

La mayor parte de las preguntas del cuestionario tratan sobre *hechos* y situaciones, aunque hay también doce preguntas sobre *aspectos subjetivos*, siguiendo la clasificación de García Ferrando y Llopis (2015). Las preguntas cerradas empleadas tienen una estructura muy variada:

- Preguntas de una respuesta que utilizan «botones de opción» en el cuestionario web, con predominio de cinco y dos categorías de respuesta. Es el tipo de pregunta más frecuente, distinguiendo entre preguntas nominales (17 preguntas y otras tantas variables) –que fundamentalmente recogen hechos– y 16 preguntas ordinales, de las que nueve versan sobre valoraciones y opiniones (aspectos subjetivos)<sup>2</sup>.
- Preguntas de respuesta múltiple, todas sobre comportamientos y situaciones, empleando dos *formatos*: preguntas con un listado de respuestas que recogen información sobre situaciones de convivencia, tipo de protección médico-sanitaria, pago de medicamentos, personas con las que emigró y familiares que se quedaron en Andalucía.

Se utilizan también respuestas con formato dicotómico en cada ítem, y versan sobre equipamiento del hogar, cambio de hábitos por problemas económicos, demanda de servicios por parte de instituciones y discapacidad.

2 Detalle del número de preguntas según el número de categorías de respuesta:

|                              | Nominal | Ordinal | Total |
|------------------------------|---------|---------|-------|
| Dicotómicas:                 | 9       |         | 9     |
| Tres categorías de respuesta | 2       | 5       | 7     |
| Cuatro categorías            | 1       |         | 1     |
| Cinco categorías             | 4       | 8       | 12    |
| Seis y más categorías        | 1       | 3       | 4     |
| Total                        | 17      | 16      | 33    |

Son 5 y 4 preguntas respectivamente, que proporcionan 30 y 31 variables<sup>3</sup>.

- Preguntas de batería; tres preguntas con 6, 6 y 8 ítems respectivamente, 20 variables, que se responden con una escala de 1 a 10. Son preguntas de opinión que recogen el grado de confianza sobre instituciones públicas y privadas, así como la satisfacción con una serie de aspectos de su vida.

El cuestionario cuenta también con doce preguntas abiertas sobre comportamientos. Seis precisan de respuestas numéricas<sup>4</sup>, una de una respuesta muy breve (municipio de nacimiento) y otras cinco se responden con varias palabras o una frase<sup>5</sup>. Se combinan, de este modo, los tipos de preguntas abiertas más sencillas y más complejas (Dillman, Smyth y Christian, 2014).

El cuestionario fue respondido —en su versión web<sup>6</sup>— en un promedio de 20,7 minutos.

El universo objeto de estudio fue la población andaluza residente en el extranjero, cifrada en 182.928 personas (Moscoso, 2010). Empleando como marco muestral el *Censo Electoral de Residentes en el Extranjero* se elaboró una muestra de 2.493 personas. La muestra, estratificada por zonas geográficas de residencia (Europa, Magreb,

América Latina, Norteamérica y Oceanía) fue contactada por correo postal y disponía de tres canales para responder el cuestionario: cuestionario por correo ordinario (a franquear en destino), cuestionario electrónico (pulsando sobre un enlace web) y entrevista telefónica CATI (sufragada por el organismo que lleva a cabo la investigación).

El proceso de implementación ha sido análogo al realizado en investigaciones similares (Dillman, Smyth y Christian, 2014), aunque no se utilizó ningún incentivo monetario ni se cambió el tipo de contacto. Concretamente, la *primera oleada* recibió por correo un cuestionario con una carta de presentación que expone los motivos por los que se realiza la investigación, se anima a responder y se explican las diferentes alternativas de respuesta. Estaba firmada por el viceconsejero de Gobernación de la Junta de Andalucía y en todo el cuestionario aparecía el logotipo gubernamental. Los que no respondieron recibieron 20 semanas después una nueva carta y un cuestionario, donde se le recordaba que todavía no habían respondido. Al no conseguir un tamaño muestral suficiente, dos meses más tarde se llevó a cabo un segundo envío a una nueva muestra, que no recibió ningún recordatorio (*segunda ola*).

El objetivo de este artículo es analizar cómo la modalidad produce diferencias entre los cuestionarios en número de respuestas, velocidad de respuesta y coste. Como hipótesis de partida se plantea que el cuestionario mediante internet obtendrá un mejor número de preguntas no respondidas, permitirá reducir el trabajo de campo y disminuirá el coste.

## RESULTADOS

Las 15.657 cartas enviadas obtuvieron 2.493 respuestas, lo que supone una tasa de cooperación del 18,9% (AAPOR, 2015). La participación fue muy superior a través del correo con 2.083 cuestionarios recogidos

<sup>3</sup> Las diferencias entre ambas, y entre estas y las preguntas de batería, son presentadas en la página 15 del texto de Azofra (1999).

<sup>4</sup> Número de personas en el hogar, número de habitaciones en la vivienda, ingresos, número de veces que ha viajado a Andalucía, año en el que emigró y año en el que comenzó a vivir en ese país.

<sup>5</sup> Ocupaciones más importantes, problemas/necesidades más importantes, motivos que le retienen en ese país, motivos por los que emigró y motivos de elección de ese país. Cada pregunta puede responderse hasta con tres respuestas.

<sup>6</sup> La versión web, en la medida en que recoge con exactitud la hora de inicio y finalización, permite conocer con precisión el tiempo empleado en la respuesta del cuestionario.

frente a los 359 recibidos a través de la web y los 51 obtenidos mediante encuesta telefónica CATI. Estos últimos no serán considerados en este trabajo dado su escaso tamaño muestral. Como puede apreciarse en la tabla 1, la muestra que responde por internet es más joven y presenta un mayor nivel de formación.

Esta situación no se explica por la falta de infraestructura o la falta de habilidades en el uso de internet, puesto que dos preguntas sobre uso de internet revelan que el 60% de los que han respondido por correo dispone de conexión a internet en su domicilio. Además, se trata de usuarios *habituales* en la medida en que un 36% se conecta a la red diariamente y un 18% todas las semanas, lo que implica que más de la mitad de los que responden por correo utilizan internet de manera habitual. Resulta sorprendente, por otro lado, que un 8% de los que responden la encuesta por internet no disponen de conexión en su hogar. No obstante, tanto la tasa de respuesta como la caracterización muestral van en la línea de investigaciones similares (entre otros, Dillman, Smyth y Christian, 2014).

### Calidad de la respuesta

El número medio de variables no respondidas por cada entrevistado fue de 11,2, una tasa relativamente baja si se tiene en cuenta que el cuestionario recoge información de 137 variables. Una primera explicación de este número de no respuesta está relacionada con el tipo de preguntas utilizadas, esperando menos respuestas de las preguntas abiertas y, dentro de esas, de las preguntas abiertas «tipo texto». De hecho, estas últimas constituyen más de la mitad de la no respuesta (5,5 variables no respondidas), siendo significativamente mayor en la encuesta por correo (5,6) que en la encuesta web (4,4).

Se aprecia menor diferencia en las respuestas abiertas numéricas, con una tasa media de no respuesta de 0,42 frente a las 0,47 de la encuesta por correo y el 0,17 de encuesta web.

Más sorprendente resulta el elevado número de no respuesta en las respuestas cerradas, que alcanza una media de 5,3 y supone un 47% de la no respuesta total (5,3/11,2). El análisis por modalidades de recogida de información evidencia una baja

**TABLA 1.** Rasgos de la muestra según modalidad elegida (porcentajes verticales)

|                   | Correo | Web |                    | Correo | Web |
|-------------------|--------|-----|--------------------|--------|-----|
| Chi-cuadr. 0,701  |        |     | Chi-cuadr. 129,9** |        |     |
| Hombre            | 49     | 51  | Sin estudios       | 21     | 5   |
| Mujer             | 51     | 49  | Est. primarios     | 20     | 10  |
| Nº de casos       | 2.083  | 359 | Est. secundarios   | 31     | 31  |
|                   |        |     | Est. superiores    | 27     | 53  |
|                   | Correo | Web | Nº de casos        | 2.083  | 359 |
| Chi-cuadr. 94,9** |        |     |                    |        |     |
| 19-29             | 8      | 15  |                    |        |     |
| 30-44             | 18     | 32  |                    |        |     |
| 45-59             | 25     | 30  |                    |        |     |
| 60 y más años     | 48     | 23  |                    |        |     |
| Nº de casos       | 2.083  | 359 |                    |        |     |

\*\* < 0,01.



tasa en la encuesta web, con un 2,3 frente a la elevada tasa obtenida en la encuesta por correo, que alcanza el 5,9. Cuando se analiza la diferencia entre preguntas sobre hechos y preguntas sobre opiniones las primeras presentan una mayor respuesta, aunque la diferencia no llega a ser significativa.

Son numerosos los investigadores que han dado cuenta de la relación existente entre la calidad de la respuesta y el momento de responder (entre otros, Olson, 2013; Kreuter, Müller y Trappmann, 2013). Estos plantean que los entrevistados que contestan inmediatamente tras ser contactados responden mejor que aquellos que lo hacen tras diversas invitaciones; hecho que queda probado también en el presente estudio. La comparación entre los que responden inmediatamente y los que lo hacen tras el primer recuerdo (columnas «sin recuerdo» y «recuerdo» en la tabla 2) desvela un aumento de más de dos variables no respondidas, diferencia que llega a tres en el caso de la encuesta por correo. Visto en términos porcentuales, los que responden inmediatamente presentan una media de 8,8%

de variables no respondidas, que aumenta al 10,3% tras recibir el recordatorio donde se le informa que todavía no han respondido el cuestionario.

La *segunda ola*, enviada a una nueva muestra unas semanas después del recuerdo, desvela la estabilidad de los hallazgos: 11,1 variables sin responder que supone un 8,8% del cuestionario (columna derecha de la tabla 2).

Lo tratado hasta el momento centra la atención en el número de no respuestas, pero no considera que puede haber entrevistados que no han proporcionado ni una respuesta, mientras que otros han respondido el cuestionario casi en su totalidad. Esta «nueva» línea de argumentación fija la atención en los entrevistados localizados y analiza el número de personas que no han proporcionado respuestas en cada modalidad de administración. Un 8% de los entrevistados en cuestionarios web han respondido TODAS las preguntas del cuestionario, y algo más de la mitad (52,6%) han dejado sin res-

**TABLA 2.** Número de no respuestas según modalidad y momento de respuesta del cuestionario

|                         | Promedio       |                            |       |                   |       |
|-------------------------|----------------|----------------------------|-------|-------------------|-------|
|                         | Sin recuerdo   | Primera oleada<br>Recuerdo | Total | Segunda<br>oleada | Total |
| Correo                  | 11,2           | 14,1                       | 12,0  | 11,9              | 12,0  |
| Web                     | 6,7            | 7,3                        | 6,9   | 6,8               | 6,9   |
| Total (media variables) | 10,6           | 12,9                       | 11,2  | 11,2              | 11,2  |
|                         | En porcentajes |                            |       |                   |       |
|                         | Sin recuerdo   | Primera oleada<br>Recuerdo | Total | Segunda<br>oleada | Total |
| Correo                  | 9,2%           | 11,2%                      | 9,8%  | 9,5%              | 9,7%  |
| Web                     | 5,8%           | 6,1%                       | 5,9%  | 5,7%              | 5,9%  |
| Total (% variables)     | 8,8%           | 10,3%                      | 9,0%  | 8,8%              | 9,1%  |
| Nº de casos             | 1.562          | 595                        | 2.198 | 285               | 2.442 |

**TABLA 3.** *Calidad en la respuesta, considerando el porcentaje de entrevistados que dejan preguntas sin responder (porcentajes verticales)*

| Preguntas no respondidas         | Número de no respuestas |      |
|----------------------------------|-------------------------|------|
|                                  | Correo                  | Web  |
| Ninguna                          | 3,0                     | 8,1  |
| Menos del 5% de las preguntas    | 30,0                    | 52,6 |
| Entre el 5,1 y el 10%            | 34,9                    | 31,5 |
| Entre el 11 y el 20%             | 22,0                    | 4,2  |
| Entre el 21 y el 30%             | 5,8                     | 0,6  |
| Entre el 31 y el 50%             | 3,4                     | 1,4  |
| Más de la mitad del cuestionario | 0,9                     | 1,7  |
| Tamaño muestral:                 | 2.083                   | 359  |

ponder menos de un 5%; porcentajes que en la encuesta por correo se reducen al 3% y al 30% respectivamente (tabla 3). Analizando conjuntamente las cifras se observa que el 60% de los que contestan a través de la web ha respondido casi a la totalidad de la encuesta (9,5 preguntas de 10), frente al 33% que lo hace en la encuesta por correo.

Las diferencias en la modalidad de administración, sin considerar otras variables, podría llevar a una impresión equivocada cuando un análisis pormenorizado permite apreciar que la encuesta por correo ha sido la modalidad elegida fundamentalmente por personas de mayor edad y bajo nivel educativo, justo lo contrario que sucede en la encuesta web, respondida mayoritariamente por personas jóvenes con alto nivel de estudios (tabla 1).

Para analizar cómo influyen los rasgos sociodemográficos y la modalidad de encuesta elegida se ha optado por una regresión múltiple (mínimos cuadrados). Se ha elegido como variable dependiente el número de no respuestas y como variable independiente —en el primer modelo— la modalidad de respuesta y —posteriormente— los rasgos sociodemográficos de los entrevistados. Se busca medir su influencia, siguiendo

la metodología empleada en otras investigaciones (entre otras, Messer y Dillman, 2011; Messer *et al.*, 2012). Los resultados, que se muestran en la tabla 4, desvelan la gran influencia de la modalidad de cuestionario en el número de variables sin respuesta (modelo 1).

Cuando en el primer modelo se introducen los rasgos sociodemográficos y la zona de residencia (modelo 2), la modalidad sigue teniendo influencia significativa en el número de no respuestas, aun cuando se controla la influencia de las variables sociodemográficas. La edad es el término más influyente, seguida de la modalidad elegida, el nivel educativo y el sexo. El signo de los coeficientes indica que a medida que aumenta la edad se incrementa el número de no respuestas, del mismo modo que la encuesta web genera una disminución del número de preguntas sin respuesta. Los entrevistados con menor nivel educativo, y las mujeres, producen un menor número de respuestas en el cuestionario. El poder explicativo del modelo ( $R^2$  0,153) es ligeramente superior al logrado en investigaciones similares (entre otros, Messer *et al.*, 2012), lo que indica el bajo poder explicativo de las variables sociodemográficas en el número de preguntas no respondidas.

**TABLA 4.** Regresión sobre los factores que influyen en el número de no respuestas según modalidad de cuestionario, rasgos sociodemográficos y zona de residencia

|   | Variable dependiente: número de no respuestas |            |                              |            |
|---|---|------------|------------------------------|------------|
|   | MODELO 1                                      |            | MODELO 2                     |            |
|   | Coefficientes estandarizados                  | t. Student | Coefficientes estandarizados | t. Student |
| Modalidad de cuestionario               |   |            |                              |            |
| (Referencia: correo)                    | -0,219  | -10,914**  | -0,132                       | -6,675**   |
| Sexo (Referencia: mujer)                |   |            | -0,064                       | -3,360**   |
| Edad (continua)                         |   |            | 0,273                        | 11,742**   |
| Nivel educativo (Ordinal <sup>a</sup> ) |   |            | -0,087                       | -3,745**   |
| Zona geográfica de residencia           |   |            |                              |            |
| Europa (Ref: otras zonas)               |   |            | 0,009                        | -0,222     |
| Latin América (Ref.: otras zonas)       |   |            | -0,085                       | -2,117*    |
| Norte América (Ref.: otras zonas)       |   |            | -0,036                       | -1,504     |
| Otros (Ref.: otras zonas)               |   |            | 0,008                        | 0,382      |
| R2                                      | 0,048   |            | 0,153                        |            |
| Número de casos                         | 2.420   |            | 2.420                        |            |

Codificación de las variables:

(\*) Significación al 0,05.

(\*\*) Significación al 0,01.

<sup>a</sup> No sabe leer ni escribir (1), sin estudios terminados (2), primarios (3), secundarios (4) y universitarios (5). Se sigue aquí, y en la zona geográfica, el modelo de codificación propuesto por Cea D'Ancona (2002).

En un tercer análisis se añadieron cuatro interacciones: sexo-estudios, uso de internet-edad, internet-estudios e internet-sexo; con una aportación nula al modelo (motivo por el que no se incluyen aquí).

De este apartado podemos extraer tres conclusiones; por un lado, que la encuesta en papel proporciona peores respuestas, que la mitad de «no respuestas» se producen en preguntas cerradas, y, por último, que considerar únicamente los cuestionarios totalmente respondidos (con el 90-95% completado) supondría reducir a la mitad el tamaño muestral de la encuesta telefónica y por correo.

### «Velocidad» de respuesta

La lentitud en la recogida de información ha sido, sin duda, otro de los grandes inconvenientes

de la encuesta por correo con cuestionario de papel. De hecho, en la presente investigación se estuvieron recibiendo cuestionarios por correo durante más de tres meses (103 días), a los que habría que añadir 43 días más tras el envío del recuerdo, lo que supone un total de 146 días hasta disponer de la información de la primera ola. Frente a esto, la encuesta web precisó de 78 días tras el primer envío (25 días menos), y otros 35 tras el recuerdo (8 menos), lo que supone una duración de 113 días, 33 días menos que los necesitados en la encuesta por correo (tabla 5).

En la *segunda oleada*, que como se ha señalado no utilizó recuerdos, la encuesta por correo precisó de 48 días frente a los 28 días utilizados para la encuesta web. Considerado conjuntamente todo el tiempo nece-

**TABLA 5.** *Tiempo de respuesta (en días)*

|                  | Diferencia (en días) desde la fecha de envío y la fecha de grabación |                            |                         |                   |                     |
|------------------|--|----------------------------|-------------------------|-------------------|---------------------|
|                  | Sin recuerdo   | Primera oleada<br>Recuerdo | Total<br>(suma de días) | Segunda<br>oleada | Total<br>(Dos olas) |
| Mail             | 103  | 43                         | 146                     | 48                | 194                 |
| Web              | 78   | 35                         | 113                     | 28                | 141                 |
| Total (nº casos) | 1.562  | 594                        | 2.197                   | 285               | 2.492               |

sario para el trabajo de campo, los 194 días de la encuesta postal se reducen a 141 en la encuesta web, lo que implica disponer de los resultados 53 días antes. Esta cifra supone una reducción del 27% en el tiempo empleado.

### Coste de la investigación

El tercero de los aspectos a estudiar, referido al coste de la investigación, considerará únicamente el importe de recepción de los cuestionarios válidos, excluyendo los costes de impresión, manipulación y envío, que no varía entre modalidades. Tampoco se tendrá en cuenta el análisis de datos, ya que no cambia en función de la modalidad empleada.

Centrando la atención en la modalidad autoadministrada debe tenerse en cuenta la diferente estructura de costes. La mayor parte del coste de la encuesta web es fijo, lo que implica que disminuye el coste unitario a medida que aumenta el número de cuestionarios recibidos. La encuesta postal, por su parte, presenta un coste variable que se incrementa en función del número de cuestionarios.

En el cuestionario web hay que considerar las horas empleadas en la programación y el test del cuestionario (50 horas), el coste del servidor web (*hosting*) donde se instaló el cuestionario (150 euros anuales), y las horas de supervisión y resolución de incidencias

(40 horas aproximadamente). Los costes del cuestionario por correo se limitan a la recepción y la grabación de la información. El prefranqueo —del sobre de 40 gramos— tiene un coste unitario de 1,33 euros si proviene de Europa y de 1,80 euros cuando proviene de otros países. En la grabación del cuestionario se empleó un promedio de 7 minutos por cuestionario.

Considerando un coste en recursos humanos de 15 euros/hora, el trabajo de campo web supuso un coste de 1.500 euros frente a los 6.853,7 del postal. Ahora bien, los 359 cuestionarios web suponen un coste unitario de 4,18 euros; mientras que los 1.151 cuestionarios por correo procedentes de Europa y los 932 procedentes del resto del mundo implicaron un coste unitario de 3,29 euros, donde se incluyen las 2437 horas empleadas en su grabación (tabla 6).

Considerando que la estructura de costes de la encuesta web es fijo independientemente del número de respuestas, conseguir otros 97 cuestionarios web (hasta los 456) hubiera logrado igualar costes entre modalidades, disminuyendo el coste por cuestionario (web) a partir de este momento. De hecho, si la opción de cuestionario web hubiera sido la modalidad principal y todos los cuestionarios autoadministrados se hubieran conseguido a

7 Número de cuestionarios por tiempo promedio dividido entre minutos de hora:  $(2.083 \cdot 7) / 60$ .

**TABLA 6.** Costes del trabajo de campo

| Concepto               | WEB            |                  | Concepto           | CORREO           |                  |
|------------------------|----------------|------------------|--------------------|------------------|------------------|
|                        | Tiempo (horas) | Coste (15€/hora) |                    | Nº cuestionarios | Coste (15€/hora) |
| Programación           | 50             | 750              | Franqueo           |                  |                  |
| Alojamiento servidor   |                | 150              | Europa (1,33€)     | 1.151            | 1.530,83         |
| Trabajo de campo       | 40             | 600              | Resto (1,8€)       | 932              | 1.677,60         |
|                        |                |                  | Grabación (1,75*€) | 2.083            | 3.645,25         |
| Total (euros)          |                | 1.500            |                    |                  | 6.853,68         |
| Coste por cuestionario |                | 4,18             |                    |                  | 3,29             |

\* Nota: cada cuestionario precisa un tiempo medio de grabación de 7 minutos; 7,57 cuestionarios cada hora. Considerando un salario de 15 euros/hora, 1,75 euros por cuestionario.

través de este medio (2.493 cuestionarios), se hubiera logrado un coste unitario de 0,60<sup>8</sup> euros, un 18,2% del coste unitario del cuestionario postal.

## CONCLUSIONES

Los resultados anteriores desvelan, en primer lugar, una mayor preferencia en responder el cuestionario a través del correo, aspecto esperado tal y como se planteó en el segundo párrafo del apartado 2. Ahora bien, resulta sorprendente que el 85% de la muestra respondiera el cuestionario de papel, mucho más cuando se constata que es un colectivo muy familiarizado en el uso de internet (el 55% de los entrevistados usan internet a diario o semanalmente).

Existe una notable diferencia entre ambos cuestionarios en el número de respuestas respondidas y en la calidad de la respuesta, presentando el cuestionario web un menor número de preguntas no respondidas —en todos los tipos de preguntas empleadas— y un «acortamiento» del trabajo de campo en 53 días. Ahora bien, debe tenerse en cuenta que el cuestionario web presenta

un coste superior a la encuesta por correo (diferencia de 0,89 céntimos por cuestionario). Cambiar el medio de contacto con el entrevistado —web en lugar de correo— podría aumentar el número de cuestionarios conseguidos a través de la red y reducir notablemente el coste de la investigación.

La escasa respuesta obtenida en la encuesta telefónica (51 encuestados) puede deberse al modo elegido para realizarla. Ante la imposibilidad de utilizar una línea 900 se optó por pedir al entrevistado que llamara a un contestador automático o cumplimentara y enviara una parte de la carta de presentación, indicando su número de teléfono y la mejor hora para contactar. El elevado coste de la llamada internacional a un número telefónico de España, y la «separación temporal» entre la intención de responder (cuando se recibe el cuestionario) y el contacto positivo con el entrevistador pueden explicar esta baja colaboración.

Terminar señalando que disponer de varios modos de respuesta y permitir que el seleccionado decida cuál utilizar puede implicar que personas que no deseaban cooperar en un primer momento se animen a hacerlo al poder utilizar un medio que consideran cómodo y accesible (Olson *et al.*, 2012), mejorando así la representatividad de la investigación.

<sup>8</sup> Costes fijos entre total de cuestionarios, 1.500 / 2.493.

## BIBLIOGRAFÍA

- AAPOR (2015). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Disponible en: <http://www.aapor.org>, acceso el 25 de mayo de 2015.
- Azofra, María J. (1999). *Cuestionarios*. Madrid: CIS.
- Bosnjak, Michael (2008). «Prenotification in Web-based Access Panel Survey: The Influence of Mobile Test Messaging Versus e-mail on response rates and sample composition». *Social Survey Computer Review*, 26: 313-233.
- Callegaro, Mario; Lozar Manfreda, Katja y Vehovar, Vasja (2015). *Web survey methodology*. London: Sage.
- Cea D'Ancona, M.<sup>a</sup> Ángeles (2002). *Análisis Multivariante*. Madrid: Síntesis.
- Couper, Mick P. y Bosnjak, Michael (2010). «Internet Surveys». En: Marsden, P. V. y Wright, J. D. (eds.). *Handbook of Survey Research*. Bingley: Emerald.
- Crawford, Scott D. et al. (2004). *Gaining Respondent Cooperation in College Web-based Alcohol Surveys*. 59<sup>th</sup> Annual Conference of the AAPOR, Phoenix.
- Dillman, Don; Smyth, Jolene y Christian, Leah (2014). *Internet, Phone, Mail, and Mixed-Mode Surveys*. New York: Wiley.
- García Ferrando, Manuel y Llopis, Ramón (2015). «La encuesta». En: García Ferrando, M. et al. (comps.). *El Análisis de la Realidad Social*. Madrid: Alianza.
- Hansen, Kasper M. y Tue Pedersen, Rasmus (2012). «Efficiency of Different Recruitment Strategies for Web Panels». *International Journal of Public Opinion Research*, 24: 238-249.
- Harmon, M. A. et al. (2005). *Does Type of Pre-notification Affect Web-surveys Response Rates?* 60<sup>th</sup> Annual Conference of the AAPOR, Miami.
- Kreuter, Frauke; Müller, Gerrit y Trappmann, Mark (2013). «A Note on Mechanisms Leading to Lower Data Quality of Late or Reluctant Respondents». *Sociological Methods and Research*, 43(3): 452-464.
- Lozar Manfreda, K. et al. (2008). «Web Surveys versus other Survey Modes: A Meta-analysis Comparing Response Rates». *International Journal of Market Research*, 50: 79-104.
- Messer, Benjamin L., Edwards, Michelle L. y Dillman, Don A. (2012). «Determinants of Item Nonresponse to Web and Mail Respondents in Three Address-Bases Mixed-Mode Surveys of the General Public». *Survey Practice*, 5(2). Disponible en: <http://www.surveypractice.org/index.php/SurveyPractice/article/view/45/pdf>, último acceso el 25 de mayo de 2015.
- Messer, B. L. y Dillman, D. A. (2011). «Surveying the General Public over the Internet Using Address-based Sampling and Mail Contact Procedures». *Public Opinion Quarterly*, 75: 429-457.
- Moscoco, David et al. (2010). *Condiciones de Vida y Expectativas de Retorno de los Andaluces en el Exterior*. IESA-CSIC/Junta de Andalucía.
- Olson, Kristen (2013). «Do Nonresponse Follow-ups Improve or Reduce Data Quality?». *Journal of the Royal Statistical Society*, 176(1): 129-145.
- Olson, Kristen; Smyth, Jolene D. y Wood, Heather (2012). «Does Providing Respondents with their Preferred Survey Mode Really Increase Participation Rates?». *Public Opinion Quarterly*, 76(4): 611-635.
- Shih, T. H. A. y Fan, X. (2008). «Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis». *Field Methods*, 20: 249-271.
- Smyth, J. D. et al. (2010). «Using the Internet to Survey Small Towns and Communities». *American Behavioral Scientist*, 53: 1423-1448.
- Tourangeau, Roger; Conrad, Frederic G. y Couper, Mick P. (2013). *The Science of Web Surveys*. Oxford: Oxford University.

**RECEPCIÓN:** 30/10/2015

**REVISIÓN:** 26/02/2016

**APROBACIÓN:** 20/07/2016