



THE QUALITY OF DATA SOURCES USED FOR RECRUITING SAMPLES FOR AUDIENCE RESEARCH

Participants in this group:

- Members of the Scientific Committee: Avner Bar-Hen, Jean Chiche, Anne-Marie Dussaix, Sébastien Lhote, Emmanuel Viennet and Thierry Vedel
- Members of CESP: Noëlie Bonnard, Valérie Morrisson and Patricia Schultz

1. CONTEXT

Today, the currency audience measurement surveys are carried out by telephone or via the Internet. This is why face-to-face, as a source of recruitment, is not mentioned in this note. Similarly, certain databases used to carry out public surveys (such as the Fidéli database) are not mentioned because they are not accessible to audience measurement studies.

The diversification of recruitment sources, that is to say the use of sources other than the phone number databases sources based on the phone directory or randomly generated databases, now concerns most of the audience surveys that CESP audits. The decline in reachability for phone surveys (with a risk of acceleration with the implementation of multi-purpose numbers by ARCEP on 1st January 2023), the increasing difficulty of reaching certain target groups (young people and lower social classes for example) and the higher costs that these phenomena entail, explain the use of new sources of recruitment by research agencies.

These new sources have one thing in common: they are qualified (by the research agency or, more often, by a service provider) and can be used to target individuals based on demographic criteria. The use of these sources can no longer be used for random or exhaustive recruitment.

In this note, CESP wanted to:

- List best practices when introducing a new recruitment source, particularly when using qualified databases other than a directory or a randomly generated database.
- Give a list of criteria for validating the quality of the recruitment source.
- Give its opinion on the appropriateness of using each of these sources.

The use of qualified sources is a subject of specific concern to CESP because it is linked to the representativeness of the recruited and interviewed sample. The main aim of audience research is to provide the indicators on which to base the purchase and sale of advertising space to the advertising industry. To guarantee a reliable currency and fair exchanges, it is necessary that that audience

measurement is unbiased, in other words it is representative of all French people, whatever their behaviour and media consumption habits.

By introducing new sources of recruitment, we can improve the structure of the samples but we must also ensure that this doesn't introduce new biases that would have an impact on the audiences.

2. DEFINITIONS

There are several recruitment sources used in currency audience studies.

It is important to define them in order to distinguish between them:

- Random base of phone numbers: there are several methods for producing random bases of phone numbers.
 - o By declination from numbers taken from the phone directory (this only concerns geographic numbers)
 - o By random generation from prefixes allocated to operators in mainland France by ARCEP. This applies to fixed and mobile numbers. The methods used to generate or clean up the databases may differ from one research agency to another, but the principle remains the same.
- Proprietary address pool: this is a database of individuals who have been questioned about a system and have agreed to respond to other surveys carried out by the research agency. These databases belong to the research agency and are managed internally. There is no reward system offered. Special case: the research agency may give itself the option of re-interviewing individuals who have already taken part in the audience research.
- Panel and Access Panel: these are individuals who claim that they will cooperate in future data collection if they are selected, generally in exchange for a bonus. There is a long-term relationship between the members of the panel/access panel - who understand the commitment they are being asked to make - and the service provider in charge of the database. These databases can be recruited from one or more sources or randomly (the good representativeness of the panel or access panel depends on these sources). We must emphasize a significant difference between a panel and an access panel: a panel is set up to be representative of a study population that can be re-interviewed over time (longitudinal analyses), whereas an access panel is a wider, non-representative pool, but which can be used to draw samples that will be used for surveys.
- File supplier: the service provider is a file or contact database aggregator. The individuals listed in these files have no connection with the address provider (they have not expressed that they wish to cooperate with this supplier) and they do not receive any compensation for their participation in a survey. In this note, we distinguish between two types of file: "directory" files, which cover a large proportion of active phone numbers, and qualified files which allow to target specific profiles (young people, high incomes...).
- Sponsorship or snowballing: an individual recruited to take part in a survey is asked (and rewarded) to provide contacts of people close to him/her who are also likely to take part in the survey (entourage: within or outside the household).

- "Intercepts": there is another category of recruitment sources not used today in currency audience research, which ESOMAR¹ refers to as "Intercepts". Interception is an approach where potential participants are invited to complete a survey (for a reward or not) while they are engaged in another activity, such as playing an online game, reading content, being active on a social network or any other online activity. The 'intercepted' participants may be unknown to the sample provider or have been pre-identified and qualified in a previous survey experiment.

3. BEST PRACTICES AND LIST OF CRITERIA

3.1. Best practices

We can list a number of good practices to be followed at several stages in the integration of these new sources of recruitment sources:

When selecting these sources, by gathering the information needed to assess the quality of the service provider supplying the contacts:

- To define the type of source involved: pool, panel, access panel, file supplier... in order to understand the link between the service provider and the individuals likely to be contacted.
- To obtain information on the sources used to create the pool, the panel, the access panel and the address file: the aim is to determine whether there is a significant lack of representativeness or whether one of the recruitment sources is related to the subject of the survey.
- To ask about the themes of the studies carried out on the source to identify if there is a link between the subject of the study and the other subjects most often dealt with by the service provider.
- To know the sampling method used by the research agency or service provider to draw up the sample to be supplied, and the selection rules (number of possible selections in the draws, maximum number of authorized solicitations...).
- To obtain information that will enable us to evaluate the contact management method and its exploitation. What are the management rules: address selection and integration into the databases, recruitment of panellists, number of selections, solicitations/participations, cleaning rules, reminder system and reward programme...? The aim is to ensure that contacts are not over-solicited, survey "experts" or "bounty hunters", which could jeopardise the success of the survey. This could compromise the quality of their responses.
- Finally, to describe the method used to qualify the profiles and give the recency and rate of updating of this information.

By carrying out analyses when the source is tested or integrated into the survey system:

- To carry out tests upstream, before introducing the new source into the recruitment or survey databases.
- To ensure that these sources are deduplicated with the survey databases used to avoid multiple solicitations.

¹ ESOMAR Global Research – March 2021 – Questions to help buyers of online samples

- To keep the "recruitment source" information in the files so that you can carry out analyses a posteriori (for example, when recruiting a panel: is there a different churn rate depending on the source?).
- To set up a profile validation system to assess the quality of sources on this criterion (are young people really young people?).
- To carry out analyses to estimate the possible impact on participation levels but also on audience levels and share these analyses with the survey sponsors.
- To include a question on participation in panels or a variable of interest.

In the use of these sources over time:

- Limit the weight of qualified sources (in favour of random sources)
- To keep the weight of each source stable over time unless it can be shown that changing the weight of the sources has no effect on the results.
- Maintain a similar sample structure from one year to the next to limit the impact on audience levels linked to these changes.

Special case of re-interviews

To avoid the risk of poor quality responses due to familiarity with the questionnaire:

- To sharply limit the use of this method
- Establish strict rules: number of authorised participations and minimum duration between resurveys
- Give yourself the means to control the participation of these individuals: length of time spent answering the questionnaire, setting a minimum acceptable duration.

3.2. First list of criteria

In this section, CESP has drawn up an initial list of criteria on which the bases can be evaluated (some criteria do not apply to all sources). There are no quantified indicators for the criteria listed, as analyses by source must be carried out on an ad hoc basis, taking into account all the information provided.

Criteria
Representativeness
Does the information provided by the service provider make it possible to check that the source used is representative of the universe that we are trying to represent?
Has the risk of using a source linked to audience measurement been evaluated?
What is the total size of the database used?
What proportion of the target audience (in the total base) does the sample provided represent?
How is the sample drawn from the database? Is a degree of randomness preserved?
What selection criteria does the service provider use to draw the sample from the database?
If the source is used regularly: are the sampling method and selection criteria stable over time?
Contact management
How many new contacts have been added each month over the last 12 months?
What is the monthly churn rate for each month over the last 12 months?
What is the maximum length of time a contact can remain in the database?
What is the maximum authorised number of selections / solicitations over 12 months, in any one month?
What is the average number of selections/solicitations observed over the last 12 months for the sample provided?
What is the average number of selections/solicitations observed over the last 30 days for the sample provided?
What are the criteria for excluding a contact?
What is the incentive protocol?
How much is the incentive for taking part in the survey?
What are the rules for follow-up after the initial contact?
How is the data updated?
How often is the demographic information updated?
What is the number of participations for each of the last 12 months for the sample provided?
What is the number of participations in "media" surveys for each of the last 12 months for the sample provided?
What is the average response rate observed in the database over the last 12 months? (Does not apply to file suppliers)
Quality of the file supplied
Has there been de-duplication with other survey databases?
Consistency between profiles sold and profiles observed: what is the percentage of consistent profiles?
What is the observed response rate for the sample supplied compared with the sample as a whole?
Impact on the quality of responses and results
What is the duration of the questionnaire for the sample provided compared to the sample as a whole?
Have any statistical tests or analyses been carried out in the last two years to identify if there is a link between the source and audience levels?

4. OPINION OF CESP

The following table gives the opinion of CESP on the use of each recruitment source in currency audience research.

Definitions :

- **Acceptable** (subject to compliance with best practices): the research agency may consider using this type of source, ensuring that quality criteria are respected.
- **To be used with caution:** CESP draws attention to the risks of losing representativeness by using this source. It recommends checking the quality of the source with the service provider beforehand and carrying out analyses to assess the impact of including this source in the sample. It also recommends limiting the use of this type of source to targets that are difficult to recruit.
- **To be avoided**

Type of qualified source	Opinion of CESP
Pool for re-interviewing: the research agency builds up a database of individuals who have already taken part in the survey with a view to being re-interviewed for the same survey.	To be used with caution and according to the studies, with sufficient time between two interrogations.
Proprietary address pool: the research agency builds up an address base with individuals who have taken part in one or more other studies carried out by the research agency.	Acceptable
Panel	Acceptable
Access panel	To be used with caution
Directory files enhanced by a service provider	Acceptable
Qualified files	To be used with caution
Recruitment of several people in the same household	To be used with caution
Recruitment by sponsorship or snowballing	To be avoided
Intercepts	To be avoided

Here are a few elements to complement the table above:

CESP points out that the use of proprietary address pools and panels, while acceptable subject to compliance with good practice, must remain limited. It should be noted that in some cases, address pools are derived from currency studies audited by CESP. In this case, their use is acceptable provided that the audit is satisfactory.

As regards to the recruitment of several people in the same household (as part of "individual" surveys), the cluster effect must be taken into account when analysing the results.

Recruitment by sponsorship is a method to be excluded for audience research: the "snowball effect" seriously undermines the representativeness of the sample. The Scientific Committee of CESP has recommended that individuals recruited via sponsorship be removed from the samples/panels.

Regarding "Intercepts" or on-the-fly recruitment, particularly on social networks: it should be possible to carry out in-depth tests in order to analyse the profiles and behaviour of individuals recruited using this method.

And finally, CESP would like to make it clear that the use of qualified sources must be limited: the sample recruited/interviewed on these sources must remain a complement to the total sample in the context of audience research.