

Mixed mode and (non)-response bias

Dominique Joye (UNIL) Caroline Roberts (UNIL)
Michèle Ernst Stähli (FORS)

August 21, 2013

1 Introduction

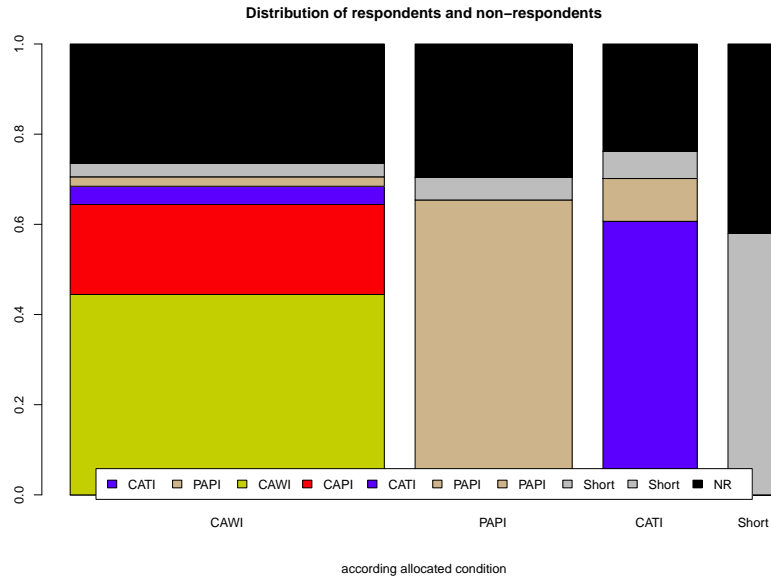
Many research programs use different surveys modes in order to acquire information. Very often too, these different surveys have very different response rates. At the end, the result could be a little bit embarrassing for the researchers as it become difficult to compare these surveys: how to know if the results are really comparable? How to know if we are facing selection or modes effects? In order to have a clearer idea of the impact of different methodologies, we have realised a mixed mode study in Switzerland at the turn of 2012 and 2013.

In this paper, we will use a very simple approach in a first time: to compare if the different surveys will give different results and lead to different scientific conclusions.

We will test also the effect of weighting by demographic characteristics, just in order to know if such a simple schema is pertinent in order to compare the results. This is a simple way to control by the selection effects of these different surveys.

2 Data and methods

On the basis of 3919 addresses obtained from the register of the Swiss Federal Statistical Office in the French part of Switzerland, we designed an experience based, in a first step, on web (2000 addresses), mail (1000 addresses), and CATI (600 addresses), keeping 319 as reserves and knowing that the ESS would give us a CAPI control. This has been followed by other modes in case of non-response in the attributed ones. The questionnaire, limited to 25 minutes of interviews (approximately 15 minutes for the main part, and 10 for the demographics) was of course related to the “Vulnerability” theme of the NCCR Lives. This was also in line with the Well-being module of the 6th wave of the ESS. In general, the final response rate was very high in each of the conditions.



In all the main modes, we end up with more than two thirds as response rate, which is rather high in general social survey. The result of the non-respondent survey allow us to end up at nearly 75%, still good but the amelioration is far to be as important than after a one hour face to face survey. In other words it seems that a maximum that can be obtained in this kind of surveys is around the three quarter of the sample.

3 Results 1: Differences in Surveys?

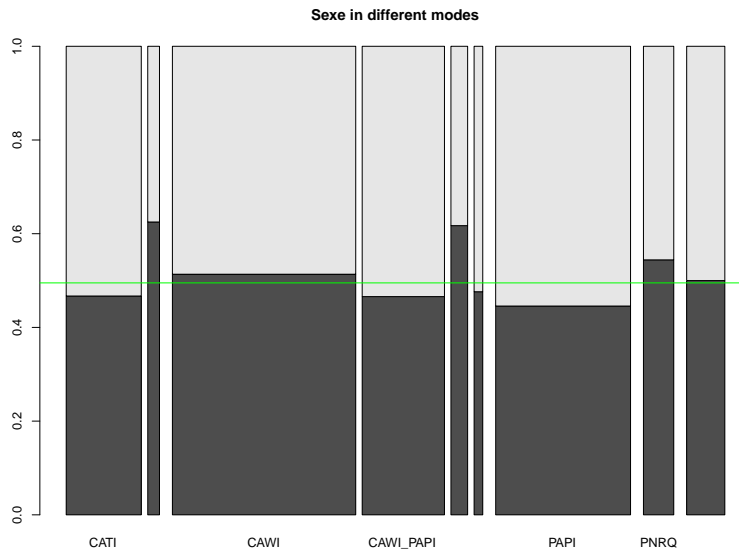
In order to test differences, we need, of course, to define a set of indicators. In the case, we have used demographic variables, but also opinion and behaviours indicators, in order to test the importance of social desirability or other mode effects.

3.1 Demographics

For the demographic variables, we have used before all sex, age and nationality. Two sets of reasons for such a choice: we suspect that the response to mode like CAWI will depend largely of the age while nationality is always a critical variable. Furthermore, these variables are also available in the sampling frame allowing a fine control of the respondents in this case.

3.1.1 Sex

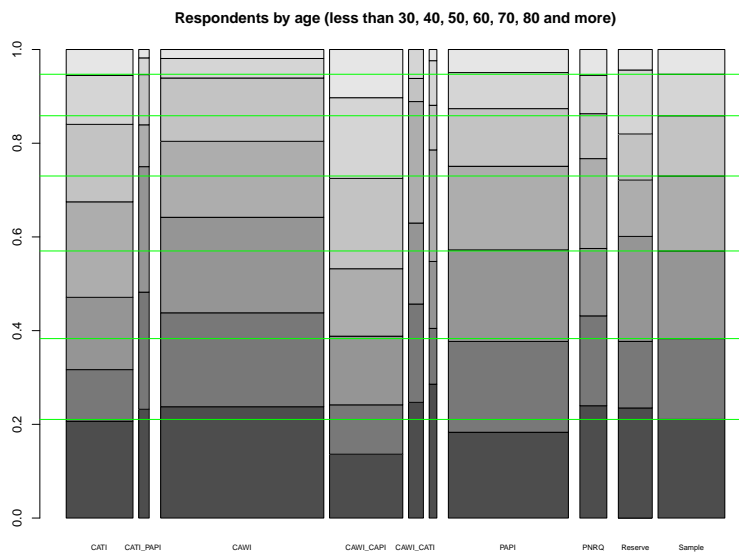
Not so much differences according to the sex of the respondent: just a general light over-representation of women, except for the CAWI where it is the correct value which is observed. There are more differences in the case of the following modes but once again the effective are small, that means that the impact on the final results will be small.



The red line indicates the level expected in case of perfect representation of men and women.

3.1.2 Age

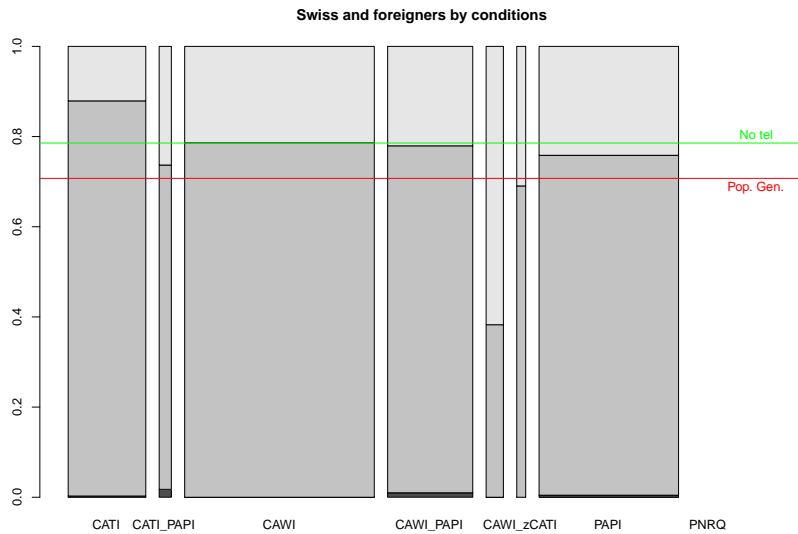
For age, it is worth to mention first the rather correct distribution in the case of the PAPI. In the case of CAWI, we have a clear under-representation of the oldest categories (more than 60 years old).



In the CATI case, the age classes between 30 and 50 are clearly under-represented but we know already the coverage problem of these categories.

3.1.3 Nationality

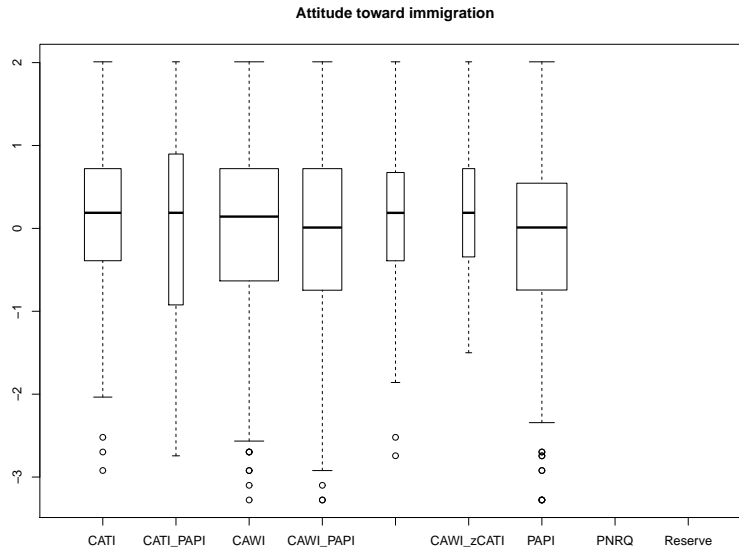
The nationality is one of the most difficult category in order of obtaining a good representativity. There are not so many differences between the PAPI and teh CAWI and to use PAPI has a second mode after a first PAPI does not change so much. CATI seems far less good but to be fair we have to compare with the proportion of foreigners that could be contacted by phone: the green line in the graph.



The red line represent the target: that means what we have to obtain in order to respect the proportion of the sample. In all cases, the first mode select too few foreigners. But the recalls with other modes, in particular the CAPI, seem very efficient in this case and really change the composition of the sample. Of course, CAPI is the most expensive mode but seems valuable in this context. As mentioned, we have also weighted the data according sex and age. This has no effect for this variable.

3.2 Political opinions

If we look at the simple indicator about political interest, we have some differences, in particular with people in the PAPI situation seeming less interested in politics. But we can also use opinions in order to measure potential differences.

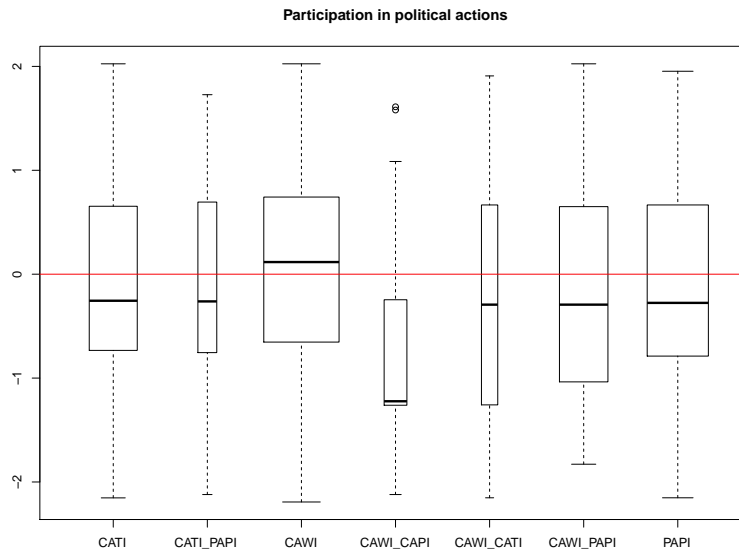


Attitude toward immigrants is a classical variables where we can expect difference between the different modes. In particular the lower response rate of the pure CAWI condition could result in a better support of immigration.

3.3 Participation

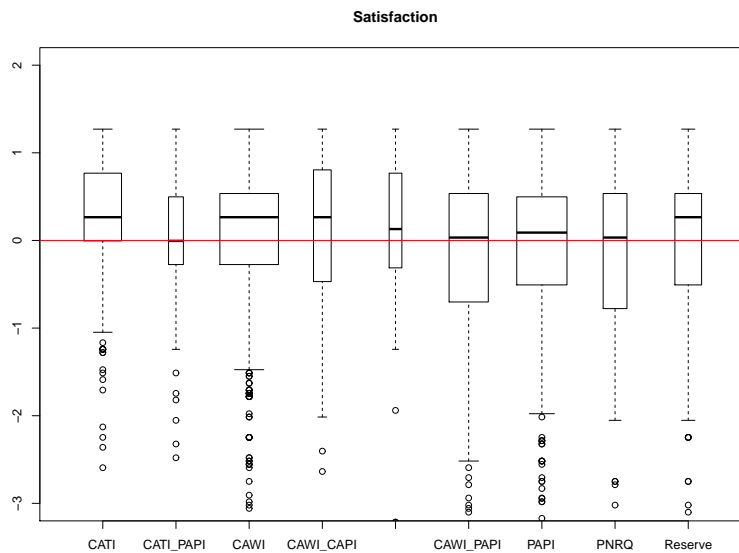
Participation is always a multidimensional concept. In the case, we propose to use a battery of question including political participation (boycott, vote and petition) as well as associative. This last category is broken up in 2 by the analysis: a first one linked to participation in leisure association (sport, culture, age driven) while the other one is in association more oriented toward action (humanitarian, religious or related to environment protection).

The weakest differences are in the general case of participation in association oriented toward leisure and the strongest one for more politically oriented participation.

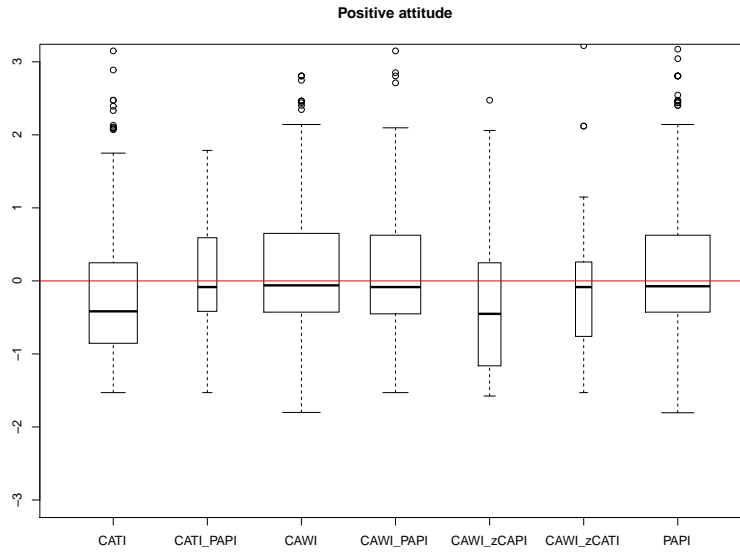


3.4 Satisfaction and Well-being

Satisfaction was measured by two variables, one referring to satisfaction with life, the second one with happiness. The correlation was strong and a common factor is fine to described them.

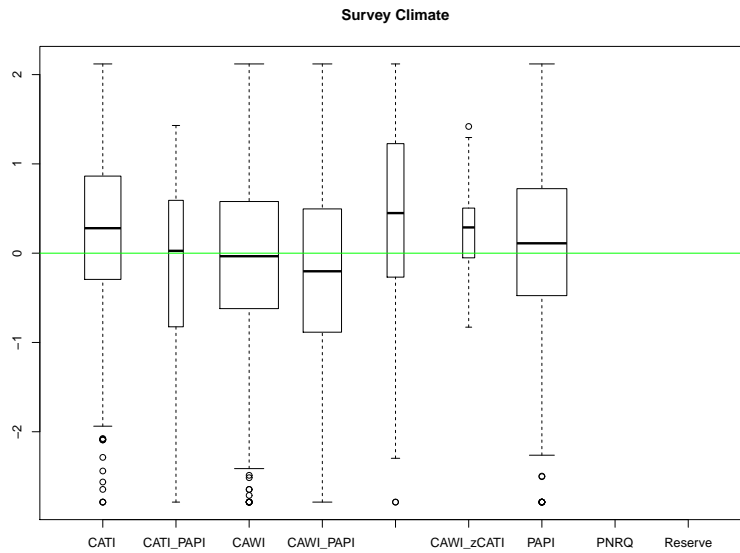


We can also consider more complex scales, some insisting on negative feelings (bad sleep, etc.) and another one considering positive attitude towards himself and the future (questions C29-C32). In the first case, there are not so many differences between conditions while the are much more important in the second.



3.5 Survey Climate

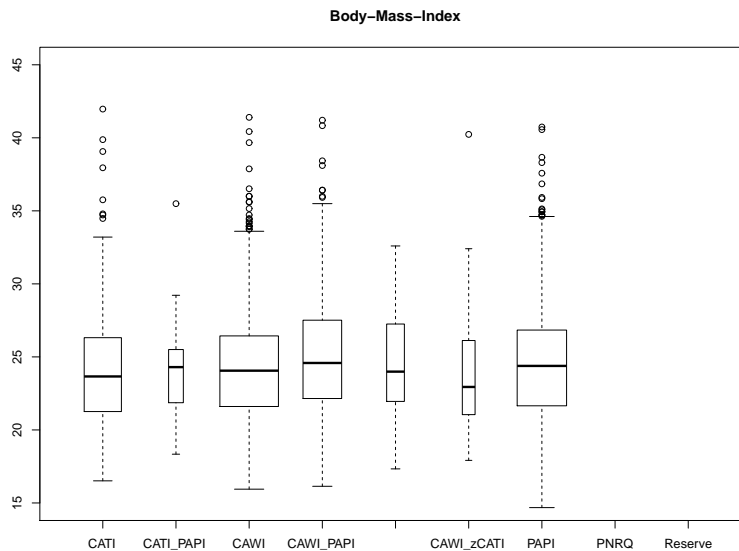
Four questions were dedicated to the survey climate: one asking about a possible intrusion in private life, a second one about the trust in the result, a third one about the interest to answer to such surveys and a fourth one dedicated to the interest of surveys for society. A principal component analysis showed that a single factor solution was satisfactory, showing the last three variables with a correlation over 0.8.



Once again, to consider a weighting schema based on age and sex does not change the results.

3.6 Social desirability

We can expect that a social desirability factor could be perceptible according the different modes. We have chosen different indicators in the questionnaire to begin with self-representation of her physical appearance: we can expect that age differences in the way to report weight and height, allowing to calculate a Body-Mass-Index (BMI).



In fact the estimation is not really different according the conditions of response and a simple analysis of variance show that the conditions explain less than one third of one percent of the variance!

It is not so different for the quantity of alcohol consumed, the only exception being that very high quantities (more than 40 "units" by weeks, that means more than 6 by day) are only reported in the absence of an interviewer (PAPI or CAWI).

4 First Conclusions

Of course such an analysis is rather descriptive and preliminary. It shows nevertheless at least three interesting results

- A difference in demographics according what we know, but even bigger than expected for the CATI
- Less significant differences than expected in the substantive results.
- Weighting with demographics like sex and age does not change the results in the case were we have tested this schema of correction.

Does it mean that methodology does not matter and results are not differing according to modes? Certainly not but that a careful design and a great attention to obtain different types of respondents can lead us to satisfying results in different conditions.

But once again we have to underline that these results are produced in a particular context, Switzerland, and a design more like the basic ISSP requirement than the longer ESS model.