# The effect of a lottery incentive on response, representativeness and data quality in the Dutch longitudinal Labour Force Survey[[1]](#footnote-1).

Annemieke Luiten

Statistics Netherlands

## Introduction

Statistics Netherlands (CBS) uses a mixed mode data collection strategy for the first wave of the Labour Force Survey. Sample households are invited to respond via a web questionnaire. Non-respondents are followed up by either telephone or face-to-face interviewers. Because of the higher costs of the interviewer modes, a high web response is desirable. To that end, CBS has introduced incentives in most social surveys, where the incentive is mostly a lottery of iPads (Luiten, 2006). From January to June 2017 10% of the LFS sample was entered in a lottery of iPads in the first wave of the LFS. When we were confident that the incentive had the desired effect of increasing the web response without notably effecting the LFS estimates, the ratio of incentive offers was gradually increased, until the entire sample was offered the incentive in December 2017. In this paper I present analyses of the effect of the incentive on first wave and later wave response, response representativeness, and data quality: the pace of filling in the questionnaire, and the number of missing items. Literature suggests that incentives have either no effects on data quality, or a positive effect, although a minority of studies also found negative effects.

Before we venture into the discussion of these issues, first a short description of the design of the Dutch LFS. The LFS uses a five wave rotating panel design in which households are interviewed in five consecutive quarters before rotating out of the sample. The survey is voluntary. Sample addresses are drawn from the Municipal Personal Records Database which contains personal details of everyone who lives in the Netherlands. The target population consists of people residing in the Netherlands, aged 15+ and living in private households. Only when all household members (15+) complete the questionnaire, the household is considered to have responded. Proxy answering is allowed.

The LFS has a sequential mixed-mode design in which web interviewing is followed by telephone or face-to-face interviewing. Whether a sample unit goes to CATI or CAPI depends on the availability of a telephone number and the size of the household: larger households are interviewed face-to-face. Non-respondents in CATI are not transferred to CAPI. Because CBS makes use of a rolling reference week, we can take our time for each mode: the web approach takes place in month t, the telephone approach in month t+1, and the face-to-face approach in month t+2. Labour force statistics for any month T are made from web data from month T, CATI data from the sample of month T-1, and CAPI data from the sample from month T-2.

Not all non-responding households of the web phase are followed up in the subsequent mode. In order to make optimal use of the cheaper web mode, and also to be able to keep the number of sample units that go to CAPI and CATI stable each month, a relatively large sample is drawn for the first web phase, of which a subsample is drawn for the follow-up phases. The subsample is stratified by interviewer region and known telephone, but is otherwise random. About half of the non-respondents in web is not followed up in other modes.

All sample units receive a letter containing the internet address of the web questionnaire and a personal login. All household members need to use the same login to gain access to the questionnaire: i.e. household members do not receive an individual login. The letter shows a prominent picture of the incentive: a number of iPads. Two reminders are sent to non-respondents two weeks and three to four weeks after the advance letter.

At the end of the first wave questionnaire, respondents are asked if they are willing to participate in the second wave of the LFS. If so, their telephone number is asked. The second wave is CATI for all sample units, regardless of the mode of the first wave. If no telephone number is given or otherwise available, respondents are excluded from the survey. The second wave takes place three months after the first. A new advance letter is sent for each new wave, but no more incentives are offered. The third to fifth waves are also telephone interviews. Households can skip one wave. If more waves are skipped, the household is removed from the sample.

## Results

### Response

Table 1 shows the wave 1 response rates for the three modes and the total wave 1 response. The data used for analysis come from three fieldwork months were there was about the same number of households with and without incentive. The total response in this table is low, as it is the response from the initial sample, not compensating for the subsampling of CATI and CAPI addresses, nor adjusting for known or unknown ineligibility. The incentive leads to a significant increase in response of almost three percentage points in web, but leads to an equally large decrease of response in CATI. In CAPI there is also a decrease in response, but this does not reach significance. Because the CAWI sample is far larger than that of the other two modes, the incentive still leads to a significantly higher overall response.

***Table 1. Wave 1 response in CAWI, CATI, CAPI and overall, by incentive.***



### Willingness to participate in further waves

Table 2 shows the percentage of households in the three modes that express a willingness to be contacted again, and provide their telephone number. The incentive leads to a substantial increase in the number of households who are recruited for wave 2 in CAWI.

***Table 2. Panel recruitment by mode and incentive condition.***

******

Table 3 shows results for the wave 2 and wave 3 response. The table shows that there is no effect at all of the incentive in the later waves. The data for waves 4 and 5 were not yet available at the time of writing, but in view of the findings for the earlier waves, differences are not to be expected.

***Table 3. Response rates by incentive in wave 2 and wave 3.***



It makes sense that an incentive that you did not win in the previous wave does not influence the decision to take part in wave 2 or 3. The lack of effect of the incentive means however, that the higher number of households that were recruited do not drop out in the second wave, for example because they were disappointed by not winning the incentive.

### Speed of responding

Households in the incentive condition respond sooner: 46% of the web response in the incentive condition is generated by the advance letter, against 41% in the control group (χ2(1) = 30.9, p < .001).

### Data quality

Three measures of data quality are discussed: the percentage of ‘don’t know’ answers in all questions answered, the percentage of refusals, and the pace of filling in the questionnaire, operationalized as the questionnaire duration divided by the number of questions answered.

Missing values in the Dutch LFS consist mostly of ‘don’t know’ answers, which makes sense in a survey where proxy answering is allowed. The missing values are expressed as a percentage of the number of questions answered. The percentage of refusals is very low in the three modes, and in neither mode is there an effect of the incentive, see table 5.

***Table 5. Percentage of refusals by mode and incentive***

******

The incentive does have an effect on the number of DK answers, see table 6. Both in CAWI and in CAPI, the number of DK answers is lower in the incentive condition. There is no effect in CATI.

***Table 6. Percentage of don’t know answers by mode and incentive***

****

### Pace of filling in the questionnaire

Respondents know that they will find out if they won an iPad immediately after filling in the questionnaire. This perspective may induce very eager respondents to speed through the questionnaire, with a risk of reduced data quality.

In CAWI a significant effect of incentive was found, indicating that questionnaire length was *longer* in the incentive condition: 15.2 minutes in the incentive condition versus 14.3 minutes in the control condition. The same pattern was found in the other modes. This is not surprising, as more substantive answers and less DK answers are given in the incentive condition.

## Conclusion, discussion, plans for the future

The lottery incentive has been successful in increasing wave 1 response, wave 2 recruitment and data quality. Not shown in this paper, but discussed extensively in Luiten and Groffen (2018), is that the incentive increased response representativeness as well. A short summary of findings:

* The lottery incentive leads to higher web response in wave 1, and higher overall response. The incentive does not affect response in the second and third wave. The effect could not yet be determined for the later waves, but we can safely assume that there will be no effect there as well.
* It is our interpretation that the incentive leads to a shift to the cheaper mode of people who would otherwise have been respondents in the other modes.
* The incentive leads to faster response, necessitating less reminders
* Not shown here, but interesting to mention: the incentive increases representativeness for several important variables, i.e., age and income. For some other variables no difference in representativeness was found. Over all variables, the design with incentives showed less variation in subgroup response propensities then the design without incentives. Subgroup variance in response propensities increases adjustment weights and therefore increases variance. If variance in the adjustment weights is decreased, sample size may be decreased as well, a third way in which costs can be reduced.
* The incentive leads a substantially higher number of households to agree to be contacted for wave 2. These households stay in the panel to the same extend as households without incentive. This indicates that the response gain in the first wave as a result of higher response and higher panel recruitment continues in the following waves. This is the fourth way in which the incentive saves costs.
* The incentive leads to higher data quality: especially the number of ‘don’t know’ answers diminishes significantly. The effect is strongest in CAWI.
* The fact that people in the incentive condition put more effort in the answering process is reflected in the speed of filling in the questionnaire: People in the incentive condition take a minute more time.

### Suggestions for alternative incentives

Statistics Netherlands has managed to find a modus vivendi to make lottery incentives work, contrary to many findings in literature. Even in LFS, a household survey, did the lottery increase response rates. The lottery incentive did not have an influence on the response propensities in the second and later waves, however. Incentive schemes can be designed with this purpose in mind: literature shows that relatively small unconditional incentives keep their effect for many waves. In order to consider alternative incentive schemes however, CBS should wait for the planned redesign of the LFS to a person sample. In the present design, with the large initial sample, no other incentive than a lottery can be economically viable.

An obvious alternative for the present lottery incentive is an unconditional incentive. CBS has extensive experience with unconditional €5 vouchers in other surveys. They have proven to substantially increase (web) response rates and can be economically viable in a design including CAPI, as is the LFS. The literature suggests that these kinds of incentives will also increase registration into the panel, and will keep their effect for several waves. The amount offered should depend on one other element: the timing of the introduction of the panel. Presently, respondents are not informed of the fact that they are in a panel: they are recruited for each next wave. It is to consider to recruit persons for the panel at the very beginning. This would probably lead to lower initial response, but presumably to higher wave response. In this scenario, an unconditional incentive with a higher value than the €5 that are presently used should be considered.

Unconditional incentives would be our preferred first choice for experimentation. Conditional incentives can be considered, as they are used in many panels, although the rewards offered in most are higher than what we would be prepared to pay. A conditional incentive that could be tried is offering one to people who stay in the panel for the entire five waves. But then again, that would entail mentioning that they are actually in a panel.

As the literature does not offer one clear cut panacea, a series of carefully designed experiments should determine which of these alternatives offers the optimal balance between costs, response, quality and attrition.

## References

Luiten, A. (2016). Incentives in official statistics. Paper presented at the 27th International Workshop on Household Survey Nonresponse, Oslo, Norway, September. Accessible at <https://view.officeapps.live.com/op/view.aspx?src=http://www.nonresponse.org/uploadi/editor/DnD148714898532LuitenIncentivesinOfficialStatistics.docx>

Luiten, A. and Groffen, D. (2018). The effect of a lottery incentive on response, representativeness and data quality in the Dutch LFS. Discussion paper, CBS (under review).

1. This work was supported by Eurostat Grant 07131.2017.003-2017.596 ‘Quality improvements for the Labour Force Survey’. This paper is an abstract of a paper that includes a literature review of incentives in cross sectional and longitudinal surveys. The longer paper also discusses subgroup response rates, representativeness and substantive variables (Luiten and Groffen, 2018). The views in this paper are those of the author, and do not necessarily reflect those of Statistics Netherlands. [↑](#footnote-ref-1)