7TH INTERNATIONAL WORKSHOP ON ADVANCES IN ADAPTIVE AND RESPONSIVE SURVEY DESIGN

Current Adaptive Survey Designs in Social Surveys at Statistics Netherlands

Kees van Berkel
Statistics Netherlands
December 2021

Abstract: Challenges that surveys are facing are increasing data collection costs and declining budgets. During the past years, many surveys at Statistics Netherlands were redesigned to reduce costs and to increase or maintain response rates. From 2018 onwards, adaptive survey design has been applied in several social surveys to produce more accurate statistics within the same budget.

In previous years, research has been done into the effect on quality and costs of reducing the use of interviewers in mixed-mode surveys starting with internet observation, followed by telephone or face-to-face observation of internet nonrespondents. Reducing follow-ups can be done in different ways. By using stratified selection of people eligible for follow-up, nonresponse bias may be reduced. The main decisions to be made are how to divide the population into strata and how to compute the allocation probabilities for face-to-face and telephone observation in the different strata.

Currently, adaptive survey design is an option in redesigns of social surveys at Statistics Netherlands. In 2018 it has been implemented in the Health Survey and the Public Opinion Survey, in 2019 in the Life Style Monitor and the Leisure Omnibus, in 2021 in the Labour Force Survey, and in 2022 it is planned for the Social Coherence Survey.

This paper elaborates on the development of the adaptive survey design for the Labour Force Survey. Attention is paid to the survey design, in particular the sampling design, the data collection constraints, the choice of the strata for the adaptive design, the calculation of follow-up fractions by mode of observation and stratum, the practical implementation of the adaptive design, and the six-month parallel design with corresponding response results.

Key words: balanced response, nonresponse bias, accuracy, data collection costs

1. Introduction

Adaptive survey design aims to get a better balanced response by putting different effort in different groups of the population. It is deployed in improving survey results, or reducing survey costs. The designs have attracted a lot of interest in recent years due to budgetary constraints and declining response rates, see Chun, Heeringa and Schouten (2018).

In the socio-economic social and economic statistics of Statistics Netherlands, adaptive survey design was first applied to the Health Survey in 2018. Its development and implementation is described in detail by Van Berkel, Van der Doef and Schouten (2020). Currently, adaptive designs are being introduced step-by-step in different surveys. Here the focus is on the most recently introduced adaptive design for the Labour Force Survey.

The paper reads as follows. Section 2 contains some methodological aspects of adaptive survey design. Section 3 is on implemented adaptive survey designs at Statistics Netherlands with elaboration for the Labour Force Survey. Finally, Section 4 provides some discussion points.

2. Methodology

In this section the four main elements of adaptive survey design are discussed: quality indicators, design features, clustering the population, and optimisation.

2.1 Quality indicators

Consider a finite population of N people, labelled by $k = 1, 2, \ldots, N$. For the survey, a probability sample with size n is drawn from the population, such that each person k has a positive inclusion probability π_k. Let a_k be the inclusion indicator for person k. This means that a_k is equal to 1 if person k is in the sample, and 0 if person k is not in the sample. The expected value of a_k is equal to the probability that person k is selected in the sample, $E(a_k) = \pi_k$.

A random response model is adopted, where each person k in the target population is assumed to have a response probability ρ_k, which is only known to person k. If person k is selected in the sample, this person is subjected to a Bernoulli experiment that results in response with probability ρ_k and in nonresponse with probability $1 - \rho_k$.

Let r_k be the response indicator for person k. So r_k is equal to 1 if person k responds and 0 if person k does not respond. The expected value of r_k is equal to the probability that person k responds, $E(r_k) = \rho_k$. The number of respondents r in the survey is a random variable $r = \sum_{k=1}^{N} a_k r_k$ with expected value $E(r) = \sum_{k=1}^{N} \pi_k \rho_k$.

The aim of the survey is the estimation of population means for several target variables. An estimator of the population mean $\bar{Y}$ of variable Y is the modified Horvitz-Thompson estimator,

$$\bar{Y}_{mHT} = \left(\sum_{k=1}^{N} a_k r_k Y_k / \pi_k\right) \Big/ \left(\sum_{k=1}^{N} a_k r_k / \pi_k\right).$$

This estimator is in general a biased estimator for the population mean $\bar{Y}$. If $\rho_k = \bar{\rho}$ for all k, where $\bar{\rho} = \frac{1}{N}\sum_{k=1}^{N} \rho_k$ is the population response mean, then $\bar{Y}_{mHT}$ is unbiased. However, this is generally not true. Bethlehem (1988) shows that

$$B(\bar{Y}_{mHT}) = E(\bar{Y}_{mHT}) - \bar{Y} \approx \frac{1}{\bar{\rho} N} \sum_{k=1}^{N} (\rho_k - \bar{\rho})\, Y_k = \frac{1}{\bar{\rho}}\, cov(\rho, Y) = \frac{R(\rho, Y)\, S_\rho\, S_Y}{\bar{\rho}}.$$

Here $cov(\rho, Y)$ is the population covariance between the response probabilities and the values of the target variable, $R(\rho, Y)$ is Pearson's correlation coefficient, S_ρ is the standard deviation of the response probabilities and S_Y is the standard deviation of the values of the target variable.

From this expression it follows that there is no bias if there is no correlation between response propensity and target variable. The smaller the variation in response probabilities or in the values of the target variable, the smaller the bias. And the higher the mean population response rate, the smaller the bias.

Since Pearson's correlation coefficient does not exceed 1 in absolute value, an upper limit for the bias can be given by:

$$|B(\bar{Y}_{mHT})| \leq \frac{S_\rho\, S_Y}{\bar{\rho}} = CV(\rho)\, S_Y\,.$$

Here $CV(\rho) = S_\rho / \bar{\rho}$ is the coefficient of variation of the response probabilities as a population quantity. Since S_Y is a population parameter that cannot be influenced, in the remainder of this paper it is attempted to minimise this upper limit for the bias by minimising $CV(\rho)$ through interfering in the process of data collection.

2.2 Design features

The focus in this paper is on the mix of survey modes. It is assumed that a mixed-mode design is used with Computer-Assisted Web Interviewing (CAWI) as starting mode. Follow-up of CAWI nonresponse is carried out by a combination of Computer-Assisted Personal Interviewing (CAPI) and Computer-Assisted Telephone Interviewing (CATI). The design features to adapt are the CAPI and CATI follow-ups.

In the sequential mixed-mode strategy CAWI → CATI/CAPI, all sampled people are first asked by letter to complete a questionnaire on the Internet. People who have not responded to this request after no more than two reminders, are contacted by telephone if a telephone number is known at the office, otherwise they are visited at home to conduct an interview. In ASD the entire sample starts with CAWI and the observation strategy of the follow-ups is adjusted as follows. To reduce the variation of response rates, more CATI/CAPI is used for groups that are less likely to respond via CAWI, and less CATI/CAPI is used for groups that are more likely to respond via CAWI. The identification of these so-called target groups is carried out using cluster analysis. Observe that the sequential strategies CAWI → CAPI and CAWI → CATI are special cases of the CAWI → CATI/CAPI strategy where, in the first case the CATI-follow-ups have been set to zero, and in the second no CAPI-follow-ups are conducted.

2.3 Clustering the population

Determining target groups is also called segmentation or clustering the population. The target groups are composed by means of response propensities of people per mode. This may mean that two target groups have approximately the same response rate at CAWI, but that their CATI or CAPI response rates differ. It is also possible that the total response rates of two target groups are approximately the same, but that their response rates differ per mode.

Clustering is performed in two steps. First a classification tree algorithm is applied, dividing people into groups based on personal characteristics. The algorithm uses the characteristics that explain the response most, and it divides each selected characteristic into categories. Second k-means clustering is applied with the selected characteristics and corresponding categories yielded by the classification tree. This is a method that divides data into groups which are homogeneous according to response probabilities, where outliers can be detected. The advantage of this method is that small groups with extremely high or low response rates can be identified as target groups. These target groups can be assigned a separate approach strategy. A disadvantage of the k-means method may be that the target groups are less homogeneous according to the characteristics used.

2.4 Optimisation

Consider a survey with the mixed-mode strategy CAWI → CATI/CAPI. Let G be a partition of the population, such that in each group $g \epsilon G$ the response probabilities do not differ much from one person to another. This partition is used to determine the target groups. Each target group is the union of one or more groups from G. For each $g \in G$, let $N(g)$ denote the population size of group g and $n(g)$ the sample size of group g. It is assumed that within each group g all people have the same inclusion probability $\pi(g)$.

The response mean $\bar{\rho}$ is estimated by the Horvitz-Thompson estimator

$$\bar{\rho}_{HT} = \frac{1}{N}\sum_{k=1}^{N} a_k r_k/\pi_k = \frac{1}{N}\sum_{g\in G} r(g)/\pi(g),$$

where $r(g) = \sum_{k=1}^{N(g)} a_k r_k$ is the estimator for the number of responses in group g. For each group $g \in G$, let $p_w(g)$ be the CAWI-response probability, $p_e(g)$ the probability of being eligible for follow-up, $p_t(g)$ the CATI-response probability and $p_p(g)$ the CAPI-response probability. Let $f_t(g)$ and $f_p(g)$ be

the fractions of CAWI-nonrespondents eligible for follow-up to be approached by telephone and face-to-face respectively in group g. So the response probability in group g equals

$$p(g) = p_w(g) + p_e(g)\left(f_t(g)p_t(g) + f_p(g)p_p(g)\right).$$

Since $r(g) = n(g)p(g)$, this allows the mean response probability and the population variance of the response probabilities to be estimated:

$$\bar{\rho} = \frac{1}{N}\sum_{g\in G} n(g)\,p(g)/\pi(g) \text{ and } S_\rho^2 = \frac{1}{N}\sum_{g\in G} N(g)\,(p(g)-\bar{\rho})^2 .$$

The following problem needs to be solved.

Minimise $CV(\rho) = S_\rho/\bar{\rho}$ under specified constraints.

Different types of constraints can be used. For instance constraints on budget, interviewer capacity, response numbers, response rates, or ratio of modes in response.

One CATI sampling fraction and one CAPI sampling fraction is used per target group. The decision variables for which the minimum can be found under the specified constraints, are the CAWI sample size n, the inclusion probabilities $\pi(g)$ and the CATI and CAPI sampling fractions $f_t(d)$ and $f_p(d)$ per target group d. Determining the partition G is a crucial part in designing the survey, and it is the starting point for clustering the population.

The minimisation problem requires a search for the numbers of people to be approached by target group and observation mode. The lower the CAWI response propensity of a target group, the more telephone and/or face-to-face observation is applied. This may lead to a smaller variation of response rates, and the ratio of the target groups in the response may be more similar to the ratio of the target groups in the population. This may be at the expense of the overall response rate.

The minimisation problem is solved with the Auglag function of the Alabama R package. On the Internet site https://CRAN.R-project.org/package=alabama, the reference manual and package source can be found. This R package uses the "Augmented Lagrangian Adaptive Barrier Minimisation Algorithm for optimising smooth nonlinear objective functions with constraints".

3. Adaptive survey design at Statistics Netherlands, 2018 -2020

3.1. An overview

At Statistics Netherlands adapted survey design was first applied in the production of official statistics for the Health Survey in 2018. For the Health Survey, people living in the Netherlands are sampled with equal probabilities and CAWI → CAPI is the observation strategy. To partition the population, demographic and regional characteristics were used that are known to explain response behaviour. Examples are ethnicity, age, income, urbanity, educational level, type of household, number of people in the household, place in the household, number of children, marital status, wealth, gender, and home ownership. Using empirical data from the Health Survey 2017, ethnicity, age, income and urbanity were found to be effective in creating target groups for the Health Survey. Further details can be found in Van Berkel, Van der Doef and Schouten (2020). In 2018 adapted survey design was also implemented for the Public Opinion Survey, where the target groups were composed by ethnicity and urbanity only.

In 2019 adapted survey design was extended to four surveys. The Life Style Monitor and the Leisure Omnibus were added. The target groups for all four surveys were defined by ethnicity, income and age with slight differences in the categories of these characteristics. Only the Life Style Monitor used urbanity additionally.

In July 2020, the adapted survey design started parallel to the regular survey for the Labour Force Survey, with the aim of moving to the adaptive design by January 2021.

3.2. Adaptive survey design for the Labour Force Survey

The Labour Force Survey (LFS) aims to provide statistics about participation of the Dutch population in the labour market. Core indicators are unemployment rate, participation rate, and job characteristics.

The survey applies a rotating panel design with five waves at three-monthly intervals. The observation strategy for the first wave is CAWI → CATI/CAPI, with different CATI- and CAPI-sampling fractions per target group. The observation strategy for the subsequent waves is CAWI → CATI.

The sampling design for the first wave is a stratified two-stage sample of people aged 14-89 years with unequal probabilities. Increasing the precision of unemployment figures, people registered at the Employee Insurance Agency (EIA) as a job seeker are overrepresented. Further, non-western migrants and 15 to 24 year olds are overrepresented. People aged 65 or over and 14 year olds are underrepresented. The relative inclusion probabilities are included in Table 1.

Table 1 Relative inclusion probabilities, first wave

EIA	ethnicity	age				
		14	15-24	25-64	65-74	75-89
no	non-western	1/4	3/4	3/4	1/4	1/8
no	western	1/4	3/4	1/2	1/4	1/8
yes	all	1	1	1	1	1

3.3 Development of adaptive survey design for the first wave.

Identifying target groups, and minimising the coefficient of response probabilities are carried out sequentially simultaneously. Constraints that have been taken into account are: (1) At most 50 percent of all CAWI nonrespondents may be followed up via CATI or CAPI, and (2) at most 40 percent of all follow-ups may be conducted via CAPI.

Target groups are defined by the same characteristics as those that are used in the sampling design: registered at the EIA as a job seeker (no, yes), ethnicity (non-western, western), and age (14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-89). With these characteristics, the population can be split up into 28 categories. After merging some near-empty ones, 24 remain.

Using 2019 LFS-pilot data, response probabilities per mode of observation and likelihoods of availability of a telephone number can be estimated for each of the 24 categories. Using Alabama in R, several of near-optimal follow-up strategies per category can be obtained numerically. For each of these strategies, categories were merged by similarity of their follow-up strategies. First by overall follow-up, then by follow-up per mode. To avoid ending up with small target groups, some groups were redistributed. This approach yielded different sets of target groups. Through elimination, the eight target groups in table 2 suited best.

Table 2 Target groups LFS

EIA	ethnicity	age							
		14	15-24	25-34	35-44	45-54	55-64	65-74	75-89
no	non-western	8	3	1	2	2	2	2	6
no	western	8	5	4	6	7	8	8	6
yes	non-western	3	3	1	3	2	2	7	7
yes	western	3	3	4	4	6	7	7	7

With these target groups, the coefficient of variation of response probabilities was minimised under the constraints mentioned above. For this purpose, the solver in R was applied with different random initial values for the CATI and CAPI follow-up fractions per target group, since the algorithm can stop at a local minimum. The optimal solution is the solution with the lowest coefficient of variation. Table 3 shows the follow-up fractions of the optimal solution, yielding a coefficient of variation of response probabilities of 0.06 and mean response rate of 44.3%. If randomly half of the CAWI nonrespondents eligible for follow-up would be selected for follow-up, of which randomly three fifth assigned to CATI and two fifth to CAPI, the CV would be 0.186 and the mean response rate 45.0%.

In Table 3, the columns p cawi, p cati, p capi, and p tot show the expected response rates for CAWI, CATI, CAPI, and the total adaptive strategy. The columns f cati, f capi and f tot represent the CATI, CAPI and total follow-up fractions as a percentage of the CAWI-nonrespondents eligible for follow-up.

Table 3 Response probabilities and selection fractions per target group LFS

group	p cawi	p cati	p capi	f cati	f capi	f tot	p tot
1	14.4	11.9	34.3	0.0	96.9	96.9	42.5
2	21.2	19.3	32.6	51.3	48.7	100.0	41.3
3	21.5	20.6	40.6	0.0	68.1	68.1	42.9
4	28.4	25.1	42.9	31.6	31.5	63.1	43.5
5	32.1	31.1	45.3	56.9	0.0	56.9	43.9
6	34.1	31.1	44.8	47.3	0.0	47.3	43.6
7	39.9	31.8	37.6	22.9	0.0	22.9	44.2
8	47.9	36.0	40.4	0.0	0.0	0.0	47.9
tot	34.0	28.1	37.4	30.0	20.0	50.0	44.3

Table 3 shows that in target groups 1 and 2, with the lowest CAWI response rates, almost everyone is followed up. In target group 3, also with a low CAWI response rate, 68.1% of CAWI-respondents eligible for follow-up are selected for CAPI-follow-up because of the relatively high CAPI response rate in this group. In target groups 5, 6, and 7, no CAPI-follow-up is used because a sufficiently high response can be obtained by CAWI and CATI. In group 8, no follow-up is needed at all due to relatively high CAWI response rate in this group.

3.4. Practical implementation of adaptive survey design for first wave of the Labour Force Survey

To determine the CAWI sample size for the first wave, it is required that the fifth wave will yield 4068 respondents every month. The conversion of this response target into a sample size for the first wave depends on a large number of parameters. The parameters for waves 1 and 2 were estimated based on the LFS pilot 2019-2020. For the subsequent waves, probabilities were estimated using empirical data from EU-SILC and the regular LFS. With these estimates, it was calculated that a sample of 3154 people should be approached weekly through CAWI.

In practice, there are some complications. Firstly, estimated response rates will differ from realised response fractions; secondly, fewer telephone numbers may be available than expected; thirdly, sample sizes for CATI and CAPI are fixed per month, due to scheduling of interviewer capacity. With estimates according to the adaptive design of the previous section, follow-up sample sizes are set at 616 and 410 people per week for CATI and CAPI, respectively. For CATI and CAPI, each month consists of follow-ups of four or five CAWI week samples. This means that in a month with follow-ups of four CAWI week samples, the CATI and CAPI samples contain 2460 and 1640 people, and in a month with follow-ups of five CAWI week samples these numbers are 3075 and 2050.

Next it is explained how the CATI and CAPI samples are drawn each month. CAWI non-respondents within the same target group are selected for follow-up with as close to equal probabilities as possible, sticking to the agreed upon CATI and CAPI sample sizes. A two-stage procedure is followed. In the first stage selections for CATI and CAPI by target group are made in accordance with the computed adapted survey design. In the second stage the selections are fitted to the agreed upon sample sizes per mode.

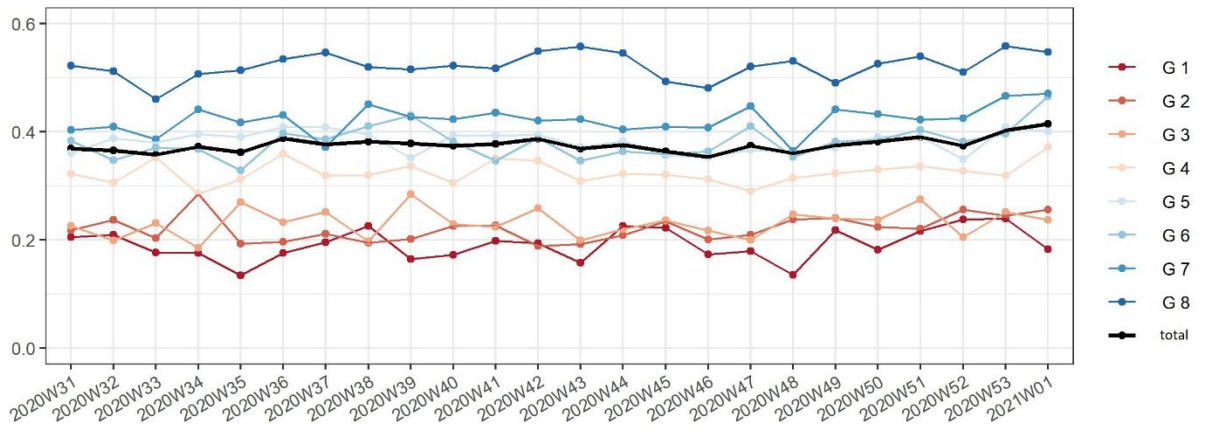
Stage 1, per target group: Split the eligible CAWI nonrespondents into CATI-eligible and CAPI-eligible people, where the groups are as close to proportional in size to the required selection fractions as possible. Non-respondents with 'best' available phone numbers are marked as CATI-eligible, the rest is CAPI-eligible. Select people from both groups with equal probabilities. If there is a lack of available telephone numbers, the CAWI-nonrespondents cannot be divided into two groups proportional to the CATI- and CAPI- follow-up fractions of the adaptive design. In this case, all CAWI- nonrespondents with telephone number are considered CATI-eligible, and selection fractions from both pools of eligible people are adjusted such that 1) the total follow-up fraction of the CAWI-nonrespondents is equal to the total follow-up fraction as included in the adaptive survey design, and if possible 2) the CATI-follow-up (as a percentage of all nonrespondents) is as close to the one given in the design. As a consequence, the CATI- and CAPI-selection fractions differ in this way.

Stage 2: Merge all selections for CATI follow-up, and merge all selections for CAPI follow-up. If both merged selections are smaller or greater than the agreed sizes, then both selections are amended by simple random sampling of the remaining eligible CAWI nonrespondents, or randomly removed. If exactly one merged selection is larger than the agreed size and the other smaller than the agreed size, then an attempt is made to transfer randomly selected elements from one to the other. After that, the previous case may occur and the same solution procedure is applied.

3.4. Starting up the adaptive survey design for the Labour Force Survey

The introduction of ASD was part of a major redesign of the Labour Force Survey. In order to be able to estimate a shift in results due to this redesign, both the new and old designs were to run in parallel in the fourth quarter of 2020. As CATI and CAPI are follow-up-modes, a start was made in July 2020 with CAWI. To this end, 24 weekly samples of 3154 people each were selected to complete the LFS questionnaire via CAWI. Figure 1 shows the weekly CAWI response rates for the separate target groups, and for the total sample.

Figure 1 CAWI response rates by target group and week



The CAWI response rates per target group over the entire period of 24 weeks are shown in Table 4. In each target group the realised response rate (r) is greater than the estimated response rate (e). The

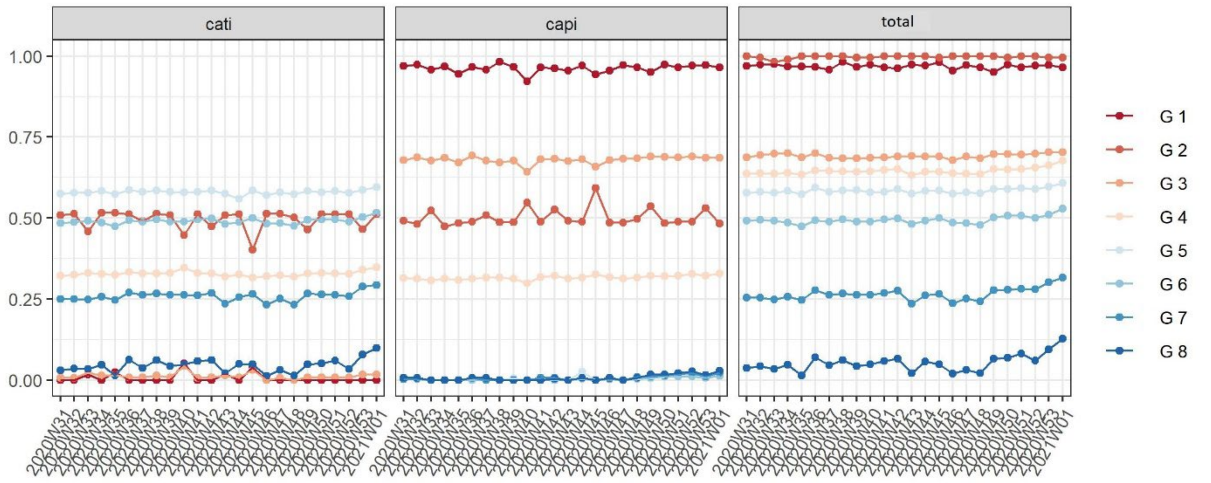
measures taken against the Covid-19 pandemic could be a reason for this. The difference (r - e) is largest in group 6 (6.1 percentage points) and smallest in group 2 (0.9 percentage points). The relative difference (r – e)/e is largest in group 1 and smallest in group 2. The overall realised CAWI response rate is 3.5 percentage points greater than estimated.

Table 4 Expected (e) and realised (r) CAWI response rates per target group

group	e	r	r - e	100 × (r – e)/e
1	14.4	19.2	4.8	33.3
2	21.2	22.1	0.9	4.2
3	21.5	23.2	1.7	7.9
4	28.4	32.5	4.1	14.4
5	32.1	38.2	6.1	19.0
6	34.1	37.1	4.0	11.7
7	39.9	42.2	2.3	5.8
8	47.9	52.2	4.3	9.0
tot	34.0	37.5	3.5	10.3

Figure 2 contains the weekly CATI- and CAPI- sampling fractions per target group. As established by design, CAPI is mainly applied in target groups with a low number and CATI in target groups with a high number. Due to a lack of telephone numbers in some weeks in group 2, shifts from CATI to CAPI can be seen there. The selection fractions for target groups 4 to 8 show an upward trend in the later portions. This is caused by the increase in CAWI response, which left capacity for random addition after selection with the fractions from the sampling design.

Figure 2 CATI-, CAPI-, and total sampling fractions by target group and week



Due to the COVID-19 pandemic in 2020, some CAPI samples were only partially or not at all observed. Therefore, both the CAPI response rates and the effect of the adaptive design on the results of the LFS are difficult to determine.

4. Discussion

This paper describes the introduction of adaptive survey design at Statistics Netherlands. The focus is on mixed-mode designs starting with Internet observation and follow-ups with interviewer modes. The coefficient of variation of response probabilities propensities was taken as the objective function in optimising the designs. There are a few limitations in this approach:

1. It has not been taken into account that bias can be caused by non-random response.
2. The possibility of mode-specific measurement bias was ignored.
3. The allocation of follow-up is posed as a yes-no decision. An alternative is to vary the effort of interviewers by proposing different numbers of contact attempts for the different target groups.
4. Explanation of response and strata were based on administrative variables that are used in post-survey adjustments. A general question is to what extent stratification should be survey-specific and to what extent a subset of general strata will always be imposed. This is especially important for regional variables as they affect interviewer workloads over multiple surveys.

Finally, some questions for discussion.

1. How to deal with the above-mentioned limitations?
2. The introduction of a new design may lead to different survey estimates. How should trend changes be communicated to the users?
3. How to monitor and adjust during data collection?
4. How often should target groups and follow-up fractions be adjusted?
5. Should the post-survey adjustment be redesigned.

5. References

Berkel, C. van, Doef, S. van der, and Schouten, B. (2020): Implementing Adaptive Survey Design with an Application to the Dutch Health Survey, *Journal of Official Statistics*, Vol. 36, No. 3, 2020, pp. 609–629.

Bethlehem, J.G. (1988): Reduction of Nonresponse Bias Through Regression Estimation, *Journal of Official Statistics*, Vol. 4. No. 3, 1988, pp. 251 – 260.

Chun, A.Y., Heeringa, S.G., Schouten, B. (2018): Responsive and adaptive design for survey optimization, *Journal of Official Statistics*, Vol. 34, No. 3, 2018, pp. 581 – 597.