Implementing Adaptive Survey Design with an Application to the Dutch Health Survey

Kees van Berkel[1], Suzanne van der Doef[1], and Barry Schouten[2]

Adaptive survey design has attracted great interest in recent years, but the number of case studies describing actual implementation is still thin. Reasons for this may be the gap between survey methodology and data collection, practical complications in differentiating effort across sample units and lack of flexibility of survey case management systems. Currently, adaptive survey design is a standard option in redesigns of person and household surveys at Statistics Netherlands and it has been implemented for the Dutch Health survey in 2018. In this article, the implementation of static adaptive survey designs is described and motivated with a focus on practical feasibility.

1. Introduction

Adaptive survey design assumes that differentiation of effort over relevant population subgroups is either effective in improving survey quality or efficient in reducing survey costs. The designs have received a lot of interest over the last decade in response to budget pressure due to gradual but persistent declines of response rates, for example Chun et al. (2018). Despite the interest in the designs, the number of implemented case studies described in the literature is still relatively small. This article discusses implementation of adaptive survey designs at Statistics Netherlands as it has been initiated in 2016.

National Statistical Offices have the task of publishing reliable and coherent statistical information that responds to the needs of society. In order to maintain a good balance between quality, efficiency and cost-effectiveness, continuous evaluation and improvement of processes and working methods are necessary. In 2016, four data collection policy decisions were made at Statistics Netherlands in order to arrive at a more efficient data collection strategy: incentives were used to increase overall response rates, a second supplier of telephone numbers was deployed so that more telephone observation is possible, follow-up sample sizes for CATI and CAPI were fixed in order to stabilise interviewer workload, and adaptive survey design became a standard design choice in sequential mixed-mode surveys. These four changes were implemented to varying degrees for a large number of surveys since 2017. Here, the fourth change is discussed.

Adaptive and responsive survey designs exist since the early days of surveys, but until about 15 years ago had never been named explicitly nor been the subject of structured and formalised research. Groves and Heeringa (2006), Wagner (2008) and Luiten and

---

[1] Statistics Netherlands, Division of Data Services, Research and Innovation, P.O. Box 4481, 6401 CZ, Heerlen, the Netherlands.
[2] Statistics Netherlands, Division of Data Services, Research and Innovation, P.O. Box 24500, 2490 HA, Den Haag, the Netherlands.

Schouten (2013) give early accounts of such designs. So why did it take so long to implement the designs in practice? At least three reasons can be given. First, design changes may have an impact on the survey results. Second, the designs require very flexible case management systems and monitoring. Third, practical and logistical constraints hamper a translation of optimal designs to implemented designs. In this article, a case study is presented that has been implemented and in which practical considerations played an important role.

Adaptive survey designs have four main elements: quality and cost objectives, design features, stratification of the target population, and an optimisation and implementation strategy. See Schouten et al. (2017) for a general overview and Tourangeau et al. (2017) for a discussion. Here, the focus is on the fourth element, the optimisation and implementation strategy, within the context of mixed-mode surveys.

Optimisation approaches range from trial-and-error to case prioritisation to advanced mathematical programming. Two approaches are confronted with each other in the case study: case prioritisation, for example Peytchev et al. (2010), Wagner (2013) and Wagner and Hubbard (2013), and mathematical optimisation, for example see Schouten et al. (2013) and Kaputa and Thompson (2017). For practical reasons, it was decided that case prioritisation should be implemented, but realised results are compared to expected results of optimised designs.

This article reads as follows: Section 2 describes the methodology behind the adaptive survey design. Section 3 discusses the application to the Dutch Health Survey. Section 4 ends with a discussion of results, limitations and future activities.

2. Methodology

In this section, the four main elements of adaptive survey design, quality and cost criteria, design features, stratification and optimisation, are discussed. This is done from an operational perspective.

2.1. Quality Indicators

The adaptive survey design is focussed on optimising balance of response through the coefficient of variation (CV) of response propensities for relevant population subgroups. See Schouten et al. (2009), De Heij et al. (2015), and Moore et al. (2018).

The CV is based on the desire to limit the risk of nonresponse bias over a range of variables. A random response model is adopted and it is assumed that each population unit has a response probability and response of the unit is independent of other population units. The response probabilities are unknown and are replaced by estimated response propensities based on a set of relevant auxiliary variables.

Consider a sample survey with a target population of N people. Let persons be labelled by k in the target population, $k = 1, 2, \ldots, N$. For the survey, a single random sample without replacement with size n is drawn from the target population. 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, $E(a_k) = n/N$. 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, belonging to the corresponding Bernoulli experiment. 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 sample survey is a random variable $r = \sum_{k=1}^{N} a_k r_k$ with expected value $n\bar{\rho}$, where $\bar{\rho}$ denotes the mean response probability in the population.

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 response mean,

$$\bar{y} = \frac{1}{r} \sum_{k=1}^{N} a_k r_k Y_k.$$

The response mean $\bar{y}$ is in general a biased estimator for the population mean $\bar{Y}$. If $\rho_k = \bar{\rho}$ for all k, then $\bar{y}$ is unbiased, but this is generally not true. Bethlehem (1988) shows that

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

Here $cov(\rho, Y)$ is the population covariance between the response probabilities and the values of the target variable. Thus, there is no bias if there is no correlation between response propensity and the target variable. Introduce Pearson's correlation coefficient:

$$R(\rho, Y) = \frac{cov(\rho, Y)}{S_\rho S_Y},$$

where S_ρ is the standard deviation of the response probabilities and S_Y is the standard deviation of the values of the target variable. Then the bias approximation formula can be written as

$$B(\bar{y}) \approx \frac{R(\rho, Y) S_\rho S_Y}{\bar{\rho}}.$$

From this expression it follows:

1. $B(\bar{y}) = 0$ if there is no linear relationship between ρ and Y.
2. The stronger the linear relationship between ρ and Y, the larger $B(\bar{y})$.
3. $B(\bar{y}) = 0$ if there is no variation of response rates or no variation in the values of the target variable.
4. The smaller the variation of response rates, the smaller $B(\bar{y})$.
5. The smaller the variation in the values of the target variable, the smaller $B(\bar{y})$.
6. The greater the mean response rate, the smaller $B(\bar{y})$.

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

$$|B(\bar{y})| \leq \frac{S_\rho S_y}{\bar{\rho}} = CV(\rho) S_Y.$$

Here, $CV(\rho)$ denotes the coefficient of variation of the response probabilities. A lower CV for response propensities defined by auxiliary variables implies smaller nonresponse

biases on these variables. When auxiliary variables are associated with survey variables, then a lower *CV* also implies smaller nonresponse biases on these variables before weighting adjustment. However, a lower *CV* does not necessarily imply smaller nonresponse biases on these variables after weighting adjustment. Nevertheless, there are three reasons to still pursue a lower *CV*. The first reason is that a more balanced response leads to less variation in adjustment weights and, as a consequence, to a more efficient sampling design. The second reason is that Schouten et al. (2016) have shown that empirically nonresponse bias on survey variables is on average still smaller for more balanced response, even after weighting adjustment using the same auxiliary variables. This finding conforms to the intuition that a more balanced data collection is a sign of a more effective design, in general. The final reason is that balancing response forces survey designers to come up with strategies to raise response rates of the strata that are harder to contact and to get participation. In the remainder of this article, an attempt is made to minimise $CV(\rho)$ by interfering in the process of data collection.

2.2. Design Features

The focus in this article is on the mix of survey modes. It is assumed that a sequential mixed-mode design is used with CAWI (Computer-Assisted Web Interviewing) as the starting mode. Follow-up of CAWI nonresponse is done through interviewer modes. Here, it is assumed that the follow-up is done by CAPI (Computer-Assisted Personal Interviewing). The design feature to adapt is the CAPI follow-up.

In the sequential mode strategy, all sampled people are first asked by letter to participate in the survey by completing a questionnaire on the internet. People who have not responded to this request after no more than two reminders are visited at home to conduct an interview. The observation strategy of the face-to-face interviews is adjusted as follows. To reduce the variation of response rates, more CAPI is used for groups that respond badly via the internet than for groups that respond well. However, the entire sample starts with CAWI. The identification of these target groups is carried out using cluster analysis.

It is assumed in this article that the answers obtained are the same in different observation modes, that is, mode-specific measurement bias is absent and can be ignored. This is a simplification, as such biases are conjectured to exist and should then be incorporated in the design decisions within the adaptive survey design. Such inclusion of measurement biases is not straightforward as they are confounded with selection biases, unless experimental designs are used to disentangle the biases. The discussion section returns to this complication.

2.3. Stratification of the Target Population

Determining target groups is also called segmentation or clustering of the target 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 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 with a classification tree algorithm. People are divided into groups based on personal characteristics. The algorithm divides the groups that differ most in response behaviour first. To ensure that reliable response rates per mode can be estimated for each target group, it is important that the target groups are not too small. To prevent this, a minimum size per target group can be set.

2.4. Optimisation

Two approaches to optimisation are explored: case prioritisation and mathematical optimisation, including expected yield of the face-to-face follow-up.

2.4.1. The Optimisation Problem

Let G be the set of groups 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. For a simple random sample of size n, it is assumed that the size of the sample in group g equals $n(g) = n \cdot N(g)/N$.

Furthermore, for each group $g \in G$ it is assumed that all people have the same CAWI response probability $p_w(g)$, the same probability $p_e(g)$ of being eligible for face-to-face follow-up and the same CAPI response probability $p_p(g)$ in the face-to-face approached sample of group g. Let $f_p(g)$ be the CAPI sampling fraction in group g, that is the proportion of people to be approached face-to-face in the CAWI nonrespondents who are eligible for face-to-face follow-up in group g. The total response probability in group g equals

$$p(g) = p_w(g) + p_e(g)f_p(g)p_p(g).$$

This allows the mean response probability and the population variance of the response probabilities to be estimated:

$$\bar{p} = \frac{1}{N}\sum_{g \in G} N(g)p(g) \text{ and } S_p^2 = \frac{1}{N}\sum_{g \in G} N(g)(p(g) - \bar{p})^2.$$

The following problem needs to be solved.

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

Different types of constraints can be used:

- *Budget*. This can be done at different levels, such as an available budget for the total observation or per observation mode.
- *Capacity*. An upper limit can be specified for the sample size to be approached face-to-face. This can be at national or regional level.
- *Precision*. This concerns requirements for the number of respondents or the number of respondents per subpopulation.
- *Response rates*. For example, a minimum response rate, or minimum response rates per mode or per subpopulation.
- *Ratio of the CAWI/CAPI modes in the response*. For example, a minimum percentage of CAPI response in the total response, or minimal CAPI sampling fractions per target group.

One CAPI sampling fraction is used per target group. This leads to the extra constraint:

For each target group d and all groups $g_1,\ g_2 \subset d :\ f_p(g_1) = f_p(g_2)$ applies.

The decision variables for which the minimum can be found, are the CAWI sample size n and the CAPI sampling fractions $f_p(d)$ per target group d.

The optimisation 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 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, however, be at the expense of the overall response rate.

2.4.2.   Optimisation Approaches

Two approaches are elaborated: case prioritisation and mathematical optimisation.

Case prioritisation is based on the rationale that the weakest performing population subgroups need the most attention and need to be allocated first. Response propensities at the end of a data collection phase, in this article CAWI, are estimated and sorted in increasing order. The sample units or sample strata with the lowest propensities are re-approached until budget is depleted and/or other constraints are met. Case prioritisation does not guarantee that the coefficient of variation is actually decreased, since expected conditional response propensities in subsequent data collection phases are not included. Such conditional propensities may have an opposite order of size and may even deteriorate balance. Such opposite ranking is, however, unusual in practice.

Mathematical programming accounts for expected yield in follow-up data collection phases, as it includes follow-up response propensities. As such, it guarantees improvement under the condition that response propensities are estimated accurately. Here, the minimisation problem is solved with the Auglag function of the Alabama R package. This R package uses the "Augmented Lagrangian Adaptive Barrier Minimisation Algorithm for optimising smooth nonlinear objective functions with constraints". The optimisation problem of Subsection 2.4 is smooth and nonlinear, because the partial derivatives of the objective function, the coefficient of variation of the response probabilities exist, and the objective function is nonlinear. The problem is also solved with the solver in Excel. This solver uses the GRG nonlinear solver method to solve the nonlinear problem and this algorithm uses the generalised reduced gradient method. Because it is a nonlinear problem and the algorithm can end up in a local minimum, different random starting points were used and the best solution was selected.

3.   Application of Adaptive Survey Design to the Dutch Health Survey

3.1.   The Dutch Health Survey

The aim of the Dutch Health Survey is to provide as complete an overview as possible of developments in health, medical contacts, lifestyle and preventive behaviour of the population in the Netherlands. The target population consists of all people living in the Netherlands who do not belong to the institutional population. The sample is a stratified

two stage sample in which people with equal probabilities are selected. This sampling design is approximately the same as the simple random sampling design. The observation starts with CAWI and the re-approach mode is CAPI. As a response increasing measure, iPads are raffled among the sampled people.

3.2. Stratification

The classification tree algorithm is implemented in R with the rpart package. Demographic and regional characteristics have been used that are known to have a different response distribution than the population. Examples are ethnicity, ethnicity of parents, age, income, urbanity of the municipality, urbanity of the neighbourhood, living in the four largest cities, 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. For more details on the characteristics used, see Section 5, Appendix. The algorithm determines which characteristics are used to split the groups and in which order. For categorical variables, the algorithm also determines where to split. This ensures that, for example, for a variable such as age, a classification can be made that best matches the response behaviour.

The results of the classification tree algorithm are the characteristics used for the Health Survey to record the target groups: ethnicity (NL resident, western migrant, non-western migrant), age (in years), income (in quintiles) and urbanity of the municipality in which the person lives (very strongly urban, strongly urban, moderately urban, few urban, and non-urban). The algorithm ensures that the characteristics are merged into larger groups. Ethnicity is divided into two groups, namely western (NL residents and western migrants) and non-western (non-western migrants). Age is divided into four categories: 0–11, 12–24, 25–64, and 65+. The income used is the standardised household income and the classification is into two categories, with the low income category consisting of the lowest 20% and the high income category consisting of the remaining 80%. Urbanity is reduced to two categories, namely very strongly urban and all others. Figure 1 shows the classification tree. The tree is read from top to bottom. In each node a division is made
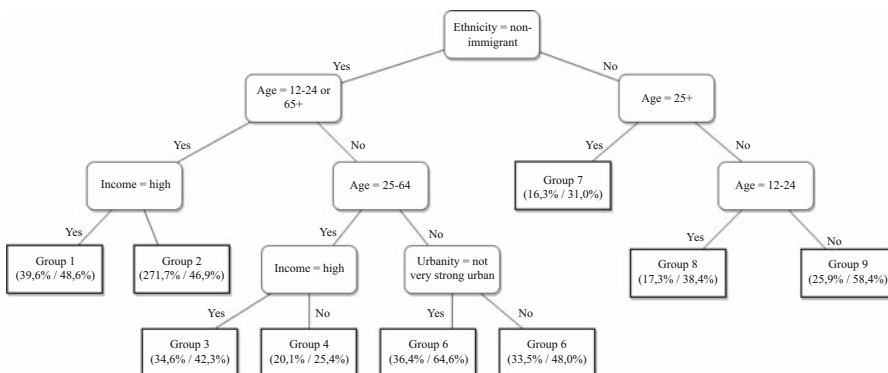

Fig. 1. *Classification tree based on results of the Health Survey 2016.*

target group	age	income	urbanity
1	12–24, 65+	high	-
2	12–24, 65+	low	-
3	25–64	high	-
4	25–64	low	-
5	0–11	-	2–5
6	0–11	-	1

Fig. 2.   Partition of NL residents and western migrants into target groups.

based on a characteristic and the group is split. At the bottom of the tree, the ultimate target groups can be found, together with the response rates of CAWI and CAPI.

The first six target groups partition the NL residents and western migrants. Figure 2 contains an overview of these target groups. Here, urbanity = 1 means very strongly urban and urbanity = 2–5 means the union of the remaining categories. A dash means that there is no restriction for the variable in question.

The non-western migrants are divided into three target groups by age: 25+ in target group 7; 12–24 in target group 8 and 0–11 in target group 9.

3.3.   The Dutch Health Survey Optimisation Problem

The set G of groups used to determine the target groups consists of 32 groups: ethnicity(2) × age(4) × income(2) × urbanity(2). The coefficient of variation of response probabilities $CV(\rho) = S_\rho / \bar{\rho}$ is estimated as described in Subsection 2.4.1. Minimising $CV(\rho)$ is carried out under the constraints:

- $n \leq n_{max}$ {CAWI sample size does not exceed n_{max}},
- $n \cdot \bar{\rho} \geq R$ {expected response size is at least R},
- $\sum_{g \in G} p_e(g) f_p(g) n(g) \leq C$ {total CAPI sample size is at most C },
- For each target group d and all groups $g_1, g_2 \subset d : f_p(g_1) = f_p(g_2)$ applies {one CAPI sampling fraction per target group}.

Here n_{max}, R and C are constants to be filled in. The parameters with which the minimum can be found are the CAWI sample size n and the CAPI sampling fractions for face-to-face observation $f_p(d)$ in the target groups d. Note that it follows from the first two constraints that $\bar{\rho} \geq R / n_{max}$.

In the case of the Health Survey 2018, the target groups with corresponding response rates per mode and probabilities of re-approachable CAWI nonresponse have been determined with data from the results of the Health Survey in January-June of 2017. The maximum CAWI sample size n_{max} has been set to 18,000 people. To be quite sure that 9,500 responses are achieved, the expected response size R has been set to 9,631 people. The maximum CAPI sample size C is 8,039 addresses, based on the available CAPI budget. The mean response rate must therefore be at least 9,631 / 18,000 = 53.5%.

3.4.   Mathematical Optimisation

First, the mathematical optimisation approach is explored, as this approach may be used as a benchmark to the case prioritisation approach.

The minimisation problem is solved with the solver in R, with different random starting values for the CAWI sample size n and the CAPI sampling fractions f_p per target group, because the problem is nonlinear, allowing the algorithm to stop in a local minimum. The optimal solution is the solution with the lowest coefficient of variation. In 100 hours the algorithm found 11 solutions. The coefficients of variation of the different solutions are between 0.1123 and 0.1217, except for one solution that had a coefficient of variation of 0.21. This solution is not considered further. It cannot be guaranteed that 0.1123 is the overall minimum of the coefficient of variation.

The remaining 10 solutions have almost the same coefficients of variation, but with different sampling fractions for face-to-face observation per target group. Table 1 shows the minimum, maximum and mean CAPI sampling fractions per target group of the 10 solutions. It is assumed that in each target group, 3% of the CAWI sample is not eligible for follow-up. It is striking that in each solution, the target groups 4 and 7 are entirely re-approached face-to-face. Also in target group 8, the CAPI sampling fraction is always relatively high.

Table 2 contains two solutions with approximately the same coefficient of variation, but with different CAPI sampling fractions per target group. The differences are largest for target groups 8, 9 and 3.

Not all solutions use the maximum allowable number of people to be approached face-to-face. For the solution with the least use of CAPI, 7,406 people are approached face-to-face with a coefficient of variation of 0.1123. The solution with the most use of CAPI, 8,039 people are approached face-to-face with a coefficient of variation of 0.1217. This solution was ultimately chosen because the variation coefficients hardly differ from each other, but the use of face-to-face observation is fully utilised.

With the adaptive survey design, the mean response rate decreases compared to a sequential CAWI → CAPI design, in which all CAWI nonrespondents eligible for follow-up are visited at home. However, both the standard deviation and the variation coefficient of the response probabilities are smaller for adaptive survey design. The CAWI part in the response increases: in the new design, approximately 64% of the response will be realised with CAWI, compared to 58% in the current design.

Table 3 shows the results of the chosen solution. The column n CAWI contains the CAWI sample size, the column r CAWI the expected number of CAWI respondents and p

Table 1. *Minimum, maximum and mean CAPI sampling fractions per target group.*

Stratum	Min	Max	Mean
	%		
1	49	57	53
2	66	88	80
3	66	76	70
4	100	100	100
5	42	47	44
6	50	77	68
7	100	100	100
8	81	100	93
9	63	91	71

Table 2.   CAPI sampling fractions per target group for two different solutions.

Stratum	Solution 1	Solution 2
	%	
1	49	57
2	77	83
3	67	76
4	100	100
5	42	47
6	77	75
7	100	100
8	81	100
9	91	72

CAWI shows the expected response rate for CAWI. Column n *elig* shows the number of CAWI nonrespondents eligible for face- to-face follow-up. Columns n CAPI, f CAPI, r CAPI and p CAPI represent the CAPI sample size, the CAPI sampling fraction n CAPI / n elig, the expected number of CAPI respondents and the expected CAPI response rate. The columns r tot and p tot indicate the total number of expected responses and the total response rates per target group. These response rates have been estimated with results of the Health Survey, January-June 2017, with an adjustment to the CAWI response rates due to the raffle of iPads among the sampled people.

Table 4 shows quality measures, in which the situations without and with adaptive survey design are compared. This table shows that the use of adaptive survey design causes the overall response rate to decrease, but the variation of the response rates is improving and the ultimate quality measure $CV(\rho)$ is also improving. This is an indication for less bias due to selective nonresponse.

3.5.   Case Prioritisation

Case prioritisation employs the same nine strata and sorts the strata after the CAWI phase by estimated response propensities. One practical complication is added. The Netherlands is divided into ten interviewer regions, each of which contains about one-tenth of the population. Each interviewer region employs 10 to 15 interviewers. Since 2016, CAPI sample numbers per month, survey and interview region have been fixed in advance. The advantage of this is that the required CAPI interview capacity can be planned easier. A disadvantage is that a possible decrease of CAWI response can no longer be compensated by increasing the number of face-to-face re-approaches. In regions where relatively many people belong to target groups that need to be re-approached via CAPI, the fixed number of face-to-face interviews is relatively large.

At the end of the CAWI observation, a sample is drawn from CAWI nonrespondents eligible for follow- up, with agreed sizes per interview region. Prior to this, a priority is defined: in order of the realised CAWI response rate in the sample portion concerned, the target group with the lowest response rate receives the highest priority and the target group with the highest response rate receives the lowest priority. The CAPI potential is then sorted per region by priority, where all elements from one target group are given the same

Table 3. Results of adaptive survey design.

Stratum	n CAWI	r CAWI	p CAWI %	n elig	n CAPI	f CAPI %	r CAPI	p CAPI %	r tot	p tot %
1	4,542	1,936	42.6	2,470	1,472	59.6	714	48.5	2,650	58.3
2	785	194	24.8	567	508	89.5	244	48.0	438	55.8
3	7,361	2,765	37.6	4,375	3,480	79.5	1,473	42.3	4,239	57.6
4	727	168	23.1	538	538	100.0	137	25.5	305	41.9
5	1,472	580	39.4	848	410	48.4	265	64.6	845	57.4
6	332	121	36.5	201	149	74.3	72	48.0	193	58.1
7	1,274	245	19.3	991	991	100.0	304	30.7	550	43.1
8	411	83	20.3	315	315	100.0	122	38.6	205	49.9
9	363	105	28.9	247	176	71.4	103	58.3	208	57.3
Total	17,268	6,198	35.9	10,551	8,039	76.2	3,433	42.7	9,631	55.8

Table 4. Quality indicators for adaptive survey design.

Adaptive survey design	$\bar{\rho}$	S_ρ	$CV(\rho)$
		%	
No	63.6	10.1	15.8
Yes	55.8	6.8	12.2

priority. The CAPI potential with the highest priority is selected, then the CAPI potential with the second highest priority is selected and so on until the fixed size per region is reached. If one target group has to be partially selected in order to reach exactly the fixed size, a systematic sample is drawn from this. For that purpose, the CAPI potential with the priority in question of this target group is first sorted by postal code, house number, house letter, addition and designation. The other target groups with the lowest priorities are not observed face-to-face at all in the region concerned.

Because a new prioritisation is made every month, the above approach implies that the CAPI sampling fractions may differ per target group and region on a monthly basis. Fluctuations depend on the CAWI response. It is possible that the fixed CAPI size in one region has already been reached at a certain priority, while in the other region people with a lower priority still have to be selected to reach the fixed CAPI size. The fixed CAPI size per region has been determined in such a way that it is expected that in each region the same target groups will be approached face-to-face.

Statistics Netherlands ultimately decided for a case prioritisation approach based on its practical simplicity and ease in handling interviewer region workloads. A full mathematical optimisation would yield allocation probabilities that, in expectation, lead to the right workloads. However, in practice we are dealing with sample variation across different months, since contact and participation of sample units cannot be controlled. Therefore, per month a subsampling is performed that exactly matches the available workloads.

Table 5 presents the realised allocations, based on six months of the Dutch Health Survey, from January-June 2018. In this period, the realised mean response rate $\bar{\rho}$ equals 57.0%, the standard deviation of response probabilities S_ρ equals 7.6% and the realised coefficient of variation of response probabilities $CV(\rho)$ equals 13.3%. Compared with Table 4, these values are between the corresponding values of the proposed adaptive survey design and the design without adaptation.

Comparison of Tables 3 and 5 shows that most of the realised CAPI sampling fractions differ from the proposed CAPI sampling fractions. This is not only caused by the method of CAPI selection, but also by differences between estimates and realisations of CAWI sample sizes, CAWI response rates and proportions of CAWI nonrespondents eligible for face-to-face follow-up per target group. The largest differences between f CAPI in Tables 3 and 5 can be found in ascending order in target groups 6, 1, 9, 3 and 5. In target groups 7 and 8 f CAPI equals 100% as estimated. Overall, the realisation of f CAPI is 0.9 percent points larger than estimated.

In all target groups, the realised CAWI response rate is higher than expected, except in target group 7. The largest differences between p CAWI in Tables 3 and 5 can be found in ascending order in target groups 6 and 2. The overall CAWI response rate is 2.7 percent points larger than estimated.

Table 5. Realisations of adaptive survey design, Health Survey January–June 2018.

Stratum	n CAWI	r CAWI	p CAWI	n elig	n CAPI	f CAPI	r CAPI	p CAPI	r tot	p tot
			%			%		%		%
1	2,441	1,079	44.2	1,250	335	26.8	157	46.9	1,236	50.6
2	305	101	33.1	200	179	89.5	82	45.8	183	60.0
3	3,725	1,490	40.0	2,187	2,166	99.0	868	40.1	2,358	63.3
4	500	145	29.0	347	328	94.5	138	42.1	283	56.6
5	758	342	45.1	413	140	33.9	88	62.9	430	56.7
6	171	78	45.6	93	31	33.3	16	51.6	94	55.0
7	672	114	17.0	547	547	100.0	173	31.6	287	42.7
8	226	57	25.2	166	166	100.0	74	44.6	131	58.0
9	200	63	31.5	135	128	94.8	66	51.6	129	64.5
Total	8,998	3,469	38.6	5,338	4,020	75.3	1,662	41.3	5,131	57.0

In target groups 4, 8, 6 and 7, the realised CAPI response rate is higher than estimated. In the other target groups less CAPI response is achieved than estimated. The largest differences between p CAPI in Tables 3 and 5 can be found in target groups 4 and 9. The overall CAPI response rate is 1.4 percent points larger than estimated.

In target groups 4, 8, 9, 3 and 2, the total realised response rate is higher than estimated. In the other target groups less response is achieved than estimated. The largest differences between p tot in Tables 3 and 5 can be found in target groups 4, 8 and 1. The overall response rate is 1.2 percent points larger than estimated.

3.6.   *Method Effects for the Dutch Health Survey*

To get an idea of the effect of the adaptive survey design on the results of the Health Survey, simulations were carried out using bootstrapping. To this end, samples were drawn with replacement from the sample of the past year with the correct numbers for CAWI and matching numbers per target group for CAPI. The response data and the survey answers were then linked to these samples. For each sample, the corresponding response was weighted using the weighting model of the Health Survey. Thereafter, estimates were made for the most important target variables and these were compared with the regular estimates.

For the bootstrapping, 1,000 samples with replacement were drawn from the 2016 sample. Each sample had the right CAWI size and the right CAPI size per target group. The numbers of responses may accidentally differ from one sample to another. The sample numbers were taken from the sampling design with adaptive survey design for the Health Survey 2018. After weighting the response per sample, target variables were estimated for both the entire population and subpopulations. These estimates were compared with the results of the Health Survey 2016. One of the assumptions to use CAPI in an adaptive survey design is that the respondents' answers do not depend on the mode in which they respond. This is a strong assumption that is not always true in practice.

The target variable smoking status is known to have mode effects. The proportion of smokers among CAWI respondents is smaller than among CAPI respondents. Thus, if relatively more are observed via CAWI and fewer via CAPI, the number of smokers is expected to decrease. With adaptive survey design, more non-western migrants are approached face-to-face and fewer NL residents or western migrants. Therefore, it is expected that the proportion of smokers among non-western migrants will increase and that the proportion of smokers among NL residents and western migrants will decrease.

The results of the bootstrapping are in line with this, see Figures 3 and 4. Figure 3 shows the smoking status for non-western migrants. The estimate and the corresponding 95% confidence interval of the Health Survey 2016 are shown with the black and dashed lines. The histogram represents the results for this variable in the 1,000 samples of the bootstrapping. Figure 4 shows the smoking status of NL residents and western migrants. For NL residents and western migrants, the proportion of smokers seems to decrease when adaptive survey design is used and for non-western migrants, the proportion of smokers seems to increase compared to the measurement from 2016.

Questions about alcohol, drug use and sexual health are asked in the face-to-face approach via Computer Assisted Self Interviewing. Therefore, fewer mode effects are
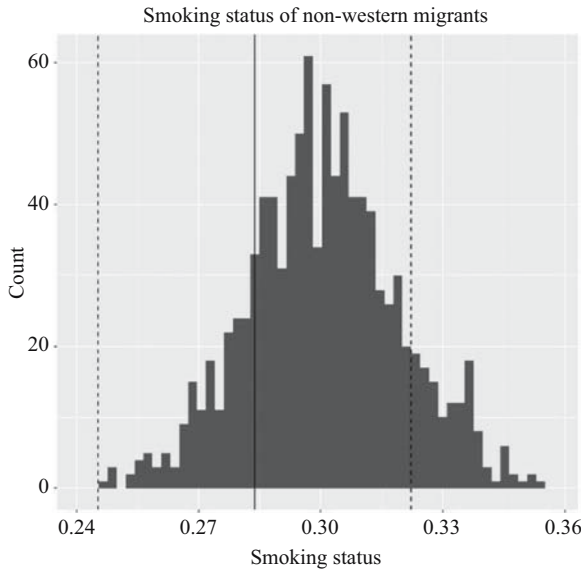
Fig. 3.   *Smoking status of non-western migrants increases with adaptive design.*

expected for these variables. Using the sample data from the bootstrapping, estimates were made for the eleven core variables: contact with general practitioner, contact with dentist, use of non-prescribed medication, experienced health, diabetes, mental health problems, disabilities, informal care, smoking, obesity and drug use, see Table 6. Columns 2016 and SE 2016 show the estimates and standard errors for the core variables of the Health Survey 2016. The last two columns contain the estimates from the bootstrapping samples.
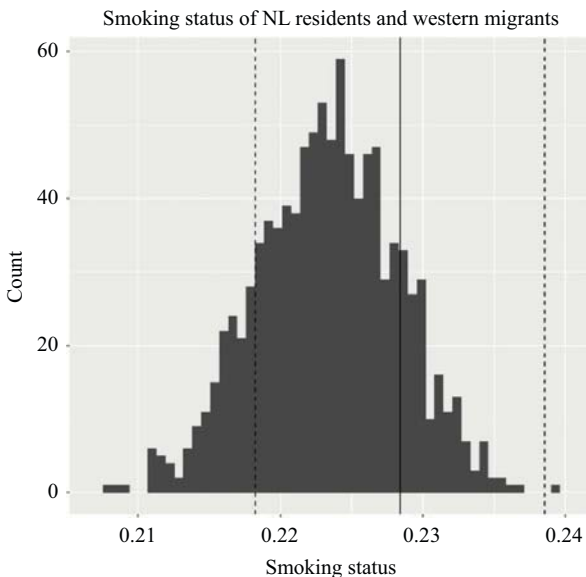

Fig. 4.   *Smoking status of NL residents and western migrants decreases with adaptive design.*

Table 6.   Estimates and standard errors for the eleven core variables for the Health Survey 2016, and for the bootstrapping samples.

	2016	SE 2016	Bootstrap	SE Bootstrap
	%			
General practitioner contact	70.9	0.6	71.2	0.8
Dentist contact	79.0	0.5	79.4	0.7
Use of non-prescribed medication	39.7	0.6	39.1	0.8
Experienced health	76.3	0.5	76.1	0.7
Diabetes	5.8	0.3	5.7	0.4
Psychologically unhealthy (MHI-5 score)	11.8	0.4	12.4	0.6
At least 1 OESO-restriction	12.2	0.4	12.0	0.5
Informal care	13.8	0.4	13.8	0.6
Smoking status	23.4	0.5	23.2	0.7
Obesity	13.6	0.4	13.8	0.4

On the basis of the bootstrapping, it is expected that most of the survey results with adaptive survey design do not differ much from those without adaptation. The greatest shifts can be seen in use of non- prescribed medication and psychologically unhealthy. Figure 5 shows the estimates of use of non- prescribed medication. The black and dashed lines show the estimate and 95% confidence interval for the Health Survey 2016. The number of people taking non-prescribed medicines is expected to decrease compared to the 2016 estimate. Figure 6 contains estimates for the variable psychologically unhealthy. It is likely that the introduction of adaptive survey design will increase the number of people who are mentally unhealthy.
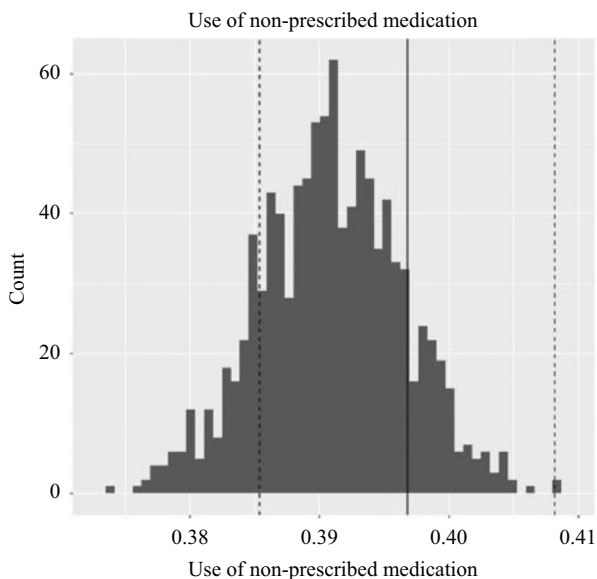

Fig. 5.   Proportion of people using non-prescribed medicines decreases with adaptive design.
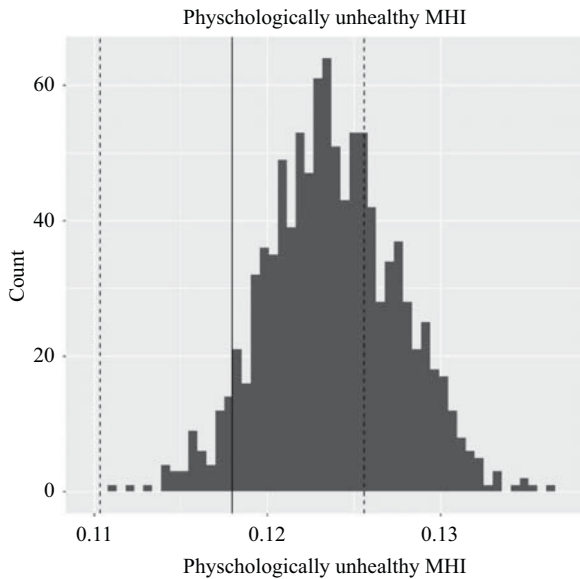
Fig. 6.   Proportion of people who are psychologically unhealthy increases with adaptive design.

4.   Dussion and Further Activities

This article describes and motivates choices that are made in the implementation of adaptive survey design at Statistics Netherlands. The focus is on sequential mixed-mode designs and the allocation of follow-up interviewer modes to nonrespondents of self-administered modes. The coefficient of variation of response propensities was adopted as the objective in optimisation of the designs. However, a range of logistical and cost constraints have been imposed and lead to a multifaceted optimisation problem. To facilitate easy management of the data collection, a case prioritisation approach was preferred over a mathematical optimisation. A case prioritisation approach is relatively easy to conduct and also is relatively robust to time change in survey design parameters such as costs and response propensities. However, improvement of the balance, that is a smaller coefficient of variation, is not guaranteed. A mathematical optimisation employing expected yield in follow-up interviewer modes does lead to improved balance, but is more sensitive to time change.

For the Health Survey case study, the implemented case prioritisation approach was compared to the mathematical optimisation approach. Results show that, as expected, on average the yield is smaller. However, balance is improved and the population strata that are allocated to face-to-face follow-up closely resemble each other. These results are promising.

There are a few limitations in this study: First, for ease of demonstration, the sampling design was restricted to simple random samples. Second, the role of mode-specific measurement bias was completely ignored. Third, the allocation of interviewer modes is posed as a simple yes-no decision, while it is clearly beneficial to also vary the amount of interviewer effort, for instance the number of contact attempts by the interviewers. Fourth,

since Statistics Netherlands' surveys are mostly repeated monthly surveys, adaptive survey designs are predominantly static, employing little paradata in making decisions.

The limitations lead to various future activities, as many surveys use unequal sampling probabilities and employ both telephone and face-to-face interview modes. Currently, the adaptive survey design framework is extended to stratified sampling designs and to multiple modes. An important decision is the choice of population strata. In this article, the focus was on explanation of nonresponse 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. Also, research into the stratification itself is conducted. Stratification of the target population can, for example, be performed using K-means clustering. This is a method that divides data into groups based on one or more characteristics, 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.

The desire to work with pre-defined monthly CAPI sample numbers per interview region makes it difficult to choose the right size per target group. Consideration can be given to whether pre-determined CAPI sample numbers for the entire country per month are sufficient for the interviewers' planning. In this case, the necessary sample numbers per target group for CAPI can be selected at random on a monthly basis from the CAWI nonresponse eligible for follow-up.

In this article and application, measurement error is neglected, while it is conjectured that mode-specific measurement biases are present in the Health Survey and also in other official surveys. Literature on the inclusion of mode-specific measurement biases in adaptive survey design is scarce. Two options have been proposed. One option is to estimate stratum mode-specific measurement biases relative to a bench measurement mode and add constraints on the absolute stratum sizes or relative sizes between strata in the optimisation, see Calinescu and Schouten (2015). Another option is to estimate stratum propensities of undesirable answer behaviour and include upper bounds to the prevalence rates of such behaviour in the optimisation, see Calinescu and Schouten (2016). Both options require extra data, either from a re- interview design or from powerful paradata or administrative data. In the application, these estimates were not available. Future work is oriented at efficient estimation of stratum biases or propensities and to include estimates in optimisation of the adaptive survey design.

Finally, the introduction of adaptive survey design requires rethinking and redesigning of the post- survey adjustment and of the estimation of precision of survey estimates. Since adaptive survey design strata are formed by variables that are present in weighting models, some adjustment may no longer be needed or could employ more parsimonious models.

5.   Appendix: Characteristics for Segmentation of the Population

1. Wealth of household: 1% groups.

2. Home ownership: owner, rent without rent subsidy, rent with rent subsidy.
3. Income: 1% groups of standardised disposable household income.
4. Socio-economic category: employee of private company, government employee, director or large shareholder, self-employed, employed other, claiming unemployment benefit, claiming income support benefit, claiming other social provision, disabled, pensioner younger than 65 years, pensioner 65 years or older, unemployed other.
5. Household size: number of people in the household.
6. Household status: child living at home, single person, partner without children, partner with children, parent in single parent household, reference person in other household, other household member.
7. Type of household: Single person household, unmarried couple without children, married couple without children, unmarried couple with children, married couple with children, married couple with children, single parent household, other household.
8. Gender: male, female.
9. Marital status: unmarried, married, partnership, divorced, widowed.
10. Age: in years.
11. Age of eldest child: in years.
12. Age of youngest child: in years.
13. Duration of stay in the Netherlands: in years.
14. Part of the country: north, east, south, west.
15. Province: the 12 provinces of the Netherlands.
16. G32: the largest 32 municipalities, other.
17. G4: the largest 4 municipalities, other.
18. Ethnicity: NL residents, western migrants, non-western migrants, unknown.
19. Ethnicity of mother: same.
20. Ethnicity of father: same.
21. Generation: NL residents, first generation migrants, second generation migrants.
22. Highest attained educational level: primary education, secondary general education, secondary vocational education, higher professional education, university.
23. Highest level of education: same.
24. Urbanity of municipality: very strongly urban, strongly urban, moderately urban, few urban and non-urban.
25. Urbanity of neighbourhood: same.

6. References

Alabama R package. Available at: https://CRAN.Rproject=alabama (accessed June 2020).

Bethlehem, J.G. 1988. "Reduction of Nonresponse Bias Through Regression Estimation." *Journal of Official Statistics* 4(3): 251–260. https://www.scb.se/contentassets/-ca21efb41fee47d293bbee5bf7be7fb3/reduction-of-nonresponse-bias-through-regression-estimation.pdf (accessed May 2020).

Calinescu, M. and B. Schouten. 2015. "Adaptive survey designs to minimize mode effects. A case study on the Dutch Labour Force Survey." *Survey Methodology* 41(2): 403–425.

Available at: https://www150.statcan.gc.ca/n1/pub/12-001-x/2015002/article/14250-eng.htm (accessed june 2020).

Calinescu, M. and B. Schouten. 2016. "Adaptive survey designs for nonresponse and measurement error in multi-purpose surveys." *Survey Research Methods* 10(1): 35–47. DOI: https://doi.org/10.18148/srm/2016.v10i1.6157.

Chun, A.Y., S.G. Heeringa, and B. Schouten. 2018. "Responsive and adaptive design for survey optimization." *Journal of Official Statistics* 34(3): 581–597. DOI: https://doi.org/10.2478/jos-2018-0028.

De Heij, V., B. Schouten, and N. Shlomo. 2015. *RISQ m2.1 manual. Tools in SAS and R for the computation of R-indicators and partial R-indicators*. Available at www.risq-project.eu. (accessed june 2020).

Groves, M.R. and S. Heeringa. 2006. "Responsive design for household surveys: tools for actively controlling survey errors and costs." *Journal of the Royal Statistical Society* Series A: Statistics in Society 169 (Part 3): 439–457. DOI: https://doi.org/10.1111/j.1467-985X.2006.00423.x.

Kaputa, S.J., and K.J. Thompson. 2017. "Adaptive design strategies for nonresponse follow-up in economic surveys." *Journal of Official Statistics* 34(2): 445–462. DOI: https://doi.org/10.2478/jos-2018-0020.

Luiten, A. and B. Schouten. 2013. "Adaptive fieldwork design to increase representative household survey response. A pilot study in the Survey of Consumer Satisfaction." *Journal of Royal Statistical Society*, Series A, 176(1): 169–190. DOI: https://doi.org/10.1111/j.1467-985X.2012.01080.x.

Moore, J.C., G.B. Durrant, and P.W.F. Smith. 2018. "Data set representativeness during data collection in three UK social surveys: generalizability and the effects of auxiliary covariate choice." *Journal of the Royal Statistical Society*, Series A, 181(1): 229–248. DOI: https://doi.org/10.1111/rssa.12256.

Peytchev, A., S. Riley, J. Rosen, J. Murphy, and M. Lindblad. 2010. "Reduction of Nonresponse Bias through Case Prioritization." *Survey Research Methods* 4: 21–29. DOI: https://doi.org/10.18148/srm/2010.v4i1.3037.

rpart package. Available at: https://CRAN.Rproject. org/package=rpart. (accessed June 2020).

Schouten, B., M. Calinescu, and A. Luiten. 2013. "Optimizing quality of response through adaptive survey designs." *Survey Methodology* 39(1): 29–58. Available at: https://www150.statcan.gc.ca/n1/pub/12-001-x/2013001/article/11824-eng.htm (accessed June 2020).

Schouten, J.G., F. Cobben, and J. Bethlehem. 2009. "Indicators for the representativeness of survey response." *Survey Methodology* 35(1): 101–113. Available at: https://www150.statcan.gc.ca/n1/pub/12-001-x/2009001/article/10887-eng.pdf (accessed June 2020).

Schouten, B., F. Cobben, P. Lundquist, and J. Wagner. 2016. "Does balancing survey response reduce nonresponse bias?" *Journal of the Royal Statistical Society*, Series A, 179(3): 727–748. DOI: https://doi.org/10.1111/rssa.12152.

Schouten, B., A. Peytchev, and J. Wagner. 2017. *Adaptive Survey Design*. Series on Statistics Handbooks. Chapman and Hall/CRC.

Tourangeau, R., M. Brick, S. Lohr, and J. Li. 2017. "Adaptive and responsive survey designs: a review and assessment." *Journal of the Royal Statistical Society*, Series A, 180(1): 203–223. DOI: https://doi.org/10.1111/rssa.12186.

Wagner, J. 2008. "Adaptive Survey Design to Reduce Nonresponse Bias." PhD diss., University of Michigan, Ann Arbor, USA. Available at: https://deepblue.lib.umich.edu/bitstream/handle/2027.42/60831/jameswag_1.pdf?sequence=1&isAllowed=y (accessed June 2020).

Wagner, J. 2013. "Adaptive contact strategies in telephone and face-to-face surveys." *Survey Research Methods* 7(1): 45–55. DOI: https://doi.org/10.18148/srm/2013.v7i1.5037.

Wagner, J. and F. Hubbard. 2013. "Using propensity models during data collection for responsive designs: Issues with estimation." Paper presented at 68th AAPOR conference, May 16-19, Boston, USA. Available at: http://www.aapor.org/AAPOR_-Main/media/AnnualMeetingProceedings/2013/Session_H-7-4-Wagner.pdf (accessed June 2020).