Experiments on the data collection process

Observing impact on response rates and data quality

Sofia Holsendahl
Andreea Bolos
Viktor Dahl

SCB

Table of contents

Abstract

A continuous objective within the Section for Survey Services of Statistics Sweden is to oversee and enhance the data collection process. In doing so, the purpose of this paper is to study the data collection process by examining the results of three types of experiments executed with the purpose of observing impact on response rates and data quality.

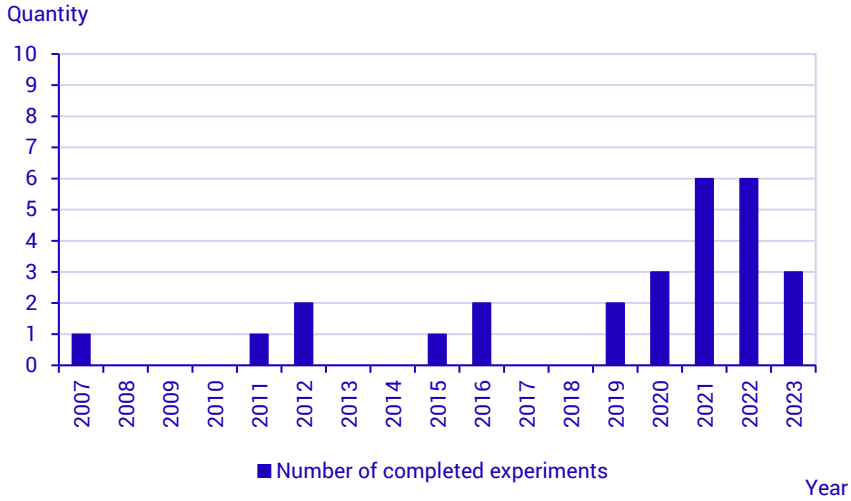
The objectives of our experiments were to: a) observe the extent to which a shortened questionnaire could enhance response rates, b) observe the extent to which *more* web-oriented contact approaches could enhance the share of digitally collected responses, and c) assess the extent to which a fifth (compared to our standard of four) send out generates a higher response rate and a more varied composition of answers.

Introduction

An overall objective within the Section for Survey Services of Statistics Sweden is to oversee and enhance the data collection process. This is especially important nowadays when the response rate is decreasing. We live in a digital world, and to constantly be up-to-date and see if we can reverse the trend, or at least stabilize the response rate, we carry out experiments to analyze the effect of various aspects of the data collection.

Figure 1 below shows the number of experiments that were carried out between 2007-2023 at the Section for Survey Services at Statistics Sweden. As the response rate has decreased over the years, the need to conduct experiments has increased. In order to be able to carry out experiments, a suitable survey is needed as well as resources in terms of time and funds.

Figure 1 Number of completed experiments at the Section for Survey Services 2007-2023.



This paper will focus on the results from five different experiments; shortened questionnaires, a *more* web-oriented contact approach, and the use of a fifth send out instead of four.

Experiments

Shortened questionnaires

In 2020, we employed two experiments to study the effect of the length of the questionnaire. Each experiment was set up in a specific survey. The first survey focused on the overall health of the population and the survey sample included all individuals in Sweden aged 16-84. The original questionnaire contained of 103 questions. The questions in the paper form were distributed over 13 pages.

The second survey focused on high-school students and their educational interests and whether they intended to continue their studies or not. The sample was comprised of high school students. The questionnaire consisted of 10 pages and about 30 questions.

The experiments tested whether a shortened questionnaire could give a higher response rate than the original questionnaire. The individuals in the experimental group got the shortened questionnaire while the control group got the original questionnaire. The experiments were carried out completely independently of each other.

Table 1 below shows the number of questions in the original questionnaires and in the shortened ones. The health survey used a shortened questionnaire with 20 questions on four pages. In the survey to people in high school, the shortened questionnaire consisted of nine questions and were distributed over two pages.

Table 1 Number of questions and pages in the questionnaires used in the experiments.

	Health survey	Study interest
Original questionnaire		
Number of questions	103	30
Number of pages	13	10
Shortened questionnaire		
Number of questions	20	9
Number of pages	4	2

The data collection strategy was identical regardless of whether a respondent got the original or the shortened questionnaire, as was the content of the various information letters in the send outs. In both surveys, the respondents could choose answering the questions on paper or on the web.

The results of the experiments are showed in table 2.

Table 2 Response rates in percent (%) with 95 % CI.

	Health survey	Study interest
Original questionnaire	28,8 ± 0,7	38,8 ± 1,7
Shortened questionnaire	33,1 ± 2,3	44,4 ± 2,7

As seen in the table, the response rate was significantly higher in the experimental group – who got the shortened questionnaire – than in the control group, in both surveys.

Web-oriented contact approach

In the same national health survey mentioned above, another experiment was conducted in 2021. Since a few years back, the survey uses a web-oriented data collection, which was the standard procedure for data collection in surveys directed toward individuals at Statistics Sweden in 2021. The strategy was called *Web intensive (WI)* and out of a total of four send outs, two of them included a paper questionnaire.

Table 3 Description of the Web intensive strategy.

	Web intensive (WI)
Send out 1	Invitation with login to the web survey
Send out 2	Reminder including a paper questionnaire
Send out 3	Reminder
Send out 4	Reminder including a paper questionnaire

In order to test some alternative and even more web-oriented collection strategies, an experiment was carried out with the goal of increasing the proportion of digitally collected data. In addition to the WI-strategy mentioned above, three other strategies were therefore used in the experiment.

Table 4 Descriptions of the different data collection strategies used in the experiment.

	Extra web intensive (EWI)	Super extra web intensive (SEWI)	Digital super extra web intensive (DSEWI)
Send out 1	Invitation with login to the web survey	Invitation with login to the web survey	Invitation with login to the web survey
Send out 2	Reminder	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire
Send out 3	Reminder including a paper questionnaire	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire
Send out 4	Reminder	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire	Reminder including information that you can contact Statistics Sweden to receive a paper questionnaire

In terms of send outs, the difference between *Super extra web intensive (SEWI)* and *Digital super extra web intensive (DSEWI)* was that in DSEWI, persons with a digital mailbox received send out 1, 2 and 4 to their digital mailbox and send out 3 by paper. Persons without a digital mailbox were treated the same way as SEWI. With a digital mailbox, people receive mail from the authorities digitally through your phone or computer, instead of via postal mail.

Table 5 shows the results of the experiment and the response rates in the four groups.

Table 5 Response rates in percent (%) with 95 % CI.

	WI		EWI		SEWI		DSEWI	
Total	44,3	± 0,5	42,7	± 1,4	38,9	± 1,4	40,5	± 1,4

The standard strategy (WI) resulted in the highest response rate, 44.3 percent. The response rates in the three more web-oriented strategies were 42.7 percent (EWI), 38.9 percent (SEWI) and 40.5 percent (DSEWI).

The differences between the standard strategy and the two strategies SEWI and DSEWI are statistically significant.

Five send outs

The last two experiments presented in this paper were conducted in 2022. As mentioned above, the Statistics Sweden standard strategy for data collection from individual populations contains of four send outs. In a survey about children's media use, where the population was parents to children aged 5-8 years, the response rate was 23.7 percent after four send outs. In another survey to assistant nurses about their work environment and health, the response rate was 24.8 percent when the ordinary data collection period was completed. As the response rates were low, a fifth send out was carried out, in order to hopefully increase response rates.

The two surveys had a different type of the fifth send out which is illustrated in table 6.

Table 6 Descriptions of the data collection strategies used in the surveys.

	Children's use of media	Assistant nurses' work environment
Send out 1	Invitation with login to the web survey	Invitation with login to the web survey
Send out 2	Reminder	Reminder
Send out 3	Reminder including a paper questionnaire	Reminder including a paper questionnaire
Send out 4	Reminder	Reminder
Send out 5	Reminder including a paper questionnaire	Reminder

The fifth send out had some effect on the response rates.

Table 7 Response rates in percent (%).

	Children's use of media	Assistant nurses' work environment
After four send outs	23,7	24,8
After five send outs	27,5	26,1

The response rates in the media survey differed by 3.8 percent and by 1.3 percent in the work environment survey. Additionally, the experiments did not indicate that the additional responses obtained through a fifth send out had a substantial impact on the estimates.

We also used the Representativity Indicator (R-indicator) to analyze the results of the experiments. The R-indicator is a measure of how representative responses are relative to the sample. As such, the R-indicator provides a measure based on the variation of response rates between different groups and can vary between 0 and 1. If all response

rates are equal, i.e., if the variation is zero, the indicator has the value 1. A value of 1 thereby represents complete balance between responses and sample. The indicator's value is dependent on the variables included in the calculation.

Table 8 The value of the R-indicator before and after the fifth send out.

	Children's use of media	Assistant nurses' work environment
R-indicator before the fifth send out	0,75	0,79
R-indicator after the fifth send out	0,77	0,79

The value of the R-indicator differed slightly in the survey about children's use of media, but the difference was not statistically significant.

Discussion and conclusion

Carrying out experiments on the data collection gives us a better awareness of how the response rate can be affected by various changes in the process. Based on the experiments described above, we can see that a shorter questionnaire has a positive effect on the response rate. We can also see that if we remove, completely, the option to answer a paper questionnaire, it affects the response rate negatively. We also learned that introducing send out to people's digital mailboxes had a positive effect on the response rate. A final discovery was that an extra send out may give a slightly higher response rate but estimates and the R-indicator are hardly affected at all by an additional fifth send out. Statistics Sweden's users and customers always require statistics with high credibility and reliability. Therefore, we must ensure that we continue to be aware of the response burden and how different choices in the data collection process affect the response rate.

Based on the results above, when we develop questionnaires, we generally promote the ones that are simpler to understand and shorter in size. After the experiment regarding web-oriented contact approach, Statistics Sweden changed the standard strategy for collection data in population surveys. Now, we always use digital mailboxes to people who has one, and we send one paper questionnaire to people aged 18-64 and two to people aged 65 or older.

What happens next? Statistics Sweden keeps conducting experiments and we plan for more in the future. An experiment has just been completed where the results are being analyzed. The experiment used a shortened information letter in the first send out instead of the standard one. During the fall this year, another experiment will be carried out in which there will be no paper questionnaire offered for people aged 18-64. It is a slightly larger experiment where the goal is to be able not only to analyze the effects on the response rate and the estimates in the tables, but also to study the reduced costs when no paper questionnaires are used to a large part of the population in the survey. How much money do we save by not sending a paper questionnaire to everyone between the ages of 18-64? Is it worth saving the cost in return for the response rate dropping? How are the estimates in the tables affected?

We also intend to conduct an experiment with reward in terms of gift cards of SEK 99. The purpose of using rewards is primarily to increase the response rate and to increase representativeness through a higher response rate of groups with low response rate. The principle is that Statistics Sweden primarily uses rewards provided prior to participation in a survey and without performance requirements. The principle is based on research on the effects of rewards.

In order to pick-up new research findings and to learn from experiments carried out by others, Statistics Sweden devotes time to external monitoring. Moreover, by continuing to conduct experiments, we can be involved and perhaps influence the response rate. It may be possible to anticipate how certain changes affect the response rate, but we don't know for sure until we have tested it.