CONFERENCE OF EUROPEAN STATISTICIANS

Workshop on Statistical Data Collection

10-12 October 2018, Geneva, Switzerland

Session 2

24 September 2018

The effect of response measures in business surveys

Ger Snijkers, Margreet Geurden-Slis, Goele Goris, Joep Burger & Lisanne van den Hombergh (Statistics Netherlands)

g.snijkers@cbs.nl

Abstract

Often asked questions concerning getting response for business surveys are:

- What will be the increase in response rate if we apply such-and-such measure(s)?
- What would be perfect timing for these measures? And,
- What will be the costs?

These questions can also be put differently: if we have a targeted response rate of x%, what measures do we need to take (and when) to achieve this? This is a quality target which may be put forward by survey sponsors or data users (e.g. Eurostat). Basically these questions ask for an efficient strategy to get response, aiming for a cost-efficient survey design both for the survey organisation (like a National Statistical Institute) and businesses alike, not chasing and burdening businesses too much.

The effects of measures to get response for business surveys have not been studied systematically as much as for social surveys. Obvious reasons for this may be the fact that business surveys are mandatory by law (which by itself has a major positive effect on response rates), business surveys run for many years using the same design, and the costs involved in getting response are not as high as for social surveys using expensive modes like CAPI or CATI. Nowadays however, with ever decreasing budgets, and the pressure to reduce response burden even more, efficient business surveys designs are required. An overview of various measures has been presented by Snijkers et al. (2013), but quantitative information to answer the above mentioned questions is to a large extend still lacking.

In a study conducted by Statistics Netherlands, the effects of various measures to get response have been analysed for a number of business surveys, without doing an experiment, as an attempt to get more insights in the effects of these measures. These measures include the obvious measures, like sending advance letters to businesses introducing the survey and soliciting survey response, sending pre-due data reminders, and after the due date sending one or more reminder letters. For one survey (the Survey on International Trade in Goods) we modelled the effects of these measures using survival analysis, to find out what would have happened without any of these measures. In the paper the results of these analyses will be presented.

References

Snijkers, G., G. Haraldsen, J. Jones, and D.K. Willimack, 2013, Designing and Conducting Business Surveys. Wiley, Hoboken. Text.

Statistics Netherlands

The effect of response measures in business surveys

Ger Snijkers[1]
Margreet Geurden-Slis
Goele Goris
Joep Burger
Lisanne van den Hombergh

Paper presented at the UNECE Workshop on Statistical data Collection "Resourceful data Acquisition"
10-12 October 2018
Geneva, Switzerland

1. Getting response in business surveys: Introduction

Often asked questions concerning getting response for business surveys are:

- What will be the increase in response rate if we apply such-and-such measure(s)?
- What will happen if I drop a measure in my strategy?
- What will be the costs? And,
- What would be appropriate timing for these measures?

These questions can also be put differently: if we have a targeted response rate of x%, what measures do we need to take (and when) to achieve this? A targeted response rate is a quality indicator that may be put forward by survey sponsors or data users (e.g. Eurostat). The assumption is that high response rates yield unbiased and precise estimates. Basically these questions ask for an efficient strategy to get response, aiming for a cost-efficient survey design both for the survey organization (like a National Statistical Institute, NSI) and businesses alike, not chasing and burdening businesses too much, and at the same time achieving good data quality.

The effects of measures to get response for business surveys have not been studied systematically as much as for social surveys (see e.g. Groves and Couper, 1998; Dillman, 2000; Groves et al., 2002; Stoop, 2005; Stoop et al., 2010; Luiten 2013, Dillman et al., 2014, to list a few references). Obvious reasons for this may be the fact that business surveys are mandatory by law (which by itself has a major positive effect on response rates), business surveys run for many years using the same design, and the costs involved in getting response are not as high as for social surveys using expensive interviewer-administered modes like CAPI or CATI. Nowadays however, with ever decreasing

[1] Contact: G.snijkers@cbs.nl
The views expressed in this report are those of the authors and do not necessarily reflect the policies Statistics Netherlands.

budgets, and the pressure to reduce response burden even more, efficient business surveys designs are required. An overview of various measures has been presented by Snijkers et al. (2013) and McCarthy & Earp (2012), but quantitative information to answer the above mentioned questions is to a large extend still lacking, although more and more studies are being conducted (see e.g. Tuttle et al., 2018; Langeland, 2018; Sakshaug and Vicari, 2018; Willimack et al., 2018).

In a study conducted by Statistics Netherlands (SN) the effects of various measures to get response have been analyzed for a number of business surveys, without doing an experiment. The aim of this study is to get more insights in the effects of these measures, hoping for a timely answer to the above mentioned questions. The measures included are the obvious ones, like sending advance letters to businesses introducing the survey and soliciting survey response, sending pre-due date reminders, and after the due date sending one or more reminder letters. For a number of surveys also telephone reminding has been applied. For one survey (the survey on International Trade in Goods) we modeled the effects of these measures using survival analysis, to find out what would have happened without any of these measures. In this paper the results of these analyses will be presented.

Section 2 provides an overview of response measures analysed, followed by an overview of the surveys analysed in section 3. Section 4 discusses the results, while in section 5 we discuss survival analysis as applied to the Survey on International trade in Goods as well as the results. Section 6 concludes this paper.

2. Overview of response measures

A variety of measures for getting response are described by Snijkers and Jones (2013). They classify these measures in pre-field measures (like pre-notifications), field measures (like sending advance letters, and reminders for non-response follow-up), and post-field measures (like enforcement measures, and re-contacts for reasons of data validation). It may be clear that none of these individual measures are sufficient to reach the targeted response rates. They conclude that a lot of effort is needed to getting acceptable levels of response. This means that a number of actions are needed to meet the targeted response rates. The question now is, as we also said in the Introduction: what is an efficient strategy to get response?

The commonly applied strategy for business surveys at SN, includes the following measures:

- Advance letters. Advance letters for electronic questionnaires are sent to the sampled businesses. At SN until now letters are sent to inform businesses about a survey. These letters introduce the survey. In general these one-page letters start with a brief introduction of SN, followed by an introduction of the survey and where more information about the survey results can be found, the due date, in case the survey is mandatory by law it is indicated that businesses may be fined if they don't respond in time, how businesses can go the questionnaire (the web address, a user name and a password), contact information in case of questions, and the letters ends with information about the consolidation cluster indicating for which business unit(s) the questionnaire should be completed, and 'thank you in advance'. An example of such a letter is presented in Figure 1: the advance letter for the Annual Structural Business Survey, to give an impression of what the letter looks like. Dutch law regulates the maximum number of days businesses get to respond, i.e. the maximum number of day between receiving the letter and responding:

- Annual surveys: 60 days,
- Quarterly surveys: 30 days,
- Monthly surveys: 10 days.

- Pre-due date reminders. At some time before the due date (about two weeks), businesses that have not yet responded will receive a pre-due date reminder letter. This is a kind reminder telling businesses that we have not heard from them yet and the due date is coming close.
- Reminder letters: after the due date, businesses receive one or more reminder letters asking them to respond. Basically the same information as in the advance letter is repeated, but stronger language for the mandatory section is used. In the first reminder a reference to the law on Statistics Netherlands is included; in the second reminder the fine is included which can be as high as 16,000 euros per day, depending on size and past response behavior. Finally a third reminder is sent, given businesses a final option to respond. If they still not respond it is indicated that the enforcement procedure will be started.
- Telephone reminder. In some cases businesses are called on the phone as a non-response follow-up measure. In this call, conducted by telephone interviewers, businesses are only reminded of the questionnaire. In response, sometimes businesses indicate that they will complete the questionnaire. Because this follow-up measure is expensive it is not applied for all non-responding businesses, but only for selected businesses, i.e. the most relevant businesses that are still missing. For the Survey on International Trade in Goods this call was not just a reminder, but interviewers tried to motivate businesses to respond. As such, this was called a 'motivation call'.

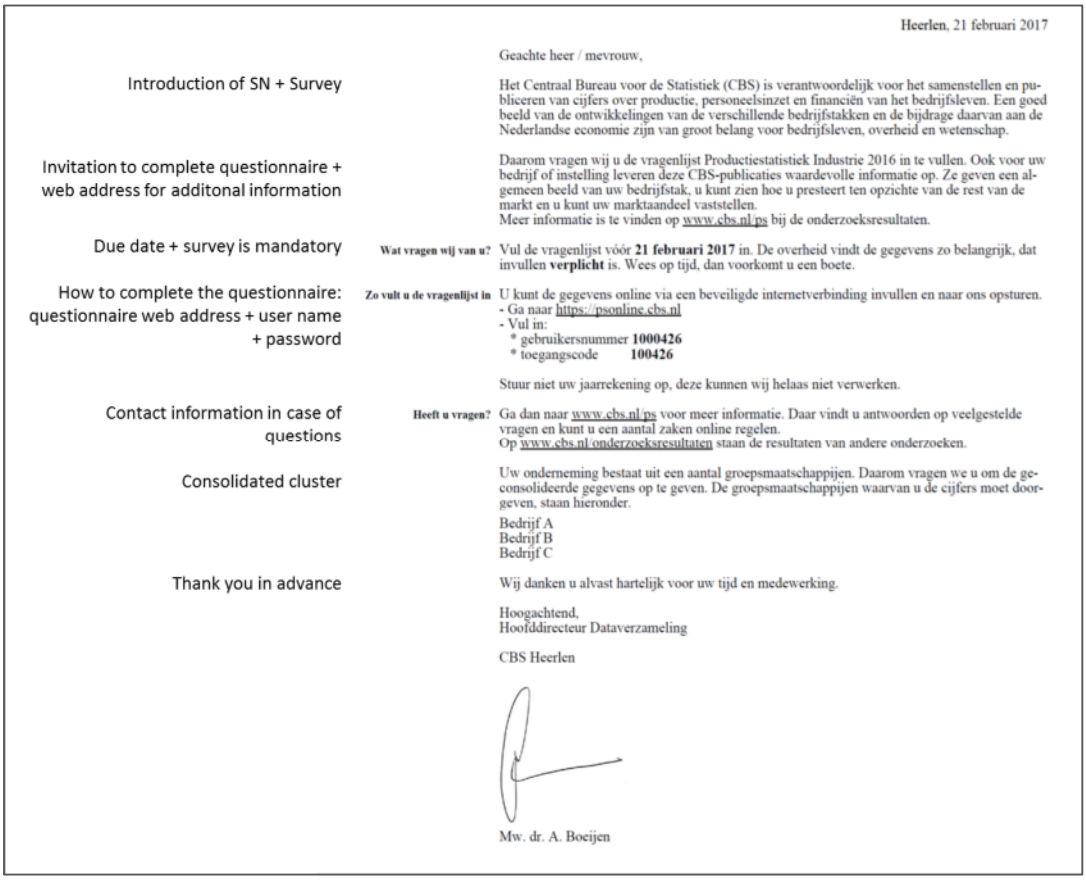

Figure 1 Advance letter for the Dutch Annual Structural Business Survey (SBS)

3. Surveys analysed

The effect of the applied field measures on the final response rate have been studied for 8 business surveys. These surveys are listed in Table 1, including a number of characteristics:

- Five surveys cover all sectors of the economy (all NACE/NAICS codes); three surveys are agricultural surveys.
- In this study we looked at annual surveys; one survey is conducted tri-annually.
- Five surveys are mandatory by law, three are voluntary.
- Six use telephone reminding to enhance response rates as a final non-response follow-up.
- All surveys use electronic questionnaires. Paper may be included in the envelope of a final reminder letter, but the policy of SN is to use as less paper as possible.

Most surveys cover all sectors of the economy (all NACE codes); three surveys are agricultural surveys. In general agricultural surveys are cumbersome to get high enough response rates. In 2016 most of these surveys have become mandatory by law (Survey on Harvest Estimates, Survey on Harvest Estimates: vegetables open soil). As a result of this fact response rates for these surveys increased with about 25 percentage points, from below 50% to about 75% (Houben et al., 2018). The agricultural survey (on Grassland Usage) that did not become mandatory stayed at a low level, with response rates below 50%.

In Table 1, the surveys are ordered according to total sample size. This is the actual number of businesses that are contacted and received an advance letter. Table 1 also includes the final total unit response rates and costs per unit response. Regarding the response rate, we have counted the number of questionnaires returned to SN. The amount of returned questionnaires is not equal to the amount of useable questionnaires, which is the actual usable response. In this paper we analyzed 'return rates' instead of actual useable response rates (Snijkers et al., 2013). As such we analyzed the effect of the measure on respondents to look into the questionnaire and return those; we will not draw conclusions on the manner the questionnaires are completed and the quality of the returned data.

Table 1 Overview of business surveys

Survey	Name	Sector of the economy	Frequency	Mandatory by law?	Telephone reminding	Total sample size	Total unit RR (%)	Costs per unit R (€)	Year analysed
SBS	Structural Business Survey	All sectors	Annual	Yes	No	71,719	74 (56)†	2.22 (2.57)†	2015
						75,518	74 (53)†	2.27 (2.71)†	2016
						78,605	- (58)†	- (2.63)†	2017
INV	Survey on Investments	All sectors	Annual	Yes	Yes	46,824	86	1.36	2018
ICTB	Survey on ICT Usage by Businesses	All sectors	Annual	No	No	12,114	74	1.93	2017
ITG	International Trade in Goods	All sectors	Annual	Yes	Yes	8,457	72	3.29	2016
ISGVC	Survey on International Sourcing and Global Value Chains	All sectors	Tri-annual	No	Yes	3,156	79	2.03	2017
HarvEst	Agricultural Survey on Harvest Estimates	Agriculture	Annual	Yes	Yes	4,443	72	2.12	2017
GrassU	Survey on Grassland Usage	Agriculture	Annual	No	Yes	2,528	40	5.01	2017
VegOS	Agricultural Survey on Harvest Estimates: vegetables open soil	Agriculture	Annual	Yes	Yes	2,453	71	2.38	2017

† At the moment of writing this paper the fieldwork of the 2017 SBS was not yet completed and the response data not yet analysed. The data are available until the 2nd reminder. For reasons of comparison, the results until this moment for the 2015 and 2016 SBS are also include in the table, in between brackets.

For most surveys the survey communication strategy to get response included the following steps:

1. An advance letter is sent.
2. About two weeks before the due date a pre-due date reminder is sent.
3. After the due date, two reminders letters are sent.
4. A third reminder may be either a telephone follow-up or a final reminder letter.

An exception to this strategy is the survey on International Trade in Goods (ITG) for the yearly reporters (Fig. 2). In 2016 a new communication strategy was applied, when the ITG Survey was entirely redesigned (Geurden-Slis and Snijkers, 2017). Businesses with trade values of over 5 million euros had to report on a monthly basis, while businesses with a total trade value between 1.5 and 5 million euros could also report on an annual basis. For these annual reporters (the smaller businesses), the communication strategy was:

1. An advance letter (8457 businesses)
2. A pre-due date appeal letter (5426 businesses)
3. A motivational call (1478 businesses)
4. A final reminder letter for those businesses that were contacted on the phone (909 businesses)
5. Enforcement (266 businesses)

It turned out that this strategy was quite effective: compared to the years before the redesign the response rate for this homogeneous group of businesses increased from about 40% in 2014 to over 70% in 2016. Where the return rate in 2014 stopped at the level of 40%, in early 2016 on the due date the return rate already reached 48%. This increase was based on the pre-due date reminder.

Although successful, in practice it turned out that for a substantial number of businesses the main route was not applied (light gray streams in Fig. 2). 91% of the 2193 final non-respondents did not receive all four interventions: 439 did not receive the pre-due date appeal letter in step 2, 1603 did not receive the motivational call in step 3, 1242 did not receive the final reminder letter in step 4, and 1717 were not enforced to respond. Some of these deviations from the strategy can be explained: no phone number was available, the reminder letter is only sent after a call, and enforcement was limited to importing businesses. Exporting businesses are liable to file a tax return, which is used to complement the survey data.

The effect of the communication strategy on the cumulative response will be analyzed in more detail through modeling in Section 5. This will reveal that the communication strategy could potentially reach almost a 90% response rate if applied to all sampling units.

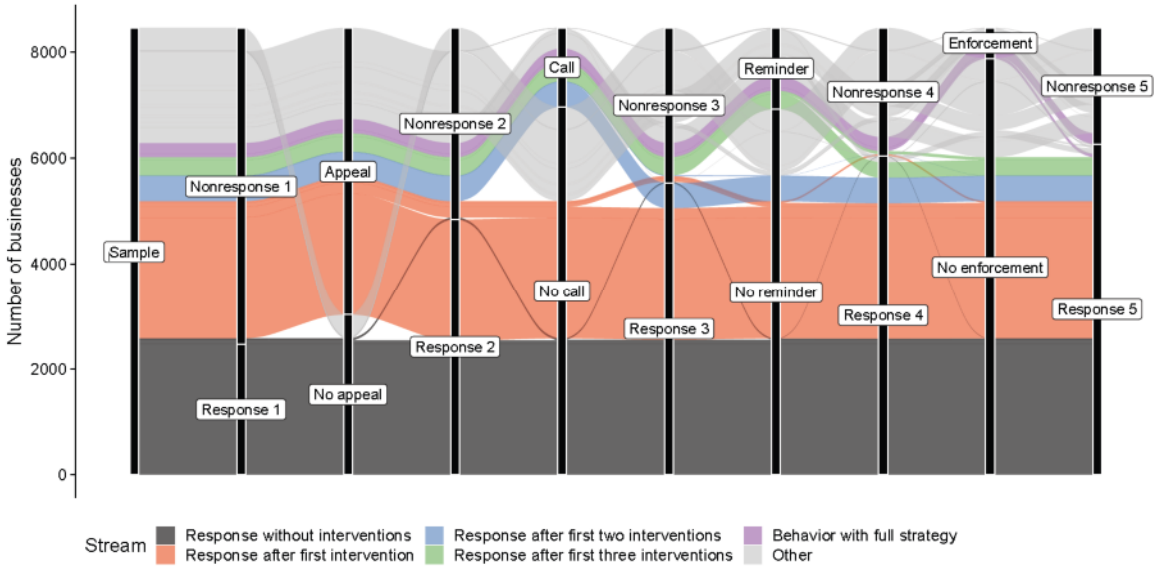

Figure 2 Alluvial plot of the response process of the survey on International Trade in Goods 2016. The first 16 streams (dark gray and colored) either follow the communication strategy or respond despite missed interventions. The other 24 streams (light gray) deviate from the communication strategy. The communication strategy comprises four interventions: pre-due date appeal letter and post due-date motivational call, reminder letter and enforcement.

4. Results

Table 2 shows an overview of the effects of the various field measures. The numbers are visualized in Figure A1.1 and A1.2 in the Appendix 1.[2] For each measure the following parameters are presented:

- p_s: the fraction of potential respondents responding in step s of the process. p_s is calculated by dividing the number of returned questionnaires in step s (r_s) by the number of businesses that were contacted in step s (n_s): $p_s = r_s/n_s$. This parameter can be interpreted as the response probability in step s, and indicates the effectiveness of each step. For comparison between surveys, it should be weighted by the number of days (d_s) a step takes (Figure A1.1).
- rr_s: the fraction of the initial sample responding in step s. rr_s is calculated by dividing the number of returned questionnaires in step s (r_s) by the total sample size (n), i.e. the total number businesses that received an advance letter questionnaire: $rr_s = r_s/n$. The sum of these response rates is the total response rate: $\sum_s rr_s = r/n = rr$. This parameter indicates the effectiveness of the strategy as a whole. For comparison between surveys, it should be weighted by the total of number of days ($\sum_s d_s$) and total costs (see Table 3 and Figure A1.2).
- d_s: the number of days in step s, i.e. the number of days to reach the resulting rr_s. The sum of all these days is the total number of days used for collecting the data (the fieldwork): $\sum_s d_s = d$. This parameter is added as the result depends in this time interval.

The higher the probability of getting response in a step (p_s, the more effective this step is. A step may have a high probability of getting response, but still the added response rate rr_s for a step may be low, e.g. when a measure is applied to a small number of businesses. This may happen when a

[2] For more visualisations see the presentation (DC2018_S2_Snijkers et al_Netherlands_P.pptx) coming with this paper, at https://statswiki.unece.org/display/Collection/2018+Data+Collection+Workshop%3A+Documents,.

step comes late in the process, or when a measure is applied to a small number of selected businesses (as is the case for telephone reminding). For instance, when looking at the International Trade in Goods Survey, a response probability of .38 for the final reminder letter indicates that 38% of the businesses receiving this letter responded; the added response rate for this step however is small: 4.1% (see Section 5 for a more quantitative analysis).

When interpreting the results as shown in Table 2, we should keep in mind that the probabilities of getting response are in fact conditional probabilities: the results of a next step are conditional on previous steps. The response rates are achieved in a sequence of measures. This may be particular relevant with regard to the interpretation of the reminding steps in follow-up to a pre-due date reminder: from these data we cannot draw conclusions for strategies without a pre-due date reminder.

Looking at Table 2 a number of conclusions can be drawn:

- The response rates for all surveys are quite high (70-80%), except for the Survey on Grassland Usage (40%), which is a voluntary agricultural survey. The mean response rate for these surveys is 76% (st.dev. is 5.7%), excluding SBS2017 and GrassU2017 (the median over all surveys is 74%).
- The response probabilities p_s for all steps are quite reasonable (on average .27 for all steps and all surveys; st.dev. is .13; median is .26), indicating that all steps have an effect to get response. Also the pre-due date reminder seems to be effective, as is telephone reminding (but at higher costs). The overall picture however is quite heterogeneous, and the variability is high. The response probabilities vary from 0.02 to 0.48.
- Diverging pictures are shown by the Survey on Grassland Usage (GrassU), and the Structural Business Survey (SBS).
    - For the Survey on Grassland Usage all steps seem to be quite ineffective with low response probabilities (ranging from 0.07 to 0.19). It striking to see that the step with the highest response probability (the 2nd reminder: 0.19) has the lowest number of days (8 days).
    - For the structural Business Survey (2015 and 2016) the response rates achieved in the first step (after the advance letter was sent) is quite low, as compared to other surveys. The effect of the pre-due-date reminder is comparable to other surveys, while the first and second reminders are less effective. The final step, after the second reminder, is quite effective again. This would indicate that in general we have two groups of businesses: those that are able to respond early in the process, and those who respond late in the process. The data do not tell why this is, but we have some ideas as we will discuss in Section 6.

Table 2 Response results for the various measures to get response for a number of surveys conducted by Statistics Netherlands

Survey	Total sample size	After advance letter			After pre-due date reminder letter			After 1st reminder			After 2nd reminder			After final reminder						Total unit RR	Total nr of days
														Final telephone reminder			Final reminder letter				
	n	p_s	rr_s (%)	d_s	p_s	rr_s (%)	d_s	p_s	rr_s (%)	d_s	p_s	rr_s (%)	d_s	p_s	rr_s (%)	d_s	p_s	rr_s (%)	d_s	rr (%)	d
SBS2015	71,719	.18	18.0	42	.34	27.9	35	.08	4.1	9	.12	5.8	30				.42	18.6	206	74.3	322
SBS2016	75,518	.17	17.1	42	.36	29.6	49	.02	1.1	6	.10	5.4	20				.44	20.4	249	73.7	366
SBS2017	78,605	.17	17.5	51	.19	15.6	14	.15	10.3	14	.26	14.8	33								
INV2018	46,824	.35	34.8	21	.38	24.8	46	.45	18.0	23	.35	7,9	11	.48	0.2	71				85.7	172
ICTB2017	12,114	.32	31.8	36	.25	17.1	13	.27	13.7	27	.21	7.8	21				.13	3.7	20	74.1	117
ITG2016†	8,457	.30	30.1	47	.48	30.8	19	.22	1.6	61				.33	5.7	24	.38	4.1	61	72.3	151
ISGVC2017	3,132	.37	36,7		.38	24,2		.58	22,8					.06	0.5					84,2	
HarvEst2017	4,443	.22	22.3	21	.25	19,7	25	.48	27.9	7				.38	2.2	56				72.0	109
GrassU2017	2,528	.07	6.9	28	.08	7.5	18	.12	10.3	10	.19	14.1	8	.16	0.8					39.6	
VegOS2017	2,453	.30	29.6	29	.20	13.7	22	.23	13.2	9	.31	13.4	13	.29	1.3					71.3	
Mean		.25	24.5	35	.29	21.1	27	.26	12.3	18	.22	9.9	19	.28	1.8	50	.34	11.7	134	76.0	
St. deviation		.10	9.6	11	.12	7.6	13	.19	8.8	17	.09	4.1	10	.15	2.0	24	.14	9.0	111	5.7	
Median		.26	26.0	36	.30	22.0	22	.23	11.8	10	.21	7.9	20	.31	1.1	56	.40	11.4	134	73.7	

†Order of interventions differs: in ITG2016 the telephone reminder came before the first reminder

Table 3 shows the costs for the surveys analysed. The following parameters are presented:

- The total number of letters sent. This is the sum over all steps of businesses contacted in each step.
- The variable costs in euros for the total number of letters. The costs include postage, setting and printing.
- For telephone reminding the number of businesses contacted on the phone is counted, not the final number of calls.
- The total variable costs in euros for contacting these businesses by phone.
- The total cost is the sum of the costs of paper and telephone reminding.
- The cost per unit response (euros) is the total costs divided by the total number of responding businesses. This parameter could be interpreted as an overall cost-efficiency indicator, indicating whether the strategy as a whole is cost-efficient. However, the number of businesses reminded by phone should be taken into account. E.g. for the International Trade in Goods Survey the costs per unit response are quite high as compared to the other surveys, but the number of businesses contacted by phone is also quite high.

The numbers in Table 3 show that:

- On average the costs per unit response are 2.51 euros (st.dev. is 1.06 euros). If we leave GrassU2017 out as an outlier, the mean is 2.20 euros (st.dev is 0.54), which is close to the median (2.22 euros).
- The most cost-efficient surveys are the surveys on Investments (INV) and ICT Usage (ICT). Also the Survey on International Sourcing and Global Value Chains is cost efficient.
- The Survey on Grassland Usage (GrassU) is not cost-efficient, as compared to the other surveys.
- For the Structural Business Survey (SBS) the cost efficiency increases after the second reminder. The numbers in between brackets show the costs per unit response after the second reminder.

Table 3 Costs for the surveys analyzed

Survey	Paper reminding		Telephone reminding		Total costs	Costs per unit response (euros)
	Total nr of letters sent	Costs (euros)	Nr of businesses	Costs (euros)		
SBS2015	236,971	118,500			118,500	2.22 (2.57) †
SBS2016	253,016	126,500			126,500	2.27 (2.71) †
SBS2017	240,522	120,250			120,250	(2.63) †
INV2018	106,740	53,400	150	1,100	54,500	1.36
ICTB2017	34,676	17,325			17,325	1.93
ITG2016	15,396	7,700	1,478	12,425	20,125	3.29
ISGVC2017	6,590	3,300	249	2,050	5,350	2.03
HarvEst2017	10,476	5,225	260	1,550	6,775	2.12
GrassU2017	8,938	4,475	130	550	5,025	5.01
VegOS2017	6,629	3,325	115	840	4,150	2.38

† At the moment of writing this paper the fieldwork of the 2017 SBS was not yet completed and the response data not yet analysed. The data are available until the 2nd reminder. For reasons of comparison, the results until this moment for the 2015 and 2016 SBS are also include in the table, in between brackets.

5. The Survey on International Trade in Goods

5.1 Method

In this section, we aim to estimate the effect of interventions that are intended to enhance the response to a business survey. Since resources are often lacking to set out a proper randomized controlled experiment, and conducting such an experiment in business surveys in itself is difficult because of confounding effects (e.g. 'trained' respondents like previously and multi-surveyed businesses), we model the response propensities over time and apply extrapolation in order to mimic the missing control group not having received the intervention. The method we used is survival analysis. This method models the number of businesses that could have responded on a daily basis. This method is illustrated with the Survey on International Trade in Goods (ITG).

The International Trade in Goods Survey (ITG) measures the value and quantity of goods traded between EU member states (Intrastat) and by EU member states with non-EU countries (Extrastat). Extrastat data can be derived from customs declarations, but since the removal in 1993 of customs formalities within the EU, Intrastat data have to be collected directly from EU-traders by the NSI of each member state. Statistics Netherlands therefore conducts a survey among the Dutch EU-traders.

As explained in Section 3, for businesses with a trade value in the range between 1.5 and 5 million euro ('annuals'), the survey communication strategy consists of an advance letter and, in case of non-response, a pre-due date reminder letter, a motivational call, a post-due date reminder letter and enforcement. Here we model the effects of these response enhancement interventions for 2016 (Table 4). Numbers slightly deviate from those in Section 3, because we excluded businesses that responded before the end of the reporting year and non-respondents that deviated from the communication strategy.

Table 4 Field protocol. Due date was Tuesday 1 March 2016.

Intervention	Date	Weekday	Time since end of reporting year, t^* (days)	n_1
	1 January 2016	Fri	1	8362
Appeal†	17 February 2016	Wed	48	5395
Motivational call‡	7–21 March 2016	Mon–Mon	67–81	1475
Reminder†	31 March 2016	Thu	91	893
Enforcement§	1 June 2016	Wed	153	253

†Assuming mail delivery in 1 business day

‡Each trader receives only one motivational call but for logistical reasons calls are spread over a 15-day period on business days only

§Enforcement is treated as a single event, but is actually a process that may take months to years

We start out by constructing a life table:

$$n_t = n_{t-1} - r_{t-1} - c_{t-1},$$

$$q_t = \frac{r_t}{n_t},$$

where n_t is the number of units (traders) at the start of time step t that can respond during time step t, r_t the number of units that responds during time step t, c_t the number of units that is censored at the end of time step t, e.g. because they receive an intervention the next time step, and q_t the response rate or the fraction of potential responders that actually responds at time step t. The initial number of units, n_1 is the initial sample size (first dataset) or the number of units subject to an intervention (subsequent datasets) (Table 4).

We then model the response rate q_t with a generalized linear model, assuming that the number of respondents is binomially distributed ($r_t \sim Bin(n_t, q_t)$) and that the complementary log-log of the underlying response rate is a linear function of explanatory features:

$$\operatorname{cloglog} q_t = \log(-\log(1 - q_t)) = X_t\beta,$$

where X_t is a p-dimensional row vector of covariates and β a p-dimensional column vector of coefficients. For $p = 3$, $X_t = (1 \quad g(t) \quad g(t)^2)$, where $g(t)$ is a transformation function of time: $g(t) = t$, $g(t) = \sqrt{t}$ or $g(t) = \log t$. The link function is chosen because it equals the log of the hazard function. The quadratic term in X_t and transformation of t allow for a more flexible relationship, including a non-monotonic one. AIC was used to select p and $g(t)$, i.e. the model with the optimal balance between parsimony and fit.

The model is fit to each of the five datasets until the next intervention is imposed (referred to as the 'short dataset') and extrapolated beyond this time point in order to estimate response rates of the lacking control group. The model fit to the last dataset (enforcement) does not have to be extrapolated since there is no next intervention. Unintentionally, some units did not receive any follow-up actions (see alluvial plot in Fig. 2). The model was therefore refit to the dataset including the serendipitous observations beyond the intervention (referred to as the 'long dataset'). The assumption here is that these units were unintentionally not exposed to the intervention, i.e. they did not differ structurally from units that were. No extrapolation of this model is required, although the model still relies heavily on data before the intervention.

Through backtransformation, the GLM yields estimated response rates:

$$\hat{q}_t = 1 - e^{-e^{X_t\hat{\beta}}}.$$

Next, the estimated response rates can be translated into the estimated cumulative response rate $\hat{F}_t$, which is the probability that a unit responds before or during time step t. It is the complement of 'survival' $\hat{S}_t$, which is the probability that a unit is still a non-respondent at the end of time step t:

$$\hat{F}_t = 1 - \hat{S}_t,$$

$$\begin{aligned}\hat{S}_t &= \prod_{k=1}^{t} (1 - \hat{q}_k) \\ &= \prod_{k=1}^{t} \left(e^{-e^{X_k\hat{\beta}}}\right) \\ &= e^{-\sum_{k=1}^{t} e^{X_k\hat{\beta}}}.\end{aligned}$$

Note that $\hat{F}_t$ only equals the fraction of the initial sample size that responds, $\sum_t r_t/n$ if all units in the initial sample have the opportunity to respond. If there is censoring (see light gray streams in

Fig. 2), this is not the case and the response rate $\sum_t r_t/n$ will underestimate the potential of the communication strategy.

The confidence interval of $\hat{F}_t$ is derived through Taylor linearization (see Appendix 2).

Finally, the cumulative response at time step t^* is estimated by stacking the cumulative response rates estimated from each dataset:

$$\hat{F}^{\text{stacked}}_{t^*-1+t} = \hat{F}^{\text{stacked}}_{t^*-1} + \left(1 - \hat{F}^{\text{stacked}}_{t^*-1}\right)\hat{F}_t.$$

5.2 Results

The observed and modeled response rates over time are given for each of the five datasets in Figures 3 through 7. The red curves show the models based on observations until the next intervention. Beyond this time point the predicted response rates are extrapolations mimicking the missing control group. The blue curves show the models based on all available data, including the units unintentionally not receiving any follow-ups. These models do not require extrapolation, but have to assume that the extra units are a random sample of the population that should have been followed up. Both models selected the logarithmic transformation of t ($g(t) = \log t$) and estimated an intercept, linear effect and quadratic effect ($p = 3$), unless stated otherwise. The consequences for our ultimate aim—the effects of interventions on the cumulative response—will be presented later (Fig. 8).

When the reporting year ends, the response rate initially increases, peaks at about 1.3% per day around t_{19} and then gradually decreases (Fig. 3). Response rates eventually approach zero according to both models, although the red model underestimates the speed of decline according to the blue model. The zero asymptote means that without any intervention, the cumulative response will never reach 100%, which makes sense in studies on response-enhancing interventions. The unintentional observations after the first intended intervention at t_{48} suggest that the response rate might peak again just before the due date at t_{61}, but the behavior cannot be seen by the red model, and the number of observations is too small to justify a more complex blue model. An experiment could shed more light on this (Tuttle et al., 2018).

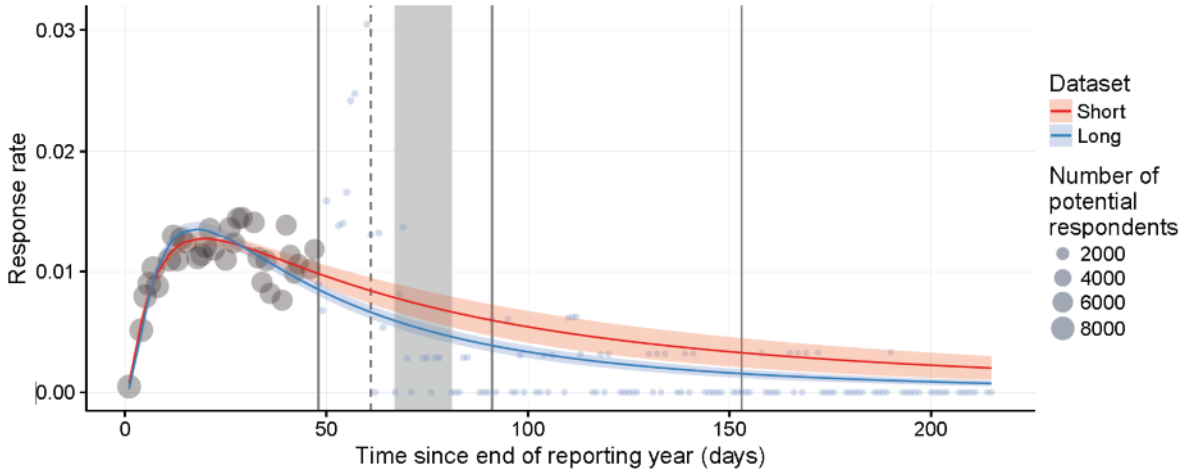

Figure 3 Response rate over time since end of reporting year. GLM (mean ± 1.96 SE) without (red) or with (blue) units that unintentionally did not receive an intervention (vertical solid lines and gray rectangle). Bubble area is proportional to denominator. Dashed line is due date.

When non-respondents receive an appeal letter two weeks before the due date, the response rate initially increases, reaching 8.4% on the due date (Fig. 4). The models do not distinguish between pre and post due date and predict a somewhat lower and earlier peak, although still considerably higher than expected without an appeal letter (compare Fig. 3 from t_{48} onwards). After the due date, the response rate quickly drops again. The red model prefers the square root transformation of t ($g(t) = \sqrt{t}$). According to the blue model, the red model now overestimates the speed of decline. Note that the number of potential respondents (bubble area) only gradually diminishes from t_{20} onwards, because the motivational calls are spread out over a 15-day period (gray rectangle).

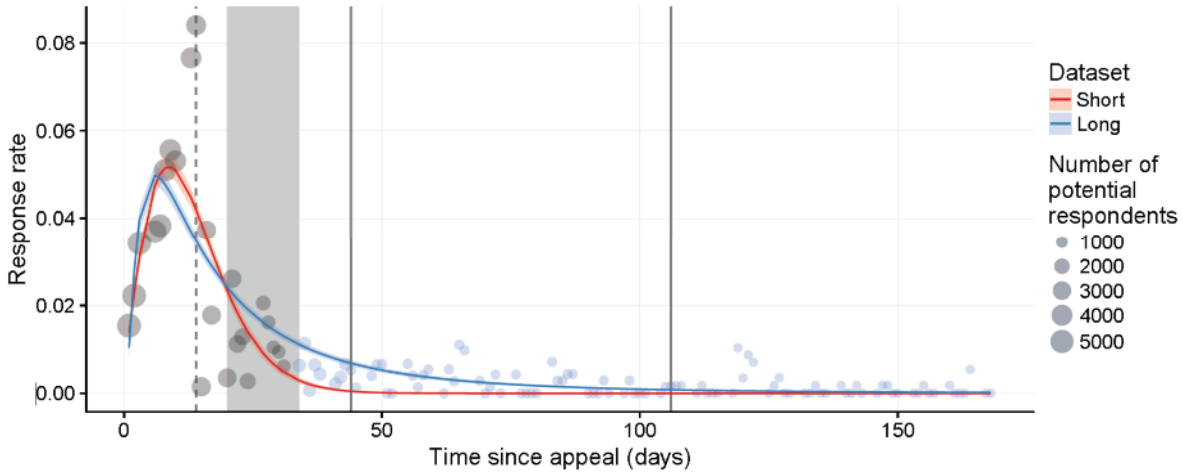

Figure 4 Response rate over time since receiving a pre-due date appeal letter. GLM (mean ± 1.96 SE) without (red) or with (blue) units that unintentionally did not receive an intervention (gray rectangle and vertical solid lines). Bubble area is proportional to denominator. Dashed line is due date.

On the day non-respondents receive a motivational call, the response rate peaks at about 3.9% (Fig. 5), considerably higher than expected without this intervention (compare Fig. 4 gray rectangle t_{20}–t_{34}). The effect directly fades away, without an initial increase. The models do not differ much about the speed of decline and agree about not having to transform t ($g(t) = t$). The red model considers the t^2 term insignificant ($p = 2$), but the blue model includes a third parameter for this effect. Note

that the number of potential respondents (bubble area) only gradually diminishes from t_{10} onwards, because the time between the motivational call and arrival of the reminder letter varies from 10 days (when called on the last calling day) to 24 days (when called on the first calling day) (left gray rectangle).

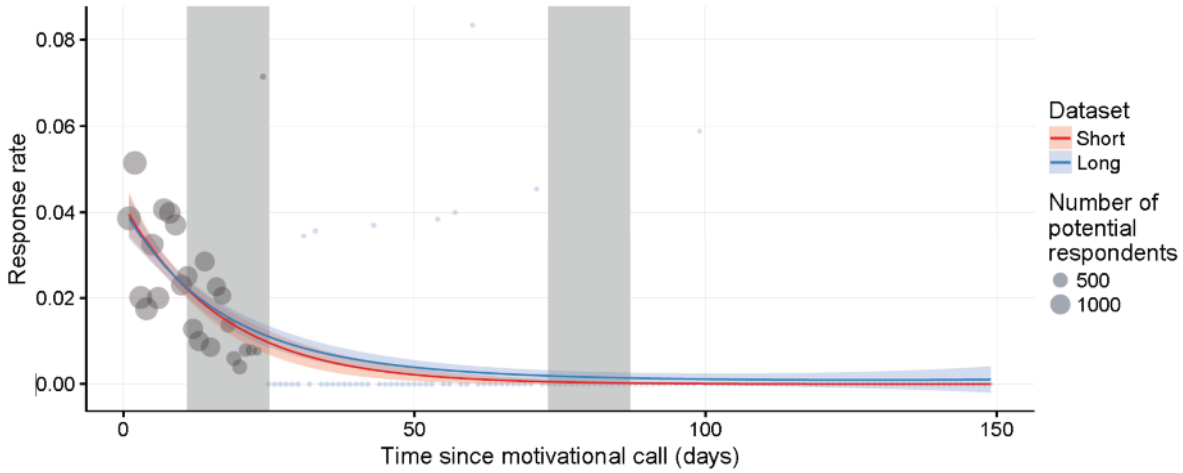

Figure 5 Response rate over time since receiving a motivational call. GLM (mean ± 1.96 SE) without (red) or with (blue) units that unintentionally did not receive an intervention (gray rectangles). Bubble area is proportional to denominator.

When non-respondents receive a reminder letter 10 to 24 days after the motivational call, the response rate initially increases, reaching 4.3% on t_6 (observed) or 3.2% on t_5 (red model) (Fig. 6). This is a moderate effect compared with the predicted response rates without a reminder letter (compare Fig. 5 gray rectangle t_{10}–t_{24}). The blue model has an even lower peak but might be compromised by three outliers at t_{76}–t_{78}. No explanation could be found for these suspiciously high response rates of over 3%.

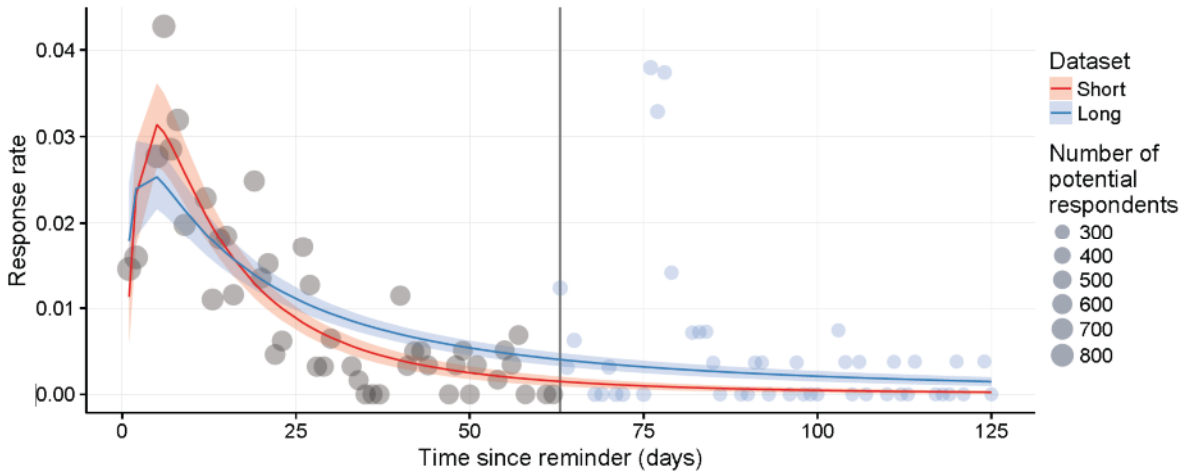

Figure 6 Response rate over time since receiving a post-due date reminder letter. GLM (mean ± 1.96 SE) without (red) or with (blue) units that unintentionally did not receive enforcement (vertical line). Bubble area is proportional to denominator.

Although enforcement starts officially at 1 June 2016, it is a process that shows a delayed effect only after about two weeks (Fig. 7). Nevertheless, the response rate reaches levels slightly above those predicted without enforcement (compare Fig. 6 from t_{63} onwards). The margins are wide because

only 253 traders are left at this stage and enforcement is treated as a single event, whereas in reality it consists of several events. Note that there is only a single model using all available data, as there is no subsequent intervention whose control needs to be modeled.

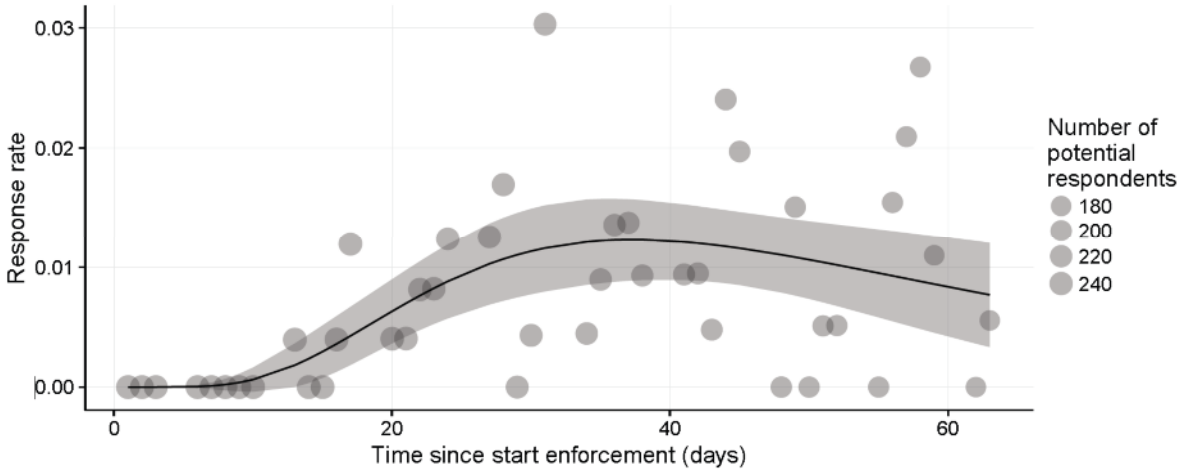

Figure 7 Response rate over time since the start of enforcement. GLM (mean ± 1.96 SE). Bubble area is proportional to denominator.

We can now construct the stacked cumulative response (Fig. 8), both observed (bubbles) and modeled (curves), including the four constructed controls. Note that color now represents the interventions instead of the two models using the short or long dataset. These two model versions are now in the column panels. The row panels become relevant when discussing the effect of the motivational call.

As noted earlier, the two models disagreed about the speed of decline in response rates (see Figs. 3 and 4). As a result, the model using the short dataset predicts that the appeal letter will have no long-term effect on the cumulative response rate (t^*_{215}: 64% vs. 60%), whereas the model using the long dataset does predict a large long-term effect (t^*_{215}: 71% vs. 50%). Fortunately, both models agree on a strong positive short-term effect of the appeal letter (t^*_{67}: 60%−38% = 21%-point according to the model using the short dataset and 58%−37% = 21%-point according to the model using the long dataset).

To assess the effect of a motivational call (after the previous intervention), we assume that all calls are made either on the first day of the calling period at t^*_{67} (top panels) or on the last day of the calling period at t^*_{81} (bottom panels). In the first case (top panels), the effect of the appeal letter has not died down yet. The short-term effect (t^*_{91}) of a motivational call is then estimated to be 76%−64% = 12%-point according to the model using the short dataset and 75%−67% = 8.3%-point according to the model using the long dataset. The long-term effect (t^*_{215}) of a motivational call is then estimated to be 79%−64% = 15%-point according to the model using the short dataset and 82%−71% = 11%-point according to the model using the long dataset. When the calls would be postponed (bottom panels), the effect of the previous intervention (the appeal letter) is exploited more. It increases the estimated long-term effect by 3%-point (to 18%-point and 14%-point, depending on the dataset).

The effect of the reminder letter (after the previous two interventions) is small: only 5.3%-point (short dataset) to 4.4%-point (long dataset) at t^*_{215} if the reminder letter would arrive 24 days after

the motivational call (top panels). If the reminder letter would arrive only 10 days after the motivational call (bottom panel), the effect would become insignificant: 1.4%-point (short dataset) to 1.0%-point (long dataset). Postponing the motivational calls is therefore only effective if the reminder letter is also postponed. The increase in cumulative response needs to be weighed against the delay in receiving the data.

Finally, enforcement (after the previous three interventions) might add another few %-points, but the effect is small and the uncertainty large.

Overall, if all interventions are applied, the cumulative response reaches about 89%, irrespective of the dataset used to fit the models and the timing of the motivational calls. This is almost 20%-points higher than the reported 71% based on the fraction of the initial sample responding ($\sum_s r_s/n$), because not all units in the initial sample experience the response-enhancing interventions (see Fig. 2).

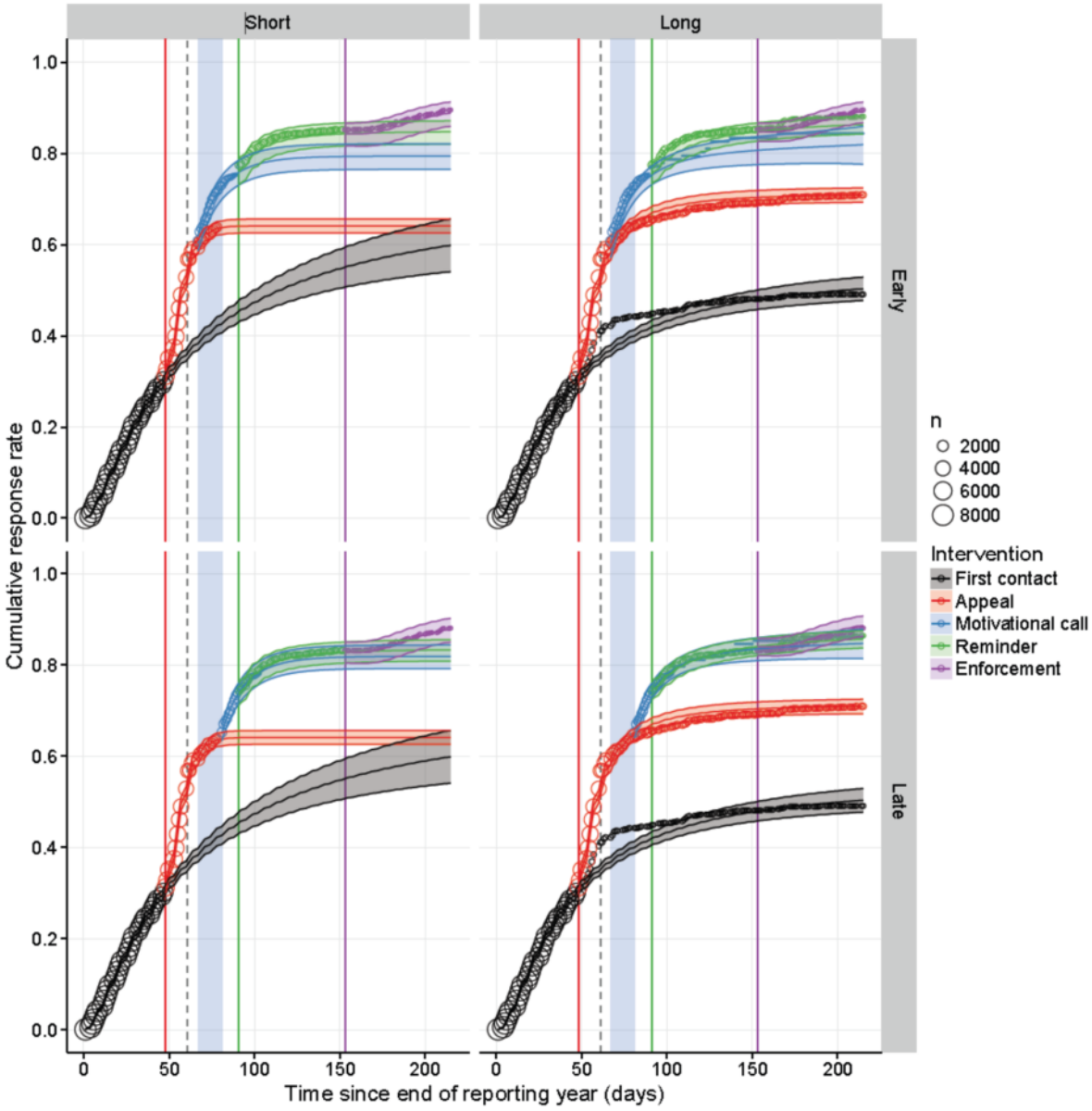

Figure 8 Stacked cumulative response over time since end of reporting year, observed (bubbles) and expected by GLM (mean and 95% CI) (curves). GLMs were fit without (left panels) or with (right

panels) units that unintentionally did not receive an intervention (vertical solid lines and rectangle). GLMs based on third dataset stacked early at t_{67} (upper panels) or late at t_{81} (lower panels). Bubble area is proportional to the number of potential respondents in the denominator, n. Dashed line is due date.

6. Conclusions and discussion

Based on the results as discussed in Sections 4 and 5 we can conclude that:

- It is difficult to draw one general conclusion from the numbers. The strategies as a whole work quite well for all surveys: final response rates are over 70%, with exception of the Survey on Grassland Usage (Table 2). The strategies are effective, in the sense that high response rates are achieved; however, some strategies can be more cost-efficient. This would require more in-depth research into the response process within businesses, in order to better tailor the strategy (see below).
- Each step in the communication strategy adds to the final response rate.
    - In general, Table 2 shows that additional steps will increase response rates, but with each next step the added response rate is decreasing. Exceptions here are the Structural Business Surveys 2015 and 2016 for which the last step adds about 20%-points to the final response rate (which is more than a quarter of the total response). A picture of the strategy that you would generally want are those from the Surveys on Investments, ICT Usage, and ISGVC, in which the added response rates in each next step still adds substantially to the total, but at diminishing rate. Also for these surveys the costs per unit response are lowest: varying from 1.36 to 2.03 euros, which is below average.
    - Also from Table 2 we can conclude that each step adds to the final response rates: in general the response probabilities for each step are between 0.15 and 0.40, but the variability is high. This variability may be explained by the variability in duration of each measure $\mathrm{d_s}$: the correlation between $\mathrm{p_s}$ and $\mathrm{d_s}$ is .40, indicating that a longer duration results in a higher response fraction. This relationship is shown in Figure A1.1. However, as we have seen from the survival analysis for IGT, this relationship is not linear: after some time the effect of a measure drops to zero. Results from studies in the USA (Willimack et al., 2018) also show this. A duration for each measure that may work well is about 2 weeks (14 days).
    - Apart from the duration, the variability in the response probabilities for each step may also be explained by survey characteristics (like information asked for; e.g. if the topic is of interest to businesses, like ISGVC (Vos and Snijkers, 2018)), and the internal response process in businesses (e.g. whether the data are available at the moment of due date, like for the SBS). This would indicate that the processes within businesses to get response may be different for the various surveys, and may also change during the field work period, which again would require more insights in these processes to better tailor the strategy. This would require response rate analysis of previous cycles aimed at identifying subgroups.
    - Finally the effectiveness of a measure depends on the how a measure is executed, i.e. flaws in the internal processes. For the ITG Survey we have seen that the final response rates could have been much higher.
- Most effective with regard to achieving high overall response rates are the measures early in the process: *the advance letter, the pre-due date reminder, and the first reminder*. Table 2

shows that these measures have high response probabilities (p_s), and for these steps the highest response rates (rr_s) are gained (again with exception of SBS2015 and SBS216).

- As for the *pre-due date reminder* the results in our studies (as described in Sections 4 and 5) are in line with experimental studies carried out in the USA. Tuttle and his colleagues (2018; Langeland and Tuttle, 2016) conclude that a pre-due date reminder accelerates response meaning that response is coming in at a faster rate. With a pre-due date reminder final response rates will be equal or higher than the final level for a strategy without such a measure. Tuttle et al. (2018; Langeland and Tuttle, 2016) conclude that at the end of fieldwork period strategies without a pre-due date reminder have somewhat (but still significantly) lower response rates. *So, if you want to get your responses fast, use a pre-due date reminder, but this will be at the costs of contacting more businesses.*
  The pre-due reminder, however needs to be designed well. We have some evidence from telephone calls that some businesses perceive this as a reminder, and feel unnecessarily chased. They have scheduled the completion of a questionnaire close to the due date, because at that time the data will be available, and/or then they have time to complete the questionnaire, or for some other reason.
- Post-due date reminders all have an effect. This is in line with other research, showing additive effects from multiple sequential contacts (Dillman et al., 2014).
- *Telephone reminding* is effective for some surveys, but for others it is not. Table 2 shows that the added response rates for telephone are very low. This has to do with the fact that for this measure a small number of businesses is selected. Without any further insights in the characteristics of these businesses, and the effect of these additional responses on the final statistical estimates of target variables, we could conclude that telephone reminding in these business surveys is not effective, and only adds to the costs. But, very likely these businesses are needed as relevant cases in order to achieve unbiased estimates.

Going back to the questions we started with, we can now conclude that:

- An effective strategy would consist of a pre-due date reminder, and at least one post-due date non-response follow-up. If time is an issue, it seems worthwhile to make an effort early in the process to get high response rates. Direct actions have an effect: strike and strike again! Each step can be seen as a nudge to get response (Thaler and Sunstein, 2009). If time is no issue, a pre-due date reminder can be skipped. If no pre-due date reminder is used, a post-due date reminder immediately after the due date would seem an appropriate strategy (Tuttle et al., 2018; Langeland and Tuttle, 2016).
- A cost-level of 2 euros per unit response or below could be considered as a cost-efficient strategy.
- Have enough time between the measures: if the time interval is to short the effect will be small, if the time interval is too large, the measure will long be exhausted. An interval of two weeks seems appropriate.

We should however keep in mind that, as said in Section 4, that we did not study the effects of these measure independently of each other: the results of a next step are conditional on previous steps. The response rates are achieved in a sequence of measures. And furthermore, we only looked at annual surveys with in general one particular contact strategy. We did study other strategies.

With regard to these conclusions, also some additional issues need to be taken into account (Snijkers et al., 2013):

- In this analysis we did not look at background characteristics of businesses, like size class, NACE (NAICS) code, panel members (being part of the survey panel, and sampled for every cycle of the survey), multi-surveyed (whether a business receives more than one survey), and the relevance of businesses with regard to the target variable (which relates to being part of the survey panel). This is related to the representativeness of the response distribution for these auxiliary variables.
- We only looked at unit return rates, not at weighted rates, like the 'revenue' response rate for the SBS and the 'trade value' response rate for the International Trade in Goods Survey. When including these weights, the added value for telephone reminding may show a different picture. By looking at the weighted response rates we take the relevance of businesses into account and monitor for bias in the final estimates.
- And finally, to be absolutely sure about the bias and precision of the final estimates for target variables, we need to look at the development of the estimates in relation to the communication strategy.

To get response in business surveys, however, more factors play a role to get a cost-efficient communication strategy. As already stated above, we recommend tailoring the communication strategy to the business context. In order to do so, it is necessary to study the response process within businesses. Getting response depends on a number of factors:

- Complexity of the reporting process: do the data need to come from different people and different departments within a business. This correlates with business size: for small businesses the reporting process is easy, for middle-size businesses the process is more complex but manageable, for the large and very large business this process can be very complex (Snijkers et al., 2013; Snijkers, 2016).
- Timing of the survey request. Whether businesses are able and willing to complete a questionnaire also depends on the moment of dispatching a questionnaire: are the data available and do people within businesses have the time to complete the questionnaire.
- Previously and multi-surveyed: businesses that have been surveyed before have established a process. Research (Snijkers et al., 2013) shows that business that have participated in surveys before are more likely to comply in the future.
- Other factors also play a role:
  - If a survey is mandatory, like we have seen for the agricultural surveys.
  - The topic of a survey, as is the fact for the ISGVC (Vos and Snijkers, 2018).
- Fancy and indirect measures like flyers do not have much of an effect, although they may have an effect on the general image of Stats Netherlands (Houben et al., 2018; Willimack et al., 2018).
- Finally, it is also the organization of the internal processes in survey organisations that effect the cost-efficiency of a strategy, and whether a strategy is exploited to the full.

References

Dillman, D.A., 2000, *Mail and Internet Surveys: The Tailored Design Method, 2nd Edition*. Wiley, New York.

Dillman, D.A., J.D. Smyth, and L.M. Christian, 2014, *Internet, Phone, Mail and Mixed-Mode Surveys:The Tailored Design Method, 4th Edition*. Wiley, Hoboken.

Geurden-Slis, M., and G. Snijkers, 2017, The effect of communication measures for the Survey on International Trade in Goods. Paper presented at 5th European Establishment Statistics Workshop, 30 August – 1 September 2017, University of Southampton, UK.

Groves, R. M., and M.P. Couper, 1998, Nonresponse in Household Interview Surveys, Wiley, New York.

Groves, R. M., D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), 2002, Survey Nonresponse, Wiley, New York.

Houben, L., W. Groot, D. Debie, G. Goris, A. Opreij, X. Geers, and G. Snijkers, 2018, A continuous search to improve targeted communication in business surveys. Paper presented at the 5th International Business Data Collection Methodology Workshop, 19-21 September 2018, Lisbon.

Langeland, J., 2018, Evaluating Mode Sequence when E-mail is used as the initial contact in establishment surveys. Paper presented at the 5th International Business Data Collection Methodology Workshop, 19-21 September 2018, Lisbon.

Langeland., J., and A. Tuttle, 2016, Improving Data Collection: Results from two Contact Strategy Experiments in Economic Surveys. Paper presented at the 5th International Conference on Establishment Surveys, 20-23 June 2016, Geneva.

Luiten, A., 2013, Improving Survey fieldwork with paradata. Doctoral dissertation, Utrecht University. Statistic Netherlands, Heerlen.

McCarthy, J.S., and M. Earp, 2012, Non-response Reduction Methods in Establishment Surveys. Introductory Overview Lecture (IOL) presented at the 4th International Conference on establishment Surveys (ICES4), 11-14 June 2012, Montreal.

Sakshaug, J., and B. Vicari, 2018, The Impact of Invitation Mode in on On-line Establishment Survey. Paper presented at the 5th International Business Data Collection Methodology Workshop, 19-21 September 2018, Lisbon.

Snijkers, G., 2016, Achieving Quality in Organisational Surveys: A Holistic Approach. In: Liebig, S., and W. Matiaske (eds.), Methodische Probleme in der empirischen Organisationsforschung: pp. 33-59 (in English). Springer, Wiesbaden.

Snijkers, G., G. Haraldsen, J.Jones, and D. Willimack, 2013, Designing and Conducting Business Surveys. Wiley, Hoboken.

Stoop, I., 2005, The hunt for the last respondent; non-response in sample surveys. Doctoral dissertation, University of Amsterdam. Social and Cultural Planning Office in the Netherlands, The Hague.

Stoop, I., J. Billiet, A. Koch, and R. Fitzgerald, 2010, Improving Survey Response: Lessons learned from the European Social Survey.

Thaler, R.H., and C.R. Sunstein, 2009, Nudge; Improving decisions about health, wealth, and happiness. Penguin Books, London.

Tuttle, A.D., J.L. Beck, D.K. Willimack, K.P. Tolliver, A. Hernandez, and C.-C. Fan, 2018, Experimenting with Contact Strategies in Business Surveys. Journal of Official Statistics (special ICES5 issues) 34(2): 365-395.

Vos, S., and G. Snijkers, 2018, Evaluation of the International Sourcing and Global Value Chain Survey 2017 – Visits to enterprises in the Netherlands. Report, Statistics Netherlands, Heerlen.

Willimack, D.K., et al., 2018, From Experimentation to Implementation: Putting the pieces together to form a cohesive strategy for the U.S. Economic Census. Paper presented at the 5th International Business Data Collection Methodology Workshop, 19-21 September 2018, Lisbon.

Appendix 1. Figures A1.1 and A1.2

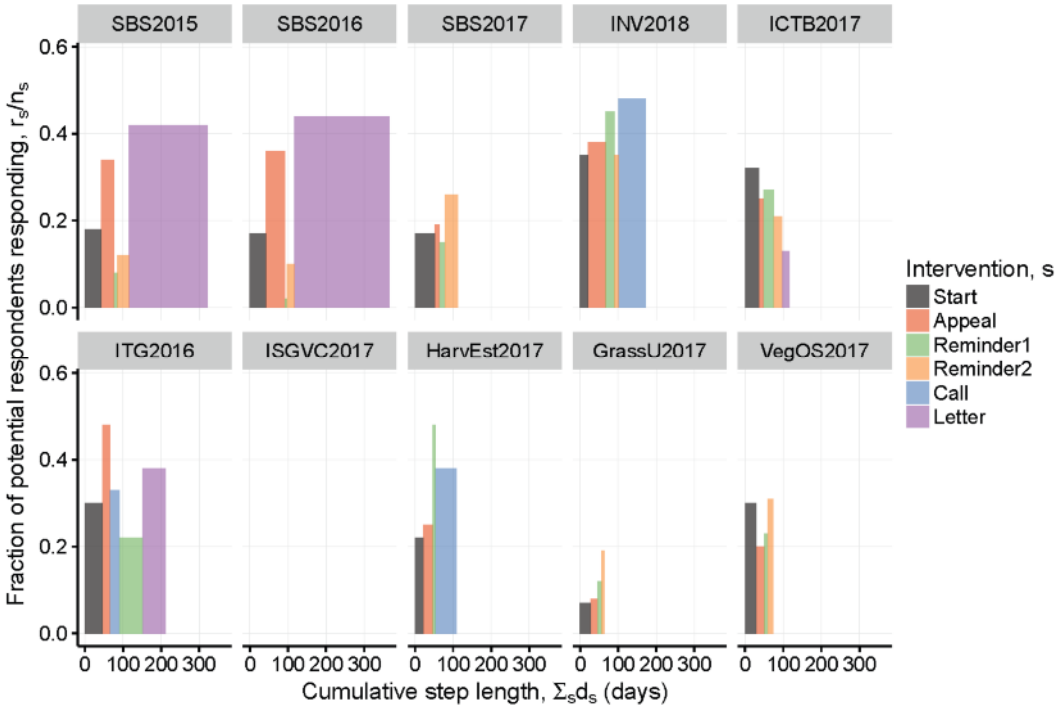

Figure A1.1 Fraction of potential respondents at the beginning of step s that responds during step s in relation to the number of days a step takes, for different surveys (panels). In ITG2016 the call becomes before the first reminder. d_s is missing for all steps in ISGVC and for the last step (call) in GrassU and VegOS.

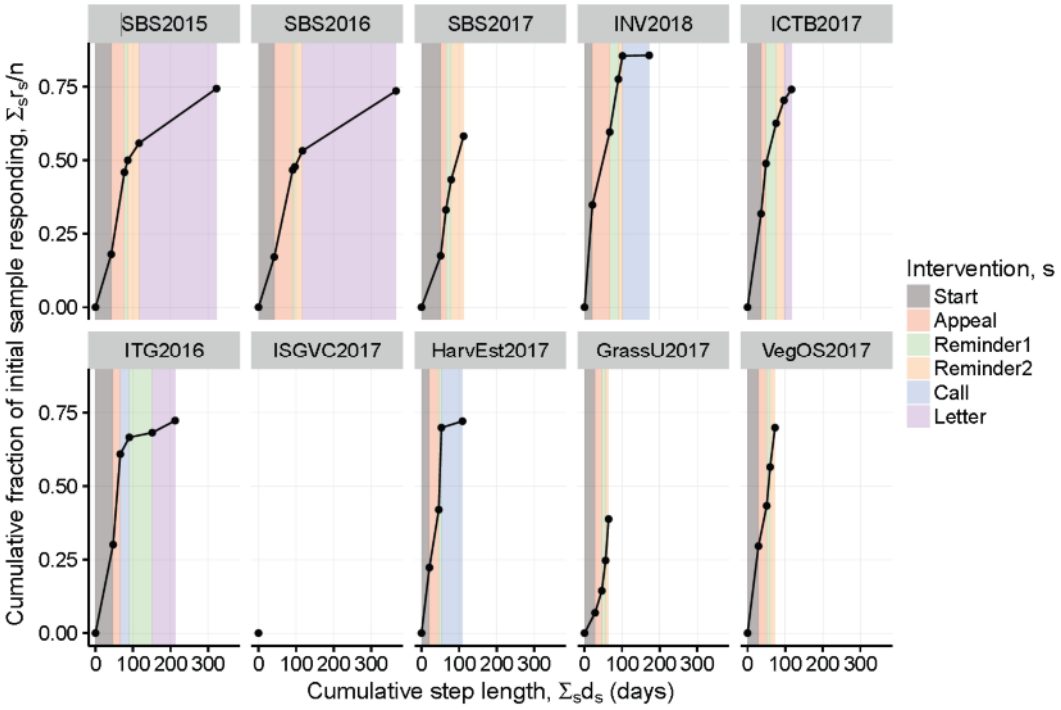

Figure A1.2 Cumulative fraction of the initial sample responding before or during step s, for different surveys (panels).

Appendix 2. Derivation of the confidence interval of the cumulative response

A nonparametric $(1-\alpha)100\%$ confidence interval of $\hat{F}_t$ could be constructed using the exponential Greenwood (Kalbfleisch & Prentice 1980):

$$1-e^{-e^{\operatorname{cloglog}\hat{F}_t \pm z_{\alpha/2}\sqrt{\hat{V}(\operatorname{cloglog}\hat{F}_t)}}},$$

where the variance $\hat{V}\left(\operatorname{cloglog}\hat{F}_t\right) = \frac{1}{(\log(1-\hat{F}_t))^2}\sum_{k=1}^{t}\frac{r_k}{n_k(n_k-r_k)}$. However, since it is nonparametric, it cannot be used for extrapolation. We therefore derive an analytical expression to approximate the confidence interval of $\hat{F}_t$ estimated by a parametric model.

Recall a first-order Taylor linearization of $f(x,y)$ in (a,b):

$$f(x,y) \approx [f]_{a,b} + \left[\frac{\partial f}{\partial x}\right]_{a,b}(x-a) + \left[\frac{\partial f}{\partial y}\right]_{a,b}(y-b).$$

Approximating $\hat{S}_t$ by a first-order Taylor linearization of $\hat{S}_t\left(\hat{\beta}\right)$ in β yields:

$$\hat{S}_t \approx S_t + \sum_{i=1}^{p}\left[\frac{\partial \hat{S}_t}{\partial \hat{\beta}_i}\right]_{\beta}(\hat{\beta}_i - \beta_i),$$

where

$$\left[\frac{\partial \hat{S}_t}{\partial \hat{\beta}_i}\right]_{\beta} = e^{-\sum_{k=1}^{t}e^{X_k\beta}}\left[-\sum_{k=1}^{t}X_{ki}e^{X_k\beta}\right] = a_i S_t.$$

Here, $a_1 = -\sum_{k=1}^{t}e^{X_k\beta}$, $a_2 = -\sum_{k=1}^{t}g(k)e^{X_k\beta}$ and $a_3 = -\sum_{k=1}^{t}g(k)^2 e^{X_k\beta}$.

The variance of $\hat{S}_t$ is then:

$$\begin{aligned} V(\hat{S}_t) &= V\left(S_t + \sum_{i=1}^{p}a_i S_t(\hat{\beta}_i - \beta_i)\right) \\ &= 0 + S_t^2 V\left(\sum_{i=1}^{p}a_i\hat{\beta}_i\right) \\ &= S_t^2\sum_{i=1}^{p}\sum_{j=1}^{p}a_i a_j C(\hat{\beta}_i,\hat{\beta}_j), \end{aligned}$$

which is estimated by:

$$\hat{V}(\hat{S}_t) = \hat{S}_t{}^2\sum_{i=1}^{p}\sum_{j=1}^{p}\hat{a}_i\hat{a}_j\hat{C}(\hat{\beta}_i,\hat{\beta}_j),$$

where $\hat{C}\left(\hat{\beta}_i,\hat{\beta}_j\right)$ is the estimated variance-covariance matrix of the estimated regression coefficients.

The $(1-\alpha)100\%$ confidence interval of $\hat{F}_t$ is given by:

$$1-\hat{S}_t \pm z_{\alpha/2}\sqrt{\hat{V}(\hat{S}_t)}.$$