Journal of Official Statistics, Vol. 24, No. 2, 2008, pp. 301–318

An Experiment on Perceived Survey Response Burden Among Businesses

Dan Hedlin[1], *Helen Lindkvist*[2], *Helena Bäckström*[1], *and Johan Erikson*[2]

In an experiment embedded in a business survey, subsamples were randomized to obtain none, both or either of two motivation-enhancing enclosures that went out with the covering letter: a sheet explaining the purpose of the survey and another sheet informing about feedback of survey results. All businesses obtained a questionnaire on perceptions of the survey. The main conclusion is that perceptions of the usefulness of the survey are improved by the enclosures while there is no evidence of an effect on perceptions of how burdensome the survey is.

Key words: Embedded experiment; official economic statistics; ordinal logistic regression; motivating respondents in business surveys.

1. Introduction

1.1. Response Burden

Response burden in statistical business surveys has come under intense scrutiny in recent years. Survey methodologists are concerned with response burden as a survey quality issue under the assumption of a negative association between burden and quality. In many countries there have since the 1970s been growing political concerns about response burden imposed on businesses, concerns which seem to have become more strident during the last ten years or so. National Statistical Institutes (NSIs) are in most countries obliged to monitor response burden incurred on businesses. However, measurement of response burden tends to focus on response time although response burden *as perceived by respondents* is not determined by time alone. In the terminology of e.g., Jones, Rushbrooke, Haraldsen, Dale, and Hedlin (2005) response burden policies have in the past focused on actual response burden rather than perceived response burden.

It has also long been recognized that a subjective response burden concept should be two-sided, with perceived burden on the one hand and perceived reward on the other. The reward may be tangible or intangible, e.g., a token financial incentive or a sense of fulfilment of one's civic or a business's corporate duty.

[1] Statistics Sweden, Box 24 300, SE-104 51 Stockholm, Sweden. Emails: dan.hedlin@scb.se; helena.backstrom@scb.se
[2] Statistics Sweden, SE-701 89 Örebro. Emails: helen.lindkvist@scb.se and johan.erikson@scb.se
Acknowledgments: This experiment was part of a project on perceived response burden in which the UK Office for National Statistics, Statistics Norway and Statistics Sweden took part. We thank our colleagues at all three agencies for their input, advice and help, and also the associate editor and three referees for making many constructive suggestions for improvement. This research has been partly funded by Eurostat (ESTAT 200292501005).

In focus groups and cognitive interviews it has been noticed that many business survey respondents do not understand the purposes of the surveys. Many of them are not familiar with, or concerned about, the statistics that the surveys produce (e.g., Willimack, Nichols, and Sudman 2002, Hedlin, Dale, Haraldsen, and Jones 2005). It is reasonable to believe that there is an association between respondents' knowledge and appreciation of the surveys and their perceptions of response burden; that is, the more they know, the less the perceived burden. Similar points have been made by Gower (1994), Tomaskovic-Devey, Leiter, and Thompson (1994), Couper, Singer, and Kulka (1998), Willimack et al. (2002), Haraldsen (2004) and others, although Willimack et al. seem to take rather a pessimistic view of the potential of raising the level of motivation among business survey respondents. In contrast, Haraldsen (2004) puts forward a model of survey burden that in terms of respondent characteristics focuses on the respondent's interest in the survey topic, time available, willingness to make an effort to respond to the survey and whether the respondent feels able to answer the questions properly. Haraldsen contends that these traits or circumstances can to some extent be influenced by survey design. In Haraldsen's model, perceived burden and reward are the result of the encounter between a survey request and the respondent. Data quality (including nonresponse in some cases) is seen as the outcome of this encounter. Hence a better balance between burden and reward should improve data quality.

1.2. The Response Process

It is helpful to evoke a model of the response process at a business to understand sources of response burden (Willimack and Nichols 2001; Jones et al. 2005):

1. Encoding in memory/record formation.
2. Selection and identification of the respondent or respondents.
3. Assessment of priorities.
4. Comprehension of the data request.
5. Retrieval of relevant information from memory and/or existing company records and/or other persons.
6. Judgement of the adequacy of the response.
7. Communication of the response.
8. Release of the data.

Haraldsen (2004, Figure 4) connects his burden model with this model of the response process.

Many business survey respondents open the envelope from the NSI with adverse preconceptions about what they are going to find. Unfortunately, one of the early steps in the response process is their realizing what information they need to retrieve to perform the (often mandatory) task of responding. The information retrieval is perceived as burdensome by many respondents (Hedlin et al. 2005).

1.3. Incentives, Appeals, and Motivation

Less arduous and frequent data requirements and more user-friendly questionnaires may be the factors that would reduce response burden the most. However, a respondent's overall perception is determined by both burden and reward. This article focuses on the

reward side, which in itself is a multi-faceted complexity. The user-friendliness of an inviting questionnaire could in itself be rewarding. As sources of respondent motivation Krosnick (1991) mentions the respondent's personal interest in the survey topic, perceived importance of the survey, her or his intrinsic rewards from mental exercises and cognitive activities, the number of questions on the questionnaire, and accountability, i.e., whether the respondent believes she or he has to justify her or his response. While the strength of several of these sources could be tested in experiments similar to the one reported here, we have concentrated on sources resembling the first two in Krosnick's account: interest and perceived importance.

Social exchange theory applied to the survey situation (e.g., Dillman 2007, pp. 14–27) is centered on three themes: cost, reward and trust. They are taken as wide concepts. Cost is what the respondent gives up or spends to respond. Reward is what she or he expects to gain. Trust is the respondent's belief that the rewards outweigh the costs in the long run. Note that the decision to participate or not is often heuristic rather than thoroughly considered. It relies on trust. In a repeated survey, trust may build up over time. The perceived reward may well be to the benefit of some "group" the respondent feels attached to. An appeal to the respondent's altruistic values can at least theoretically tip the balance. As an example of experimental appeals in the U.S. population census, we mention Dillman, Singer, Clark, and Treat (1996). However, they found that an altruistic appeal had no effect. As pointed out by Hak, Willimack, and Anderson (2003), Dillman's concept of cost overlaps the concept of perceived burden. They, too, contend that cost can be offset by perceived rewards.

Turning to other sources of motivation, Couper et al. (1998) suggest that willingness to respond to the U.S. population census is related to the "sense of attachment and civic duty toward the society at large" (p. 68). Their analysis of the 1992 census participation data corroborates this, but shows also that if a variable that summarizes attitudes towards the census is included in the model, more general attitudes towards government and the polity lose explanatory power. Social exchange theory may include the perceived reward one may experience from being well integrated into the community. Thus there is some overlap between the idea of "civic duty" as one driving force and social exchange theory. However, there is also a structuralist view of attachment to the polity in which some people are marginalized through their actual position rather than their attitudes.

Theories of what makes people respond have been formulated and tested mainly in surveys of individuals and households. It is not clear how they translate to business surveys. It is sometimes forgotten that respondents to business surveys are people. Mechanisms in surveys of individuals should therefore also be in force in business surveys, at least to some extent. On the other hand we play different roles when we work for a living than we do in our private lives. It is also true that what constitutes perceived rewards and sense of civic obligation is a more complex issue at the workplace where our personal stance may differ from that of our employer and colleagues. Also, some rewards may be more appreciated by somebody else at the business than by the actual respondent.

In the survey context the term incentive is often confined to denote something that appeals to the respondent's self-interest. Use of incentives in this sense has been investigated extensively in surveys of individuals and households but less so in business surveys. See references in e.g., Singer, Van Hoewyk, Gebler, Raghunathan, and

McGonagle (1999) and Shettle and Mooney (1999) on theories of what makes incentives effective. James and Bolstein (1992) report on one experiment with monetary incentives in a survey of small businesses. Appeals, as distinct from incentives, are also rare in business surveys research. See also the overview of business survey practices in Willimack, Lyberg, Martin, Japec, and Whitridge (2004). The business survey experiment reported here is novel in that it involves randomized treatments that allow measurement of motivational effects on self-perceived response burden.

1.4. Purpose of the Article

The hypothesis about an association between perceived burden and respondents' knowledge and understanding of the survey and its purposes has largely been formulated on the basis of observations made in focus groups and interviews with business survey respondents (see Willimack et al. 2002 and other references above). This article draws conclusions from data collected in a survey. The reasons why there may be a discrepancy between results obtained in a survey and in a focus group include the fact that the setting of a focus group is social, the fact that a survey is isolated and while that participants in surveys and focus groups tend to be selected differently (cf. O'Muircheartaigh 1999).

We wanted to see if we could alleviate some of the negative perceptions of the survey request among many respondents. In particular, it was of interest to see whether appealing to either businesses' self-interest with a nontoken incentive or respondents' interest and sense of responsibility (by explaining the use and importance of statistics) could be proven to change the views of the respondents.

We opted for a cost-effective way to operationalize these pursuits. In an experiment embedded in the annual Structural Business Statistics (SBS) survey conducted by Statistics Sweden, subsamples were randomized to obtain none, both or either of two enclosures that went out with the covering letter. Both enclosures were intended to motivate respondents. All businesses in the sample obtained a questionnaire on perceptions of the survey as well as the regular SBS questionnaire. We refer to the questions on perceived response burden as "*PRB questions*." These questions were constructed in a project on perceived response burden in which the UK Office for National Statistics, Statistics Norway, and Statistics Sweden took part. The PRB questions are analysed in Hedlin et al. (2005).

The next section discusses the experimental treatments and introduces the research questions. In Section 3 the SBS survey and the design of the experiment are described. Section 4 goes on to the rationale behind the response variables and model, while Section 5 reports on results. The article ends with a discussion in Section 6.

2. Treatments and Research Questions

Two enclosures were designed with a view to enhancing the motivation of the recipients to respond to the survey request. One of the enclosures targeted the understanding of why the survey is conducted while the other one appealed to the respondent's self-interest. The former enclosure was a sheet explaining the purpose of the survey and giving some facts obtained from the previous wave of the survey. It consisted chiefly of a description of how the results are used. They are for example important contributions to the National Accounts. Examples of conclusions drawn in previous waves of the SBS survey were also

given. These were given as short text passages together with three tables and one graph. One of the striking facts highlighted was the high turnover per employee among electricity producers: more than double that of the industry with the second-highest ratio. The electricity-producing industry also had a far larger year-on-year mean increase in turnover per employee than any other industry. The revenues in this industry were a topical issue at the time. This enclosure was intended to enhance both the respondent's personal interest in the survey and her or his perception of the importance of the survey (cf. Krosnick 1991, as discussed above). The enclosure was a folded four-page green sheet. In the sequel, we refer to this sheet as "*How the data are used*," which is an approximate translation of the Swedish title of the sheet.

The other enclosure was a beige sheet informing about the feedback of survey results that respondents could expect. All essential information was put on the front of the sheet, while contact details together with information on other key ratios that can be ordered from Statistics Sweden were on the back. The feedback allows respondents to compare their key ratios with those of the industry within the same size bracket. In total, there were 19 key ratios to be expected, including gross and net margin, return on adjusted equity, return on total assets, liquidity, return on working capital and rate of inventory turnover. A table on the front of the sheet gave the recipient an idea of what the feedback information would look like with the key ratios of the business in the first column, the lower quartile, the median and the upper quartile of the reference group in the next three columns. There were prominent overlaid text boxes that read, for example, "in this column you will find the key ratios of your business computed from the data you submit." We refer to this enclosure as "*Key ratios*."

Obviously, the key ratios could not be computed ahead of the response. So, unfortunately, it was not possible to adhere to the large body of evidence that shows that incentives are more effective if they are given without conditions to the prospective respondent (see the literature review in Singer et al. 1999; also Dillman 2007, pp. 167–170). For example, in the experiment of James and Bolstein (1992) a conditional promise of 50 USD to prospective business survey respondents had no significant effect on response rates while an unconditional 5 USD given up front increased the response rate by 30 percentage points after the first mailing as compared to that of the control group who received no incentive. However, feeding back survey results is different from offering money. This article sheds some light on how feedback compares to the types of incentive studied in the literature.

Our experiment concerned two research questions:

1. Can we influence respondents' appreciation of
   (a) their own (potential) use of the statistics
   (b) the importance of the statistics to society
   by enclosing supplementary information (either "Key ratios" or "How the data are used") with the covering letter?
2. Can we influence respondents' perceptions of the survey with the same supplementary information?

Four response variables were identified to inform on the research questions. They are discussed in Section 4.

3. Description of the Survey and the Experiment

The annual structural business statistics are central to Sweden's official economic statistics. The target population comprises all active businesses except those in the financial sector (NACE J 65–67 Financial intermediation). The survey is mandatory. This was stated in fine-print on top of the questionnaire and in the covering letter. The survey was reengineered in the period 2002–2004. The 2004 Survey was new in several respects, including the enlargement of the survey population to encompass also small businesses, a new name of the survey, a redesigned questionnaire and a change of data collection mode from paper to an electronic questionnaire. In 2004, about 800 of a selection of the largest businesses formed a self-representing stratum. These businesses were sent a long questionnaire in the spring. A Pareto πps sample (Rosén 1997) of some 8,000 businesses selected from the complementary part of the frame obtained a shorter questionnaire in the autumn. On this questionnaire tax records from the year before were preprinted. The respondents were asked to break down the tax data on all types of costs and turnover in some specified detail. The πps sample was split into two groups, the first of which comprised businesses with employees less than 20. They were sent a paper version of the questionnaire with preprinted tax data together with the covering letter. The second group, businesses with at least 20 employees, got a covering letter with information on how to download a computerized self-administered questionnaire (CSAQ) from a website. The 3,677 businesses that were included in the CSAQ part of the sample were also included in the experiment.

The businesses that responded to the PRB questions reported having spent on average about two hours on retrieving the information necessary to respond to the survey and about 70 minutes on completing the questionnaire. The lower quartile, median and upper quartile for retrieval were 30, 60, and 120 minutes, while these numbers were 30, 60, and 60 minutes, respectively, for completing the questionnaire. About 70% of the businesses reported that one person responded to the CSAQ single-handedly, for 22% there were two people taking part, while three or four people participated in the response process for 5% and 1% of the businesses, respectively.

The CSAQ was constructed in Microsoft Office Excel. On the first of four sheets there were general questions about the business. The second sheet was the main one for the SBS survey. The PRB questions were on the third sheet, with a statement at the top that this part of the CSAQ was voluntary. Finally, on the fourth sheet the businesses were invited to give further comments. An Excel macro produced a dialogue box with the help of which the respondents could save the completed questionnaire and transfer it to Statistics Sweden by pressing a button.

Prior to the experiment we expected a slightly negative association between perceived burden and size, since most large businesses have special staff and information systems in place that should facilitate their response. Hence the businesses included in the experiment were grouped in four blocks (strata) by number of employees. Within each block the businesses were randomized to the four treatments with sample sizes as given in Table 1. The businesses took part in the experiment without their knowing. Staff at Statistics Sweden working with the daily operation of the survey were not involved in the experiment and had no immediate access to the treatment codes.

Table 1. Number of businesses in sample by treatment and block (employment bracket)

Block	Employment bracket	Treatment				
		1. No special enclosure	2. The enclosure "How the data are used"	3. The enclosure "Key ratios"	4. Both enclosures	Total
A	20–29	209	209	209	208	835
B	30–49	260	260	260	259	1,039
C	50–99	219	219	219	218	875
D	100 +	232	232	232	232	928
Total		920	920	920	917	3,677

4. Response Variables and Model

To avoid the uncertainties of post hoc interpretation of data, four response variables as well as the analyses of these were specified ahead of the experiment (Wahlström 2004). Table 2 lists the prespecified response variables. The variables Y_1 and Y_2 are responses to the questions that are given in English in Table 2. The variable Y_3, "perceived burden," is a combination of responses to two PRB questions, "Do you think it was quick or time-consuming to collect the necessary information to complete the SBS questionnaire?" and "Did you find it easy or burdensome to complete the questionnaire?" Both questions have a five-point response scale. There was no "don't know" alternative. A respondent who has ticked very or quite quick (the first or the second option) at the first question *and* very or quite easy (the first or the second option) at the second one is defined as perceiving a light response burden while perceived heavy burden is defined analogously, i.e., if either the fourth or the fifth alternatives are ticked. In the case of item nonresponse for at least one of the variables, Y_3 is defined as item nonresponse. Other respondents are defined as "neutral."

Table 2. Prespecified response variables. Response categories in parentheses

Variable	Description and response categories
Usefulness for business (Y_1)	Response to "Do you think that the statistics based on the SBS survey are of great or little use to your business?" (Very useful, Fairly useful, Neither useful nor useless, Fairly useless, Very useless and Don't know)
Usefulness for society (Y_2)	Response to "Do you think that the statistics based on the SBS survey are of great or little use to society?" (Very useful, Fairly useful, Neither useful nor useless, Fairly useless, Very useless and Don't know)
Perceived burden (Y_3)	Combined response to two items on perceived response burden. (Light, Neutral and Heavy)
Punctual response to SBS (Y_4)	Whether the business responded to the SBS survey by 7 November 2004 or not (the Sunday after deadline). (Yes and No)

The variables Y_1 and Y_2 inform on research questions 1a and 1b, respectively; the variables Y_3 and Y_4 on research question 2.

To summarize, Y_1 and Y_2 operationalize the reward side of the concept of burden, where Y_1 has a bearing on the self-interest of the business while Y_2 is more altruistic. The response variable Y_3 operationalizes perceived burden through a combination of burden associated with mainly steps 4–7 in the response process (Section 1.2). Unlike Y_1, Y_2 and Y_3, the variable Y_4 is an indicator of actual behaviour. As Haraldsen (2004) points out, nonresponse is in itself not the same as perceived burden. Nonresponse is rather a potential effect of perceived response burden. Nevertheless, we use it as one of four response variables as it is common in studies on efficacy of incentives. In a complex model following Haraldsen (2004), treatments may affect Y_1 and Y_2 which may have an effect on Y_4 through Y_3. As will be seen below, we have chosen a more straightforward model.

The response variables Y_1, Y_2, and Y_3 generate ordered categorical data. The variable Y_4 is binary. Through four cumulative logit models, one for each of the four response variables, we investigated if the treatments have any effect on the response variables. Ease of interpretation makes the proportional odds model (e.g., Agresti 1990) attractive. Aiming at a parsimonious model, treatment, block and their interaction were the only factors included:

$$\log\left(\frac{\Pr(Y \le d|i,j)}{1-\Pr(Y \le d|i,j)}\right) = \alpha_d + \beta_{1i} + \beta_{2j} + \beta_{3ij}, \quad 1 \le d \le \Delta, \quad 1 \le i \le 4, \quad A \le j \le D \tag{1}$$

where Y is the response variable which can attain categories 1, 2, . . . , Δ, $\Delta + 1$ and $\Pr(Y \le d|i,j)$ is the probability of a respondent marking category d or lower for Y given that this respondent has received treatment i and belongs to block j. The d's are usually referred to as cut-off points. Thus the model in (1) expresses the ratio $\Pr(Y \le d|i,j) \times [1-\Pr(Y \le d|i,j)]^{-1}$ as being proportional to $\exp(\alpha_d)$ times a proportionality factor consisting of the number e raised to the sum of parameters β_{1i} for treatment $i = 1, \ldots, 4$, β_{2j} for block $j =$ A, B, C, D and β_{3ij} for the interaction. The parameterization $\beta_{11} = 0$, $\beta_{2A} = 0$, $\beta_{31j} = 0$ and $\beta_{3iA} = 0$ was chosen. There are other attractive features of the model including the fact that it is essentially invariant under change of the labels of the levels and of the order of direction of the levels. For the variable Y_4, for which $\Delta = 1$, the model in (1) simplifies to the ordinary logistic model.

By rewriting the model we obtain (for block j) the following cumulative odds ratio for cut-off d, $1 \le d \le \Delta$, and treatment i', relative to treatment i'':

$$\frac{\Pr(Y \le d|i=i',j)}{1-\Pr(Y \le d|i=i',j)}\left(\frac{\Pr(Y \le d|i=i'',j)}{1-\Pr(Y \le d|i=i'',j)}\right)^{-1} = \exp(\beta_{1i'} - \beta_{1i''} + \beta_{3i'j} - \beta_{3i''j}) \tag{2}$$

Note that the right-hand side of the expression does not depend on the cut-off d which is one of the features of the proportional odds model.

Also, note that if the interaction is omitted, the expression in (2) simplifies. In this application, the β_{1i}'s (treatments) are of prime interest. The cumulative odds ratio for treatment i', relative to treatment i'', is in a model without interaction terms

$$\frac{\Pr(Y \le d|i = i')}{1 - \Pr(Y \le d|i = i')} \left(\frac{\Pr(Y \le d|i = i'')}{1 - \Pr(Y \le d|i = i'')} \right)^{-1} = \exp(\beta_{1i'} - \beta_{1i''}), \quad 1 \le d \le \Delta \tag{3}$$

If $\exp(\beta_{1i'} - \beta_{1i''}) > 1$ there is an advantage in terms of Y for treatment i' in comparison with treatment i''. For example, if treatment 2, "How the data are used," is beneficial to the attitude towards response burden Y_3 (in comparison with the treatment 1, "No special enclosure"), we would expect $\exp(\beta_{12})$ to be larger than unity (recall that $\beta_{11} = 0$). If the treatment has no effect at all, the odds ratio equals 1.

We view the parameters in the proportional odds models as model parameters and not as quantities in a finite population. Hence the survey weights were not taken into account when estimating model parameters. See further discussion of this issue in Chambers (2003). The models were fitted with PROC LOGISTIC in the SAS System® Version 9.1 for Windows. The standard 5% significance level was used in all hypothesis testing.

5. Results

5.1. Response Rates

The number of responses as of February 2005 is given in Table 3. For example, 920 businesses were given treatment 1, of which 620 sent back the SBS questionnaire (65 of those printed the questionnaire on paper and returned it by mail). 477 of the 620 businesses responded to the nonmandatory PRB questions. None responded to the PRB questions without having returned the SBS questionnaire. The overall response rate to the CSAQ was rather disappointing. A telephone follow-up after February pushed the response rate up to 84%. A subsample of the businesses were reminded about the survey and its mandatory status. The PRB questions were left out at this stage.

5.2. Premeditated Hypotheses

Table 4 summarizes the outcome for response variable Y_1 (see Table 2 for a description). There were only six observations in the "very useful" category; hence the categories

Table 3. Response rates as of February 2005 by treatment and respondents' choice of mode

	Treatment				Total
	1	2	3	4	
Sample size	920	920	920	917	3,677
Number of responses to the SBS survey	620 (67%)	630 (68%)	623 (68%)	624 (68%)	2,497 (68%)
Electronic (CSAQ)	555	565	574	559	2,253
Returned by mail	65	65	49	65	244
Number of responses to the PRB questions	477 (52%)	514 (56%)	497 (54%)	492 (54%)	1,980 (54%)

Table 4. Relative distribution of Y_1 (Q: "Do you think that the statistics based on the SBS survey are of great or little use to your business?") by treatment

Y_1	Treatment			
	1	2	3	4
Very useful	0%	0	0	1
Fairly useful	5	7	8	9
Neither useful nor useless	21	21	26	22
Fairly useless	20	20	17	20
Very useless	54	52	49	48
TOTAL (number of responses excl "don't know")	100% (397)	100% (417)	100% (398)	100% (407)
Don't know (number of responses)	67	82	87	74
Item nonresponse (number)	13	15	12	11

"very useful" and "fairly useful" were joined when the proportional odds model was applied. Only the main factors were included as the interaction factor (adjusted for the main factors) was found not significant (see Table 5). This was established with a standard likelihood-ratio test. Block adjusted for treatment is also not significant but the *p*-value for treatment adjusted for block is 0.11. For this variable, we were only interested in comparing the effect of treatments 3 and 4 (thus "Key ratios" was enclosed) with the effect of Treatments 1 and 2 since the other enclosure, "How the data are used," was prior to the experiment not assumed to have an appreciable effect on this variable. Table 4 indicates that businesses given Treatments 3 or 4 tend to mark lower categories on the question (i.e., believe more strongly in benefits of using the statistics for the purposes of management of their business) than do businesses given Treatments 1 or 2. As can be seen in Table 6, the estimated odds ratio of the effect of the enclosure "Key ratios" on Y_1 was 1.24. This is significant on the 5% level. Hence the estimated odds (controlled for block) for respondents who have been given this incentive to believe more strongly that the statistics the survey eventually will produce are of use to their own business are 1.24 higher than the odds for respondents not given this incentive.

Table 6 lists estimated odds ratios for the final models for treatment effects on the remaining response variables Y_2, Y_3 and Y_4. These will be discussed below. All tests behind Table 6 were specified ahead of the experiment (Wahlström 2004).

Table 5. Test statistics and p-values of factors in the model using a likelihood ratio test

Response variable	Treatment (adjusted for block)		Block (adjusted for treatment)		Interaction (adjusted for main factors)	
	Test statistic	*p*-value	Test statistic	*p*-value	Test statistic	*p*-value
Y_1	6.05	0.11	0.48	0.92	9.81	0.37
Y_2	4.39	0.22	5.35	0.15	9.20	0.42
Y_3	3.82	0.28	9.91	0.02	15.80	0.07
Y_4	6.03	0.11	5.65	0.13	11.44	0.25

Note: The test statistics for the main effects and the interactions are compared to a chi-square distribution with 3 and 9 degrees of freedom, respectively.

Table 6. Selected odds ratios and corresponding estimates and 95% confidence intervals

Response variable	Treatment	Odds ratio	Estimate	95% confidence interval
Y_1	3 and 4	$\exp(0.5(\beta_{13}+\beta_{14})-0.5(\beta_{11}+\beta_{12}))$	1.24	1.03–1.49
Y_2	2 and 4	$\exp(0.5(\beta_{12}+\beta_{14})-0.5(\beta_{11}+\beta_{13}))$	1.17	0.97–1.41
Y_3	2	$\exp(\beta_{12}-\beta_{11})$	1.19	0.93–1.53
	3	$\exp(\beta_{13}-\beta_{11})$	0.96	0.75–1.23
	4	$\exp(\beta_{14}-\beta_{11})$	0.97	0.76–1.25
Y_4	2	$\exp(\beta_{12}-\beta_{11})$	1.03	0.85–1.24
	3	$\exp(\beta_{13}-\beta_{11})$	0.89	0.73–1.08
	4	$\exp(\beta_{14}-\beta_{11})$	1.12	0.93–1.36

Table 7 reports on the outcome for response variable Y_2. The interaction was again not significant and was therefore dropped when estimating the treatment parameter. Furthermore, neither treatment nor block is significant (Table 5). For this variable, Y_2, the effect of Treatments 2 and 4 as one group was compared with the effect of Treatments 1 and 3 as another group with similar arguments as for Y_1. As can be seen in Table 6, the 95% confidence interval for the odds ratio covers unity. While the point estimate 1.17 is similar to the estimated odds ratio pertinent to Y_1, it is a consequence of our formal strategy to declare it not significantly different from 1.

The results for response variable Y_3 by treatment are displayed in Table 8. Strictly speaking, the additional effect of the interaction (adjusted for treatment and block) was not significant (see Table 5). However, due to the low p-value ($p = 0.07$) the interaction effect was tentatively included in the model. Since further analyses did not result in any clear interpretation, it was omitted from the model. The odds ratios for Y_3 computed without the interaction are reported in Table 6.

Unlike the treatment effect on Y_3, block (adjusted for treatment) was significant, as can be seen in Table 5. Hence respondents at larger businesses perceive the survey as less

Table 7. Relative distribution of Y_2 (Q: "Do you think that the statistics based on the SBS survey are of great or little use to society?") by treatment

Y_2	Treatment			
	1	2	3	4
Very useful	2%	3	1	3
Fairly useful	28	32	34	31
Neither useful nor useless	29	29	29	31
Fairly useless	19	18	17	19
Very useless	22	18	19	16
TOTAL (number of responses excl "don't know")	100% (340)	100% (366)	100% (345)	100% (358)
Don't know (number of responses)	125	134	138	123
Item nonresponse (number)	12	14	14	11

Table 8. Relative distribution of Y_3 "perceived burden" by treatment

Y_3	Treatment			
	1	2	3	4
Heavy	24%	29	23	25
Neutral	59	55	61	57
Light	17	16	17	18
TOTAL(number of responses excl "don't know")	100% (458)	100% (497)	100% (486)	100% (483)
Item nonresponse (number)	19	17	11	9

burdensome than do respondents at smaller businesses. This is borne out by the estimated odds ratios $\exp(\beta_{2j})$ for Block $j = $ B, C or D relative to Block A that can be obtained in a fashion similar to what was done in (3). The 95% confidence intervals for these odds ratios for Block $j = $ B, C, D were 1.04–1.75, 1.12–1.90 and 1.11–1.85, respectively. For example, the estimated odds for businesses in the 100+ employment bracket (Block D) to perceive a burden less than or equal to level d, irrespective whether d is 1 (light), 2 (neutral) or 3 (heavy), is 1.43 higher than that of businesses in the 20–29 employment bracket (Block A).

Finally, Table 9 exhibits the outcome for Y_4 (responded by November 7). The table splits nonrespondents by November 7 into those who responded to the SBS survey between November 7 and February and those who had not responded by February. This split is not utilized in the model. With the interaction factor omitted ($p = 0.25$), the corresponding p-values for treatment and block are $p = 0.11$ and $p = 0.13$, respectively (Table 5). The estimates in Table 6 indicate that Treatment 3 has a surprising negative effect on the response rate (in comparison with Treatment 1) while Treatment 4 has a positive effect (again in comparison with Treatment 1). However, all confidence intervals cover unity.

5.3. Technical Discussion of Model and Analyses

The aim of the analyses in Section 5.2 was to quantify the relationship between treatment and the four responses controlled for the block factor, rather than finding the model that would offer the best prediction of the four responses given a number of potential

Table 9. Relative distribution of Y_4 (responded on time) by treatment

	Treatment			
	1	2	3	4
Responded punctually	36%	37	34	39
Responded late	31	31	34	29
Nonrespondents (as of February)	33	32	32	32
Total sample size	100% (920)	100% (920)	100% (920)	100% (917)

covariates. This is the reason why we have not drawn the most from available data by considering a large number of models. Moreover, one should of course only use such factors or covariates that are not suspected to change after treatment is given. One variable that we considered was "Have you responded to the previous questionnaire?" with categories "Yes, more than once," "Yes, once," "No" and "Don't know." However, we found from our data that this variable is associated with size of business, i.e., the block factor, so there was no point in adding it to the model. For other available variables, such as "Time to complete the questionnaire," one can imagine that the answer could depend on what treatment the respondent was randomized to. For Y_3, there are exogenous variables that could conceivably be added to the model, such as certain key ratios, but none of these altered the main conclusion about Y_3: no treatment effect.

It should be noted that eight confidence intervals have been constructed (reported in Table 6) and a number of hypotheses have been tested. The issue of performing multiple significance tests should be taken into account when drawing conclusions from one significant result among some number of nonsignificant results. However, to choose a stricter but still conventional level of a test, a 99% confidence interval of the effect of the enclosure "Key ratios" on Y_1 just covers unity (the lower and upper limits being 0.97 and 1.58).

Turning to the risk of bias incurred by differential nonresponse over treatment groups, there does not appear to be any connection between nonresponse to PRB questions and treatment, conditional on block. We have seen from tests concerning Y_4 (Tables 5 and 6) that there is no significant treatment effect on the probability of responding to the SBS survey. The same conclusion can be reached by considering response rates for PRB questions.

The assumption that the odds ratios are independent of the cut-off of the response variable is inherent in the proportional odds model (see (2)). The SAS procedure PROC LOGISTIC outputs results from a score test of this proportionality assumption. In the score test, the generalized logit model (which in this case is the saturated model) is compared to the proportional odds model. For the models underlying Table 6 the p-values were 0.20, 0.74, and 0.66 for Y_1, Y_2, and Y_3, respectively. Also, differences between observed frequencies and expected frequencies under either the proportional odds models or the generalized logit models did not show any noticable pattern. Thus, the fit of the proportional odds models seemed adequate.

5.4. Exploratory Analysis

The statistical analyses of the experiment reported so far consist of tests of a limited set of premeditated hypotheses about some contrasts. In addition to this we have done more exploratory analyses with the aim of identifying subgroups with deviant distribution of certain variables, especially the four prespecified response variables. One intriguing finding is that there seem to be associations between time spent on retrieving the relevant information for the survey, appreciation of the societal use of the results of the survey and, to a lesser extent, the treatment of having been sent the enclosure "How the data are used." As can be seen in the top left cell of Table 10, 29–35% of respondents with relatively immediate access to the required data (<60 minutes for retrieval, i.e., less than the

Table 10. Distribution of responses to Y_2 (Q: "Do you think that the statistics based on the SBS survey are of great or little use to society?") within levels of time required for retrieving data necessary to complete the survey by Treatments 2 and 4 and Treatments 1 and 3

Time spent on data retrieval	Fairly or very useful	Neither useful nor useless	Fairly or very useless	Don't know	Item nonresponse	Row total (Number)
0–60 min	35%	23	19	22	1	100% (269)
	29	20	23	28	0	100% (246)
60+ min	24	22	28	25	1	100% (561)
	24	21	28	26	2	100% (538)
Item nonresponse	13	19	27	32	10	100% (176)
	13	20	28	30	10	100% (190)
TOTAL						(1,980)

Note: The first entry in each cell is a percentage based on respondents who have been given the enclosure "How the data are used," the second entry is based on the complementary group of respondents.

median) believed that the structural business statistics are fairly or very useful to society. The higher percentage refers to those who had been sent "How the data are used" and the lower percentage to those who had not been sent that enclosure. There was no difference in the <60 and the 60+ minutes groups in terms of number of items filled out in the SBS part of the CSAQ.

In the group of respondents who spent less than one hour on data retrieval there is a difference between those who obtained "How the data are used" and those who did not. There is no treatment effect in the other two groups displayed in Table 10, i.e., 60+ minutes and item nonresponse. The fact that this particular table has been preceded by exploratory data analysis makes hypothesis testing too tenuous to be viable.

Respondents who have not responded to the question on how much time they spent on data retrieval are generally negative about societal use. Only 13% of them believe that the statistics are fairly or very useful. This is part of a general pattern; respondents with more than one item nonresponse among the PRB questions tend to be negative to the SBS survey.

Businesses were given specifications of costs and sources of income to report. Through satisficing (Krosnick 1991) some businesses might unduly simplify their burden by reporting on fewer items than they should. This would manifest itself in few nonzero values filled out in the SBS part of the CSAQ, i.e., in the mandatory part of the survey. Interestingly, the response to the question "Do you think that the statistics based on the SBS survey are of great or little use to society?" and time spent on retrieving data are not appreciably associated with the number of nonzero values filled out in the mandatory part. As just mentioned, item nonresponse in the voluntary part of the survey, i.e., the PRB question set, is positively associated with perceived burden, Y_3. However, the number of nonzero values filled out in the mandatory part is not negatively associated with heavy perceived burden. The direction of the association, although weak, is rather the opposite. Thus respondents who feel a heavy burden do not appear to satisfice by making their response overly simplistic.

There are some other variables that one could suspect to be associated with the response variables. A rather surprising finding is that Y_2, perceived usefulness to society, and time

spent on data retrieval are only weakly associated with size of business. Recall that perceived burden appears to be negatively associated with business size (Table 5). An interpretation of this is that large businesses feel the burden less than do smaller businesses although their actual burden is as least as heavy. Profit is another obvious potential explanatory variable. Tomaskovic-Devey et al. (1994) found by examining a particular survey that profit relates negatively to likelihood to respond. We have not found any association between profit and Y_2 but a positive association between profit and likelihood to respond.

We know from company visits that many respondents have bespoke systems and other procedures that allow them quick access to data needed for repeated surveys; on the other hand many other businesses have no such systems or procedures (Hedlin et al. 2005). Those who need little time to retrieve data have procedures and systems in place probably because they have anticipated the survey. Their record formation has been effective (cf. the response process outlined in Section 1.2). They have high professional standards and the capacity to organize procedures that allow them comfortable data access. We hypothesize that these respondents take it as part of their professional responsibility to commit themselves to tasks that relate to society outside their own workplace. This is a business survey parallel to the "sense of civic duty" that Couper et al. (1998) focus on. However, along a different line of thought, since retrieving data is relatively convenient for these respondents, they may perceive the survey as a straightforward exercise that should lead to reliable statistics and hence be useful.

6. Discussion

A cost-effective way to test and measure amendments to a repetitive survey process is to conduct experiments embedded in the regular survey. As a basic experimental design one part of the sample is exposed to the new survey process (e.g., a redesigned questionnaire) and the other part to the old survey process. Responses to PRB questions and other data will indicate the effectiveness of the new survey process and the perceived response burden among businesses in both groups. Further use of this and alternative PRB question sets will be discussed elsewhere.

Feedback of survey results was offered as an incentive to respondents. It had no significant effect on the response rate, which is consistent with the literature that clearly shows that incentives that come with obligations have little or no effect. Our finding is nevertheless novel as feedback as an incentive is different from the types of incentive studied in randomized experiments in other surveys, mainly surveys of individuals and households.

Another conclusion of the experiment may appear to contradict the first result. We have seen that respondents at businesses that are given the enclosure informing about feedback of survey results tend to believe that the statistics from the survey will be more useful to their business than do those at businesses that are not given that enclosure. For a start, this confirms the view generally held among Statistics Sweden's business survey practitioners that survey feedback is appreciated by businesses. However, this result goes further than a logical tautology of appreciation of survey feedback: it is interesting to see that a

document enclosed with the covering letter can make a difference when the respondent fills in the questionnaire a few weeks later. There is no mention of survey feedback in the PRB questionnaire; the link is established by the respondents themselves, perhaps semiconsciously. Hence this experiment constitutes an indirect observation of the respondent's cognitive process. We believe that this indicates that the "Key ratio" enclosure does change perceptions of the survey among some of the respondents. This is not to say that more deep-seated attitudes or values are necessarily influenced; indeed, respondents who on the fourth sheet of the CSAQ have commented belligerently on the survey and the resulting statistics are equally spread over treatment groups.

An early analysis, which was reported in Hedlin et al. (2005), was confined to 1,520 responses to the PRB questions up to November 21, 2004. The differences between those results and the ones reported here, which includes also late responses up to the end of February, are surprisingly small. For example, the odds ratios corresponding to those of Y_1 and Y_2 in Table 6 were 1.30 and 1.11, with confidence intervals 1.05–1.60 and 0.89–1.37, respectively. Thus the period of time elapsed between obtaining the covering letter and responding does not seem important.

We asked the respondents whether they felt it was quick or time-consuming to retrieve the necessary information to fill in the questionnaire and whether they thought it was easy or burdensome to complete the questionnaire. The fact that there was little treatment effect on responses to these questions indicates that perceived response burden as such was not affected by the treatments. We can only speculate on potential reasons for this. First, contrary to what we like to believe there may not be a strong inherent link between perceived rewards and perceived response burden. In fact, social exchange theory does not imply such a link. What the theory says is rather that the respondent's impression of burden and reward interplay to produce some result.

Second, we do not know how many of the actual respondents had the chance to see the enclosures. These may have been retained by "gatekeepers," i.e., staff who obtain the covering letter and forward the survey task within the organization. About 70% of the respondents have indicated that only one person was involved in the response, although gatekeepers may reasonably have been excluded. From a pragmatic point of view, since gatekeepers do take part in the response process (Step 2 in the model response process in Section 1.2) an experiment of this kind should encompass their actions.

Third, the annual SBS survey asks for ten or eleven months old data with the main purpose of delivering input to the National Accounts. Many business respondents would find it easier to see the societal or corporate use of some subannual surveys, in particular price indices. Although the enclosure "How the data are used" pointed out various types of applications of the structural business statistics, short-term surveys may allow a different set of motivational material. Naturally, for both annual and subannual surveys, there is scope for improvement in the contents and design of enclosures. Also, a truly web-based survey would offer more potential for putting the information across to the actual respondents. For example, text boxes could be created to be accessible from a web-questionnaire (Clayton and Werking 1998, in particular p. 551).

Fourth, this article reports on effects of a single exposure to the treatments. Multiple exposures, perhaps of different kinds, and measurement of burden over time is an area for future research.

Fifth, rightly or wrongly, there are reasons for not trusting the resulting statistics that we have not addressed; this is the respondents' own data inaccuracy, which many respondents believe is transferred to or even exacerbated in the aggregated statistics. Statistics Sweden gives on its web site (Ribe 1996) nontechnical information about how aggregation of individual numbers will tend to cancel out random measurement errors, but not many business survey respondents will have seen it.

In our experiment we found that it is the reward side of the two-sided concept of response burden that was influenced by our treatments. In appealing to either the self-interest of businesses or the sense of taking part in a process important to the community, self-interest had a larger effect. While this may not be very flattering to businesses it must be borne in mind that in most successful co-operations there is a fair amount of mutual self-interest, often referred to as a "win–win" situation.

In sum, NSIs should pay particular attention to how their surveys are introduced to the respondents and what information about the use and usefulness of the survey are offered to the participating businesses. Statisticians may need to convert their views about the role of the respondent. As Colm O'Muircheartaigh has put it in the context of response error models: ". . . the respondent is largely disregarded, seen as an obstacle to be overcome rather than an active participant in the process" (O'Muircheartaigh 1999, p. 43). The current article shows that it is possible to work together with respondents to achieve goals, although the respondent and the NSI may see different uses of the same information that they collate in concerted action.

7. References

Agresti, A. (1990). Categorical Data Analysis. New York: Wiley and Sons.

Chambers, R.L. (2003). Introduction to Part A. In Analysis of Survey Data, R.L. Chambers and C.J. Skinner (eds). New York: John Wiley and Sons, 15–28.

Clayton, R.L. and Werking, G.S. (1998). Business Surveys of the Future: The World Wide Web as a Data Collection Methodology. In Computer Assisted Survey Information Collection, M.P. Couper, R.P. Baker, J. Bethlehem, C.Z.F. Clark, J. Martin, W.L. Nicholls, II, and J.M. O'Reilly (eds). New York: John Wiley and Sons, 543–562.

Couper, M.P., Singer, E., and Kulka, R.A. (1998). Participation in the Decennial Census: Politics. Privacy, Pressures. American Politics Quarterly, 26, 59–80.

Dillman, D.A. (2007). Mail and Internet Surveys: The Tailored Design Method (2nd Ed). New York: John Wiley and Sons.

Dillman, D.A., Singer, E., Clark, J.R., and Treat, J.B. (1996). Effects of Benefits Appeals, Mandatory Appeals and Variations in Statements of Confidentiality on Completion Rates for Census Questionnaires. Public Opinion Quarterly, 60, 376–389.

Gower, A.R. (1994). Questionnaire Design for Business Surveys. Survey Methodology, 20, 125–136.

Hak, T., Willimack, D.K., and Anderson, A.E. (2003). Response Process and Burden in Establishment Surveys [CD-ROM]. Proceedings of the American Statistical Association, Section on Government Statistics, 1724–1730.

Haraldsen, G. (2004). Identifying and Reducing Response Burden in Internet Business Surveys. Journal of Official Statistics, 20, 393–410.

Hedlin, D., Dale, T., Haraldsen, G., and Jones, J. (eds) (2005). Developing Methods for Assessing Perceived Response Burden. A Joint Report of Statistics Sweden, Statistics Norway and the UK Office for National Statistics. Available at epp.eurostat.ec.europa.eu/pls/portal/url/ITEM/1CA1E9AC26242D43E0440003BA9322F9

James, J.M. and Bolstein, R. (1992). Large Monetary Incentives and Their Effect on Mail Survey Response Rates. Public Opinion Quarterly, 56, 442–453.

Jones, J., Rushbrooke, J., Haraldsen, G., Dale, T., and och Hedlin, D. (2005). Conceptualising Total Business Survey Burden. Survey Methodology Bulletin (SMB 55 3/05). Office for National Statistics, UK, 1–10.

Krosnick, J.A. (1991). Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys. Applied Cognitive Psychology, 5, 213–236.

O'Muircheartaigh, C. (1999). CASM: Successes, Failures, and Potential. In Cognition and Survey Research, M.G. Sirken, D.J. Herrmann, S. Schechter, N. Schwarz, J.M. Tanur, and R. Tourangeau (eds). New York: John Wiley and Sons, 39–62.

Ribe, M. (1996). Fel hit och fel dit. . . Statistikskolan. Välfärdsbulletinen, nr 3. [In Swedish] Retrieved from the Web September 27, 2007 http://www.scb.se/Grupp/Klassrummet/_Dokument/Felhitochfeldit.pdf

Rosén, B. (1997). On Sampling with Probability Proportional to Size. Journal of Statistical Planning and Inference, 62, 159–191.

Shettle, C. and Mooney, G. (1999). Monetary Incentives in U.S. Government Surveys. Journal of Official Statistics, 15, 231–250.

Singer, E., Van Hoewyk, J., Gebler, N., Raghunathan, T., and McGonagle, K. (1999). The Effect of Incentives on Response Rates in Interview-Mediated Surveys. Journal of Official Statistics, 15, 217–230.

Tomaskovic-Devey, D., Leiter, J., and Thompson, S. (1994). Organizational Survey Nonresponse. Administrative Science Quarterly, 39, 439–457.

Wahlström, H. (2004). Projekt Perceived Response Burden. Analysplan. Internal Report. Statistics Sweden. [In Swedish]

Willimack, D.K. and Nichols, E. (2001). Building an Alternative Response Process Model for Business Surveys. Paper presented at AAPOR, Montreal, Canada. Retrieved on December 29, 2006 from http://www.amstat.org/sections/srms/proceedings/ at '2001'.

Willimack, D.K., Nichols, E., and Sudman, S. (2002). Understanding Unit and Item Nonresponse in Business Surveys. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley and Sons, 213–277.

Willimack, D.K., Lyberg, L., Martin, J., Japec, L., and Whitridge, P. (2004). Evaluation and Adaptation of Questionnaire Development, Evaluation, and Testing Methods for Establishment Surveys. In Methods for Testing and Evaluating Survey Questionnaires, S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds). New York: John Wiley and Sons, 385–407.

Received November 2005
Revised November 2007