UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
UNECE Expert Meeting on Statistical Data Collection – Towards a New Normal?
26 to 28 October 2022, Rome, Italy

Session 2
10 October 2022

Reduction of response burden for individual businesses

Lianne Ippel, Jonas Klingwort, Marc Smeets, Remco Paulissen, Geert Gommans, Lars Burgerjon, Goele Goris, Ger Snijkers, and Leanne Houben (Statistics Netherlands)

gje.ippel@cbs.nl

Abstract

Statistics Netherlands aims to keep both actual and perceived response burden on businesses as low as possible. In order to achieve this, we have set seven ambitions, which are laid out in the Position Paper 'Focus on the Reporting Business' (multi-year plan, reduction response burden 2021-2023), They include: (1) the building of a portal, (2) importing RCSFI*-data to the Structural Business Survey, (3) the reduction of the sample size by enhanced enforcement, (4) the implementation of the changes in European regulations with minimum increase in burden, (5) strengthening cooperation with sectors and businesses, (6) the development of a new strategy on business data collection, and (7) to avoid, where possible, multi-surveying of businesses. In the presentation, we will discuss these ambitions with a focus on the 7th ambition.

Statistics Netherlands operates a sample coordination system for business surveys. The system's purpose is to reduce the (perceptual) response burden of individual businesses by evenly spreading the surveys among businesses. Generally, the number of surveys for the smaller companies is rather small. However, despite the sample coordination, several businesses still receive a more than proportionate number of questionnaires. Especially the smaller companies suffer from this as most of them do their own administration. In some cases, these are very specific companies providing a unique contribution to the survey, in other cases however, businesses are multi-surveyed by chance. Either way, this can cause peak pressure in response burden: Hotspots. For this reason, businesses, trade associations, and politicians have requested Statistics Netherlands to come up with a solution.

In May 2021, we started with a pilot for small businesses (with up to 20 employees): hotspots were identified and removed from the samples as much as possible. In June 2022, we started evaluating this pilot approach with regard to number of hotspots removed, effects on response behavior, and effects on final statistics. First results show that hotspots have similar or slightly improved response rates compared to non-hotspots. Using the R&D survey, we analyzed the effect of omitting hotspots in the final estimates. First results show that removing hotspots from the sample, has a small negative effect on the precision of the total estimates and led to larger standard errors. Apart from analyzing the effectiveness of this approach, this evaluation analysis is aimed at establishing a longer term, systematic solution to lower (perceived) response burden while ensuring methodological soundness. During the presentation, results will be discussed more in detail.

Reduction of response burden for individual businesses

Lianne Ippel, Jonas Klingwort, Marc Smeets, Remco Paulissen, Geert Gommans, Lars Burgerjon, Goele Goris, Ger Snijkers, and Leanne Houben

summary An important ambition of Statistics Netherlands (SN), is to reduce the actual and perceived response burden for the Dutch businesses. In the Netherlands, most of the business surveys are mandatory. Yet, small businesses with up to 10 or 20 employees, receive none, or just a few SN-surveys. Nevertheless, some of these businesses receive three or more surveys. For these businesses, this can result in a problematic peak in response burden.
In this paper, we will discuss our approach to avoid multi-surveying of small businesses as much as possible within the boundaries of what is feasible. Sometimes business are so unique, that we need their data to make adequate statistics. We will discuss the applied strategies, the initial results, and next steps.

topic Respondent care: managing and reducing statistical burden

1. Introduction

Statistics Netherlands (SN) uses a sample coordination system for business surveys (Smeets & Boonstra 2018). The system's purpose is to reduce the response burden of individual businesses by evenly spreading the surveys among businesses. Currently, this system is still in limited use. Generally, the number of surveys for small businesses is rather low. However, despite the sample coordination, several businesses still receive a large number of surveys, resulting in a considerable increase in response burden, leading to complaints from businesses and sector organizations (van Straalen, 2020). In addition, multi surveying may have a negative impact on data quality (Lorenc, Kloek, Abrahamsson, & Eckman, 2013). This multi-surveying of businesses might be due to unique characteristics of a business, we therefore need the data of such a business to produce adequate statistics. However, due to random sampling, it might also just be coincidence. In the current project, we focus on the small businesses which receive a large number surveys. We refer to these businesses as hotspots. In the current project, hotspots are defined either as:

- businesses with less than 10 employees that receive more than three surveys in the last 12 months, or
- businesses with 10 to 20 employees that receive more than four surveys in the last 12 months.

SN aims to keep both actual and perceived response burden on businesses as low as possible. To achieve this, SN has set a multi-focus approach with seven ambitions (Houben, 2021). Minimize multi surveying of small businesses where possible is an important part of this approach.

In May 2021, we started with a pilot: hotspots were identified and manually removed from the samples as much as possible: Hotspots were identified and together with the statistics departments it was decided whether these businesses could be removed from the sample. In June 2022, we started evaluating this pilot approach with regard to the number of hotspots removed, the effects on response behavior, and the effects on final statistics. First results show that the number of hotspots is indeed reduced. However, analysis of the lasting effect of the used strategies and the effects on the outcome is still taking place. Using the Research and & Development (R&D) survey, we are analyzing the effect of omitting hotspots on the estimates of population and domain totals. Apart from analyzing the effectiveness of the applied strategies, this evaluation analysis is aims to establish a longer-term, systematic solution to lower (perceived) response burden while maintaining methodological soundness.

In the following chapter, we discuss the SN multi-focus approach to reduce response burden in more detail. Next, the hotspot pilot and its tentative results are discussed. The last chapter highlights the lessons learned so far and future steps.

2. A multi focus approach to reduce response burden

The mission of SN is to publish reliable and coherent statistical information that responds to society's needs. If this statistical information cannot be compiled from administrative data sources such as government registers, the data are collected from businesses, citizens, and / or households, whereby businesses can be obliged by law to supply data. This obligation for businesses to respond, can result in a large response burden for businesses. Therefore, SN aims to perform its tasks as well as possible while keeping the response burden as low as possible.

The information we now collect from businesses is approaching the minimum level required to produce good statistics, although we continue to search for alternatives and improvements together with sector organizations and businesses. SN aims to make it as easy as possible for businesses to submit data, now and in the future. As mentioned in the introduction SN has set seven ambitions to reduce the response burden, which are laid out in the Position Paper 'Focus on the Reporting Business' (multi-year plan, reduction response burden 2021-2023, Houben, 2021).

In summary, these ambitions are: (1) to avoid, where possible, multi-surveying of businesses by adjusting sample designs and upgrading the SN sample coordination system, (2) to facilitate the response process by import data directly from their business records into the Structural Business Survey with the push of a button, (3) facilitate businesses in reporting data and accessing information by developing a business portal, (4) implementing a more effective data collection process by improving the communication strategy and intensifying enforcement, resulting in higher response rates. Because of the higher response rates, sample sizes can be reduced, resulting in the reduction

of the total survey burden, (5) implementing the changes in European regulations with minimum increase in burden by launching a program to explore possible innovations for the various statistics (6) strengthening cooperation with sectors and businesses, by intensifying consultations with business representatives (sector organizations), improving complaint handling, measuring the actual burden and gaining more insight in the perceived burden, (7) the development of a new strategy on business data collection, by putting businesses first, being transparent and predictable in what businesses can expect from SN, and the reuse of automatically retrieved data from business systems, with minimal interference from the businesses themselves.

SN monitors progress on these seven ambitions on a quarterly basis, providing feedback to the sector organizations and consulting these stakeholders about the follow-up.

There is a clear link between these ambitions. They reinforce each other. For example, reducing the sample size as a result of an improved data collection process ensures that fewer businesses receive multiple questionnaires, resulting in a somewhat reduced number of hotspots. Importing data with the push of a button will make responding easier, thus reducing the response burden, for the businesses that have to fill in the Structural Business Survey. By building a portal, we can inform businesses at an earlier stage when they have to comply to survey obligations, making it easier for them to plan. Implementing a new strategy on our business data collection process that is in line with the working methods of companies as much as possible will lead to lower perceived and actual response burden.

3. Hotspots

In an ongoing project, we aim to reduce the multi surveying of small businesses, mentioned as our first ambition to reduce response burden. Over the course of a year, different strategies to reduce the number of hotspots have been applied. In the following chapter, we discuss the applied strategies in more detail.

3.1 Applied strategies

Basically, there are two strategies to reduce the number of hotspots: reduction of the total response burden or a more evenly spread of the total burden over the businesses. During the pilot, multiple approaches have been applied to circumvent hotspots receiving (even) more questionnaires, given the time and options available to the statistical teams and the team responsible for drawing the samples.

This has resulted, among others, in the following strategies: 1) For several surveys, hotspots could be removed from the sample if these businesses did not have a crucial contribution to the statistic according to the domain experts. 2) The second strategy is to replace hotspots with other businesses by identifying, prior to sampling, the expected number of hotspots. Using a ‘trial sample’, one could estimate the expected number of hotspots for each stratum and, when possible, oversample from these strata where a substantial number of hotspots is expected. After the sample is drawn, the hotspots could be removed (manually) from the sample, such that the effective sample size remains the same to the situation where no hotspot strategy would have been applied. 3) A third strategy is excluding hotspots from the sampling frame of the survey. However, this is only applicable for

businesses that do not strongly affect the statistic or for businesses that can be estimated by other sources or prior knowledge. This strategy is applied to the R&D survey in combination with the second strategy. This strategy was by far the most labor and time intensive, because prior to drawing the sample, all crucial businesses and the (potential) hotspots had to be identified.

3.2 Results

First, in Table 1, we present an overview of the Dutch business population categorized by the number of employees. Both the number of small businesses and the number of large businesses have grown from 2021 to 2022: the number of small businesses has increased with about 120,000 to a little over 2 million, and the number of large businesses has increased with about 1,000 businesses to 36,000. Next, in Table 2, an overview is given about how many businesses were part of one (or more) of the SN business surveys. It is noteworthy that the number of surveys was lower in the pilot year, compared to the year prior to the pilot (42 to 35). This is due to the fact that some surveys are conducted only bi-annually or even only once every 5 years. Yet, the sample sizes of these surveys of the pilot year were larger, resulting in more questionnaires being sent out (612,960 vs. 675,339). The number of hotspots is decreased to about half its initial size (768 to 376).

Table 1: Business population

	Employees		May 2021	May 2022
Small businesses				
Size 0	0		405,897	444,299
Size 1	1		1,143,251	1,221,033
Size 2	2-4		279,628	284,718
Size 3	5-9		66,126	67,499
Size 4	10-19		33,617	34,890
		Subtotal	1,928,519	2,052,439
Large businesses				
Size 5	20-49		20,155	21,040
Size 6	50-99		7,078	7,161
Size 7	100-199		3,768	3,862
Size 8	200-499		2,521	2,540
Size 9	≥500		1,602	1,645
		Subtotal	35,124	36,248
Total			1,963,643	2,088,687

Table 2: Sampled businesses

	06-2020 until 05-2021	06-2021 until 05-2022
Number of surveys	42	35
Total sampled businesses	612,960	675,339
Uniquely sampled businesses	181,986	201,331
Smaller than 10 employees	125,453	139,510
10-20 employees	24,084	26,995
>20 employees	32,449	34,826
Unique Hotspots	768	376

In Figure 1, the number of unique hotspots is presented. There is a clear downward trend visible. Two remarks can be made regarding this figure. First, there is a disturbance in the trend around the end of 2021/beginning of 2022. This is due to the fact that, among others, the structural business survey, the largest sample of the business surveys, was postponed for a couple of months. Consequently, this resulted in a number of businesses 'losing' their hotspot label as it was more than 12 months ago that they took part in this large survey, however, when the large survey did take place, the number of hotspots increased again. Secondly, there seems to be a 'quarter effect', where every three months the line slopes somewhat down further. This is due to the dates that the samples are drawn for the surveys and alongside the planning of the surveys.

In Figure 2, the number of surveys received by hotspots is presented. From May 2021 to May 2022, the overall distribution is much lower as a result of reducing the number of hotspots. At the same time, the average of surveys received is rather stable: 5.1 in May 2021 to 4.9 in May 2022. Also the median is equal to 5 in both years, although the number of businesses that received 5 surveys dropped from 360 to 206.

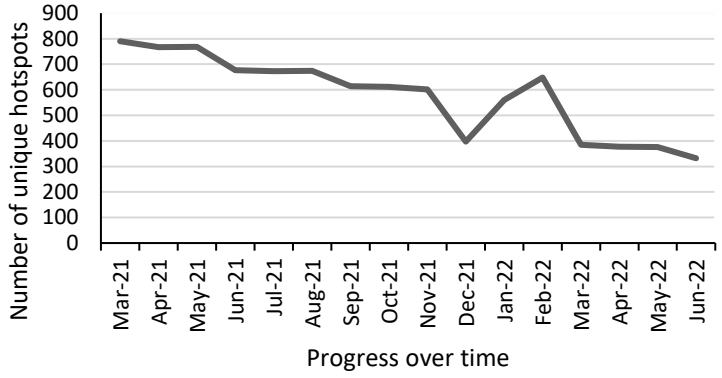

Figure 1: Number of unique hotspots over time

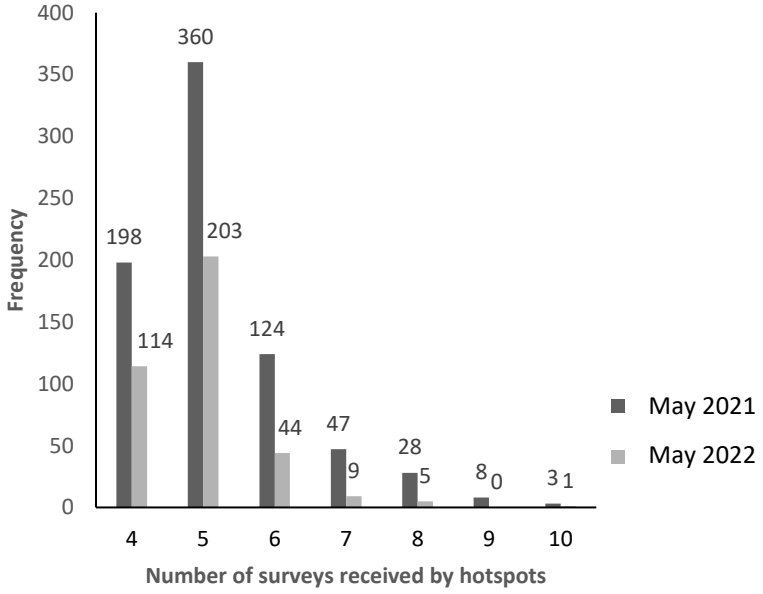

Figure 2: Number of surveys received by hotspots

Lastly, we analyzed the results of one of the studies, in addition to the pilot to learn more about the effects of removing hotspots from the sample. The R&D survey covers, among others: total expenditure, total employment, and total FTE. The R&D survey is a good use case as 1) hotspot labels were already assigned, 2) none of the hotspots were removed, and 3) the percentage of hotspots in the response data is one of the highest compared to the other surveys. In Table 3, we give an overview of the responding units of the R&D survey. The total response to the survey is 7,340 of which 235 are hotspots (3%). For this statistic, only class 4 and larger receive a questionnaire, while the contributions of the smaller businesses are estimated using register data.

The R&D response can be divided in four groups: 1) crucial businesses, 2) crucial R&D businesses, 3) no R&D businesses, and 4) the sample. The first group (1087) consists of businesses which have a complex organizational structure and a significant contribution to the Dutch economy. Consequently, for most surveys these businesses have to be observed to get an accurate and precise estimate of the statistics. In practice, these businesses will probably not be removed from the sample, even if they are hotspots. The latter also holds for the second group. The second group (1) is particularly important for R&D though not necessarily for other statistics. The third group (452) consists of businesses from which we already know they do not have R&D and therefore do not receive the questionnaire for this statistic. Their response is set to zero.

In Table 4, the results of removing the hotspots from the response of three main outcome variables are presented: total expenditure, total employment, and total FTE. The results consist of the estimate of the total, the standard error (SE), and the relative SE (rSE), where rSE is the relative standard error. The subscript h is used to indicate that hotspots are included. The estimates of the totals hardly differ whether the hotspots are included or not. Both the SE and rSE are slightly larger when the hotspots are removed.

Currently, we are investigating the effect of removing hotspots from the sample on more detailed publication cells. The publication cells are based upon the combination of the business size (i.e., the number of employees) and the Standard Business Categories (i.e., economic activities). At first inspection of the publication cells, there is a distribution of the changes in rSE, which is summarized as follows: Some rSE become smaller after removing the hotspots as the minimum of the distribution is equal to -10.5%. The first quartile is equal to 0.0%, the median is also 0.0%, the third quartile 3.0% and the maximum is 67.4%. The mean is equal to 7.7% increase of the rSE. Whether these changes are acceptable or not is ongoing research, and if need be, approaches to mitigate unacceptable changes.

Table 3: Research and Development: overview of responding business units (BU)

Type	Response	Hotspots	Percentage hotspots	Distribution of hotspot types
Crucial BU	1087	41	3% of crucial BU	17% of hotspots is a crucial BU
Crucial R&D BU	1	0	0% of crucial R&D BU	0% of hotspots is a crucial R&D BU
No R&D BU	452	22	4% of no R&D BU	9% of hotspot is a no R&D BU
Sample	5800	172	2% of sample	73% of hotspots is sample
Total	7340	235	3% of total	

Table 4: Effects of excluding hotspots on target variables

Target variable	Including hotspots			Excluding hotspots			
	$Total_h$	SE_h	rSE_h	Total	SE	rSE	$(rSE/rSE_h)-1$
Expenditure	13,944,054	208,161	0.0149	14,006,335	215,663	0.0154	0.033
Employees	191,890	4,831	0.0252	192,849	4,908	0.0255	0.012
FTE	133,527	2,758	0.0207	134,111	2,839	0.0212	0.024

4. Conclusion and discussion

In the Netherlands, for most business surveys participation is mandatory. Even though, on national level, the number of surveys businesses receive is rather low, for some businesses the number of surveys can result in a problematic peak in response burden. Especially for smaller businesses, this might result in a problematic situation. The ambition of Statistics Netherlands is to reduce response burden. While the most effective strategy is sending out less questionnaires to these businesses and/or using other data sources, the question remains how to do so without lowering the quality of the business statistics.

From May 2021 till May 2022, several strategies have been tried out and analyzing these results is still ongoing work. The first results show a positive picture, with an impressive reduction of hotspots from 768 to 376. This drop can partly be explained by the manual removal of hotspots from the samples. However, it is not a guarantee that the number of hotspots will remain the same or decrease once new or returning surveys will be conducted in the upcoming period. Additionally, the manual removal of businesses from the sample is labor intensive and may result in somewhat biased estimates of the statistics and larger standard errors since the realized sample might differ from the specified sample. This is due to the fact that the removal of businesses from the sample is a nonresponse and could negatively affect the quality of the estimates. Additionally, replacing hotspots with other businesses, can eventually result in an increase in the number of hotspots because a larger proportion of the business population is used, making it increasingly more difficult to find proper replacements which are not hotspots.

Currently, the definition for hotspots is based on the number of surveys a business participates. In addition, it may be worthwhile to account for the length and difficulty of the questionnaires. In such case, the completion time can be distributed as fairly as possible among the businesses. Lastly, by labeling businesses as hotspots by counting the number of surveys they participated in during the last 12 months can have a negative consequence: businesses can become hotspot for one month and 'lose' their label again next month resulting in an instable and unpredictable procedure where businesses are never sure when they do or do not take part in the survey. Additional studies should inform us how to deal with these situations, e.g., label businesses as hotspots for at least a number of months.

Therefore, it is especially relevant to have a multi-focus plan to lower the actual and perceived response burden. For instance, by improving the communication and enforcement strategy response rates are likely to go up, allowing a reduction of the sample size. This, in turn, will reduce the number of multi-surveyed businesses. When responding to the survey is only a push of a button, it is less of a burden to respond to surveys.

Our next steps are, among others, studying approaches to formalize the lessons learned from the applied strategies, to prevent the manual removal of hotspots from samples. At the same time, research on sample coordination is ongoing at SN to realize a wider use of the sampling coordination system and a more targeted application to hotspots. This will also enable us to take the chosen hotspot strategy into account during the estimation phase. Given the number of surveys that need to be conducted and as long as there are no alternative data sources to collect the required information, there will always be hotspots, even when we apply sample coordination as much as possible. However, by meeting our multi-focus ambitions, Statistics Netherlands aims to reduce the number of hotspots to a minimum and lower the overall response burden and spreading the response burden as fairly as possible among businesses.

References

Houben, L (2021). Position Paper 'Focus on The Reporting Business', multi-year plan reduction response burden 2021-2023 (in Dutch). Statistics Netherlands, The Hague/Heerlen.

Lorenc, B., Kloek, W., Abrahamsson, L. & Eckman, S. (2013) Description of actual response burden and response behaviour with Swedish Register. Chapter 2 in Giesen, D., Bavdaž, M. & Bolko, I. (eds) *Comparative report on integration of case study results related to reduction of response burden and motivation of businesses for accurate reporting.* BLUE-ETS deliverable 8.1. doi 10.13140/RG.2.2.24741.63207

Smeets, M. and Boonstra, H. J. (2018). Sampling coordination of business surveys at Statistics Netherlands. In Lorenc, B., Smith, P., Bavdaˇz, M., Haraldsen, G., Nedyalkova, D., Zhang, L.-C., and Zimmermann, T., editors, *The Unit Problem and Other Current Topics in Business Survey Methodology*, pages 127–137. Cambridge Scholars, Newcastle-upon-Tyne.

Straalen, van M.A. (2020). Letter from the Strategic Commission on Better Regulations for Businesses, demanding Statistic Netherlands to minimize hotspots where possible (in Dutch).