

**UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (iv): Data collection using mixed modes and multiple sources

ORGANIZATIONAL CHALLENGES USING MIXED MODE

Contributed Paper

Prepared by Jeroen van Velzen¹, Hank Hermans²

I. Summary

1. In recent years Statistics Netherlands faced serious budget cuts and most likely more will follow in the near future. By introducing the use of mixed mode in data collection, Statistics Netherlands succeeded in overcoming these challenges. In the last three years Statistics Netherlands has introduced mixed mode data collection for almost all social surveys. This year, 2012, we introduced mixed mode for the Labour Force Survey.
2. The use of web questionnaires especially decreased the need for capacity for face to face interviewing, traditionally the most expensive collection mode. The introduction of mixed mode also caused some side effects that needed some time to resolve.
3. To cope with these mixed mode design several measurements had to be taken. Measurements focussed on all aspects of data collection: process, organisation, methods and cooperation with other governmental institutes.
4. This multi-dimensional approach helped Statistics Netherlands to benefit optimality from the new mixed mode designs.

II. Introduction

5. In recent years Statistics Netherlands – like all other government institutions – was confronted with ongoing budget reductions due to the intention of central government to increase efficiency of government agencies in general and to reduce the number of government employees. In order to meet the requirements a number of programs of change were started. One of these programs of change was the centralization of data collection. The drivers for the centralization and harmonization of data collection activities have been twofold: internal and external reasons. An internal reason is a more cost-effective organization of the data collection. The external reasons include the facts that by law (2004), Statistics Netherlands was commissioned to reduce response

¹ Statistics Netherlands: j.vanvelzen@cbs.nl

² Statistics Netherlands: hjcm.hermans@cbs.nl

The views expressed in this paper are those of the authors and do not necessarily reflect the policies of Statistics Netherlands.

burden even more, and make use of available register data as much as possible, without affecting the quality of the statistics. On the basis of this law, a new strategy on data collection was put in place. This strategy dictated the ways of data collection in general: use register data as much as possible; only if additional data are needed, surveys can be conducted, using Electronic Data Interchange (like XBRL³), web surveys, paper, CATI⁴, and CAPI⁵ (in this order). The result of this strategy is multi-source/mixed-mode data collection designs.

6. In the same period Statistics Netherlands started a total redesign of the data collection for its social statistics. Potentially the redesign involved all surveys, including the surveys that were done under commission. It did, however, involve all data collection that is conducted by Statistics Netherlands itself.
7. The motivation for the redesign comes from the ever increasing demands for faster and more coherent statistics. The timeliness of Statistics Netherlands statistics is constantly under pressure, and more and more cross-relations need to be made between different variables of interest. This implies that Statistics Netherlands need to be able to react more flexible and faster to changes in user needs and to new user needs. Next to these demands there is a strong incentive to collect data more efficiently and to reduce the costs of data collection. The old design was not flexible enough and was also too costly.
8. A number of means was distinguished for the design that better would meet the objectives mentioned above.
 - (a) The use of registers and administrative data:
 - as a surrogate for data that are collected by surveys;
 - as auxiliary information for sampling, data collection strategies, processing and analysis.

The main advantages of secondary data collection are costs and size. Drawbacks are a loss of independence and risks in timeliness of statistics.

 - (b) The use of the cheaper data collection modes: internet and paper. These modes are cheap and to some extent respondent-friendly. Also response bias may be reduced in surveys that are sensitive for social desirable answers. However, response rates are low and interviewing of complete households becomes more difficult. Furthermore, these modes may lead to mode effects on other survey items.
 - (c) Combinations of survey modules and themes. Some question modules or blocks may be included in more than one survey, thereby enhancing coherence, transparency and efficiency. Coherence is improved as more cross-relations can be made. Transparency is better as definitions are harmonized. Efficiency may be improved as the survey items for which accuracy demands are highest, i.e. employment and education, are spread over different surveys.
 - (d) Re-use of respondents. Respondents can be randomly selected for a follow-up survey, possibly with the aid of screener questions. There are a number of obvious benefits and drawbacks in terms of costs and survey error.
9. Means (b) and (c) evolved into a new standard collection strategy for social statistics with the use of sequential mixed mode. At first we approach persons or households with an advance letter in which we invite them to participate in the survey by filling in an internet questionnaire. After one week respondents receive a reminder letter. After two weeks, the internet mode is switched off and non-respondents are approached by telephone (in case of a known telephone number) or approached by face to face interviewers otherwise. This strategy was first used in 2009 in the

³ eXtensible Business Reporting Language

⁴ CATI: Computer Assisted Telephone Interviewing

⁵ CAPI: Computer Assisted Personal Interviewing

housing survey. Since 2012, also the Labour Force Survey is conducted in a fully mixed mode approach.

10. Introducing the internet mode resulted in some serious challenges with capacity planning and with maintaining a nationwide coverage of the fieldwork force because of the strong reduction of the overall capacity needed resulting from our regular working program. This paper focusses on these organizational aspects of introducing mixed mode.

III. The challenges we faced

11. After experimenting a couple of years with internet questionnaires for social surveys Statistics Netherlands introduced in 2009 the first fully mixed mode survey: the housing survey. The survey as a whole showed good results, but in practice we faced some side effects that were repeated in other surveys. This chapter describes these effects.
12. The use of sequential mixed mode introduces dependencies in the planning process. First, the internet (CAWI⁶) response determines the total amount of sample units (non response) that will be put through to the other two modes CAPI and CATI. The second dependency is the rate of telephone numbers that can be found for these sample units. Only when a telephone number is found, the sample unit will be approached by CATI. All remaining sample unit will be approached by using the –most expensive- CAPI mode. So, the amount of work in face to face interviewing is strongly affected by internet response and enrichment rate of telephone numbers. This planning of field work capacity starts one month ahead, which means that fluctuations during the survey period are hard to handle and results in extra costs.

A. Internet response

13. In the last years we found the internet response rate varying between different surveys. For example, the mobility survey accomplishes an internet response rate of approximately 21 percent, the labour force survey accomplishes 25 per cent and the health survey accomplishes 35 per cent. In 2010 the youth survey accomplished an internet response rate of 45 per cent. We suppose that the subject of survey strongly determines the willingness of respondents to participate in the specific survey. For planning purposes it is important to have a good prediction of the expected response rate and in the case of mixed mode surveys with dependent modes it is even more important. The more surveys we did, the better the expectations became.

B. Enrichment of telephone numbers

14. The other dependent variable, the enrichment of telephone numbers, also showed a whimsical pattern in the last two years. There are a few causes. First, we see that a lot households move from traditional landlines to mobile telecom providers and cable network providers. The coverage of phone numbers therefore decreased in the last years. This results in a greater portion of sample units that are being put through to face to face interviewing. Second, the telephone number provider made various changes in their production process that resulted in lower accuracy of the phone numbers received. Especially mobile phone numbers proved to be of poor quality. This quality problem is different from the coverage problem, because the logistic preparation time of face to face interviewing is much longer than telephone interviewing. Therefore it is not possible to put these sample units through to face to face interviewing in the same month. It directly results in a lower total response rate.

⁶ CAWI: Computer Assisted Web Interviewing

15. Summarizing, the above response by internet and fluctuating enrichment rate of telephone numbers can be described as a challenge to cope with fluctuations in the capacity need for face to face interviewing.

C. Nationwide coverage of the field workforce

16. One of the means for redesigning the social surveys was to increase the use of the cheaper data collection modes like internet and paper. The increase of these survey modes results directly in a decrease of the more costly data collection modes like telephone and face to face interviewing. Especially the last mode has a certain minimum in capacity need for maintaining a nationwide coverage of the field workforce. The more an interviewer has to spend time on travelling, the less effective time remains for interviewing. Hence, there is a specific point where the available interviewing time is too low for covering the amount of work. This can only be solved by expanding the available interviewing time (by hiring more interviewers or increasing the contract hours). Research showed that a cost effective size of the field workforce is found at 110 full time equivalents. Knowing this, the total working program for the field workforce has to more than this minimum. The actual working program was in the last years heavily affected by the introduction of mixed mode surveys. In 2011 and 2012 the regular working program (without commissioned work) required only 80 fte. In other words, the introduction of mixed mode surveys was successful in reducing the need for costly data collection modes, but introduced a new challenge to maintain the nationwide coverage without increasing the total cost of responses.

D. Complex survey designs

17. The last challenge we want to mention in this paper has to do with the application support for our survey process. As we stated earlier, the use of mixed modes introduced new dependencies into our planning process. A mixed mode survey design consists of many parallel and sequential survey steps to take. These designs are far more complex than the traditional survey designs that only make use of the CATI or CAPI modes. Our survey management application however was not equipped for the newer survey designs. Besides that, the survey management application already was an incrementally built application with complex source code and badly documentation. With the perspective of a completely new survey management application it was unwise to redesign or adjust the application. Workarounds were found in manual tasks. The increase in manual tasks however leads to higher risk of failure and increase in the workload.

E. Challenges

18. Taking all the above mentioned issues into account the challenges of the introduction of mixed mode surveys can be summarized as:
 - (a) How can we achieve a flexible field workforce that can cope with fluctuations in capacity needs and also provides a nationwide coverage for their activities.
 - (b) How can we achieve a manageable survey process.

IV. Applied measures

19. Several measurements are implemented to improve the robustness for fluctuations of workload.

A. Logistical processes

20. The maintenance of the logistic processes is optimized. In the last 5-10 years Statistic Netherlands introduced the use of business architecture. The use of Business Architecture helps to harmonized the processes and makes the interfaces of the subsequent processes more transparent. This increased transparency makes it possible to determine the impact of changes

within processes en survey quicker and more accurate. This helps us to become more adjustable. In the coming years further improvement is necessary especially to reduce the planning cycle of CAPI from a monthly cycle to a weekly cycle. In 2013 one centralized logistic survey management system will be introduced. This system enables the handling of complex survey designs and monitors planning en response on a detailed level. The processes become less labour-intensive, but more important the response can be optimized and it enables us to adapt new surveys. The workload of the more labour intensive data collection modes of face to face interviewing and telephonic interviewing benefit limitedly from these measurements.

B. Labour force

21. To be able to adjust for fluctuation in workload for more labour intensive data collection the legal status of the employers is revised. Furthermore, Statistic Netherlands started a corporation with an employment agency to be able to work with more flexible contracts. To manage this flexible workforce regional managers are appointed. In 2012 Statistic Netherlands experimented with a co-operation with commercial partners. The new approach lead to a higher turnover of staff and decreased the efficiency. In 2013 we will investigate several scenarios with different combinations of permanent and temporary staff and/or co-operation (contracting out data collection) with commercial partners.

C. Instruction of agents

22. To maintain the high quality standards, the instruction of agents has to meet high standards. Statistic Netherlands introduced blending learning techniques to maintain the high standards although the training period is reduced. The instruction of the tooling and knowledge of the different surveys is trained by an e-learning module. The competence to persuade the respondent and other communication skills are trained interactively in small groups. The introduction period is reduced from 3 month to less than 2 months.

D. Regional allocation

23. For face to face interviewing the regional allocation is a complicating factor. The elements within the survey sample are firstly addressed to a certain region and subsequently within the region over the agents. This allocation is optimized to minimize the travel distance (travel time) for all agents. By directly allocating the elements to the agents the allocations become more optimal.

E. CATI at Home

24. New technology enables agents to do telephonic interview at home. The face to face agents are now trained in order to be re-allocated to telephonic interviewing so fluctuation in CATI-CAPI workload can be managed. The workload for CAPI is decreasing. This causes a relative increase of travelling time for the agents and therefore a decrease in efficiency. Therefore Statistic Netherlands will investigate the possibilities for full integration; in which all agents are doing both CATI and CAPI.

F. Thinning Techniques

25. The workload of the field organisation is managed on forehand by using thinning techniques. The design contains two stages. In the first stage the full survey is carried out using CAWI. In the subsequent stage a number of non-responding units are appointed to each region and to CATI interviewer. This number is determined on forehand. Thereby the productivity of the field organisation is optimized and the response strategy becomes more predictable. Within a survey design the first modes are less costly. Statistic Netherlands is now investigating which design in terms of initial sample size in relation to the portion of CATI and CAPI is optimal.
26. In the coming years Statistics Netherlands is aiming to introduce adjustive designs to improve the effectiveness of our surveys.

G. Benefit of scale

27. Statistic Netherlands aims to centralize the data collection for all governmental surveys. The need for commissioned work is increased in order to maintain the nationwide coverage of our field workforce and become more robust. In the last 5 years up to 25% of the portfolio contained surveys for third parties (especially ministries and agencies). Most of the surveys are (semi-) continuous which makes it possible to build a long term relationship with these partners.

V. Concluding remarks

28. Statistics Netherlands benefits from the full implementation of mixed mode designs for most person and household surveys. The foreseen increase of effectiveness is realized. The implementation of the new data collection strategy created the need to take several measurements to be able to cope with the increasing fluctuation and complexity of data collection processes. In the coming years we see new possibilities for further improvement of both processes and data collection designs.

VI. References

Hermans, H., G. Snijkers, and M. Roos (2011). CBS Data Collection Strategy, version 2011: "Collection and delivery" (in Dutch: CBS-brede waarneemstrategie, versie 2011: "Halen en brengen"). Report Statistics Netherlands, Heerlen/The Hague.

Ivanova, I. (2010), Optimizing interviewing efficiency as a function of number of interviewers and their locations, Postmasters study program Mathematics for Industry, prepared for Statistics Netherlands, ISBN 978-90-444-0926-0

Schouten, B. (2008), Quality indicators and the re-design of household data collection at Statistics Netherlands, Statistics Netherlands, The Hague.

Snijkers, G., R. Göttgens, and H. Hermans (2011), Data collection and data sharing at Statistics Netherlands: yesterday, today, tomorrow. Paper presented at the 59th plenary session of the Conference of European Statisticians, 14-16 June 2011, Geneva, Switzerland.