 

|  |
| --- |
| **Modular design for the Dutch Crime and Victimization Survey**  Peter Lugtig, Vera Toepoel Utrecht University  & Annemieke Luiten, Statistics Netherlands |

**Introduction**

One obvious way to reduce breakoffs in mobile web surveys, is to reduce the length of questionnaires. Breakoffs increase with the length of the questionnaire about linearly (Bakker, 2016), meaning that shortening the questionnaire by half could reduce breakoff rates by half as well. This effect is perhaps even stronger when respondents anticipate that the questionnaire is short. A literature review focusing on mail surveys showed that reducing the expected duration of the survey in an advance letter increases response rates, although the effects were not strong (Bogen, 1996).

This paper discusses a practical question: how to shorten questionnaires from a survey practice perspective? As an example case for studying this question, we use the Dutch Crime and Victimization Survey. This is an annual survey that takes respondents 20 minutes on average to complete. In this paper, we set out different ways to shorten the questionnaire without losing information on key statistics. The way to do this is to not ask all respondents all survey questions. This procedure is also known as a planned missingness design, matrix sampling, data chunking or modularization (Graham et al., 2006, Rhemtulla & Little, 2012). The key question answered in this paper is how modularisation should be implemented in practice so that 1) the respondent receives a good questionnaire and 2) as little statistical information is lost as possible.

**Motivating example: the Dutch Crime and Victimization Survey**

The Dutch Crime and Victimization Survey is an annual, cross-sectional survey focusing on feelings of safety and victimhood among the Dutch population aged 15 years and older. It is carried out by two fieldwork organisations: Statistics Netherlands draws a nation-wide sample of about 67.000 individuals from the Dutch population register. Local councils can boost sample sizes so that they can get a more fine-grained picture of safety and victimhood in their area. The fieldwork for these boost samples are carried out by I&O Research, and include a mixed-mode component. In this paper, only CAWI data are used that are collected both by Statistics Netherlands and I&O Research (total sample size 53,605) for the year 2015. All analyses use unweighted data.

The questionnaire of the Dutch Crime and Victimization Survey consists of 14 sections. There is little routing in the questionnaire, apart from a section on victimization. The questionnaire is not extremely long, but probably too long for mobile completion.

**How to create a modular design: 2 approaches**

The Crime and Victimization Survey will serve as a motivational example. The goal of this study is to reduce the interview duration in the Crime and Victimization Survey by about half, leading to a new average duration of about 10 minutes for each respondent.

The core idea of modularization in planned missingness designs is that missing data are generated by the researcher and that therefore there is full control of the correlated of missingness. Figure 2 illustrates the idea of planned missingness: the questionnaire is split into a certain number of components (5 in the Figure). Each respondent receives the questions in the core component, and is assigned randomly to 2 out of 4 remaining modules. In Figure 2, there are six versions of the questionnaire, but researchers may opt to have fewer or more versions. In Computerized Adaptive Testing, the idea of planned missingness is extended in such a way that individualized questionnaires result. What is important in any design is that there should be overlap between questions so that covariances between all items can be estimated. Questions in the core component are answered by all respondents. Questions used for routing are therefore typically put into the core component, and so are variables that are deemed to be important for recovering the missing data that is created by modularization.

Missing data resulting from modularization is usually Missing Completely At Random (MCAR), meaning that respondents are assigned to one of the planned missingness conditions randomly. Given MCAR, missing data can be easily recovered by imputation. For each variable, a missing data model is created, and missing values are predicted per individual. Modern methods use Markov Chain Monte Carlo methods to semi-automatically specify the imputation model, and impute multiple values to account for uncertainty in the imputed values (van Buuren, 2012). The stronger the covariates in the imputation model predict the missing values, the lower the uncertainty in the imputation model, and the smaller the variance between the imputed values over the multiple imputations. Under MCAR, large proportions of missing data in a variable are unproblematic when strong predictors of that variable are available. The reverse is however also true. Small proportions of missing data can actually be problematic when there are no good predictors in the model.

The first approach to create a planned missingness design follows the logic of statistical efficiency. The general idea is that variables are split into modules according to the inter-item correlations[[1]](#footnote-1). Any two Items with high inter-item correlations should appear in two different modules so that good predictors are available in the imputation process.

This statistical approach that relies on the strength of correlations to create modules is ignorant of any survey-context. The assumption is that items can be easily moved from one component to another without consequences. This assumption is strong and will seldomly hold in practice. Almost all surveys contain routing questions used as filters for example. Filter questions cannot easily be moved around in the questionnaire without messing up the routing of the questionnaire completely.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Questionnaire version | Core module | Module A | Module B | Module C | Module D |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |

*Figure 1: hypothetical planned missingness design. Cells in grey are asked to respondents, cells in white are skipped.*

The second approach to creating planned missingness designs addresses the weakness of just focusing on statistical efficiency alone. It takes the survey context into account, by trying to preserve the logic of the questionnaire as much as possible. If a questionnaire consists of sections that each cover a separate topic, it for example makes sense to modularize questionnaire sections, instead of individual questions. In other words, in this approach the components in figure 2 usually correspond to the questionnaire sections. Another intuitive approach is to reduce the number of items in multi-item scales in each section. So when a scale consists of six items, the items are split in 2 times three items that will appear in different modules. The most important principle of this approach is the missingness design is adapted to the questionnaire and population under study.

The two approaches outlined above should be seen as two extreme scenarios. In reality, the two approaches can and should be mixed. We now turn to our example to show how the two approaches would work for a short section of the Crime and Victimization Survey.

**A simulation study: how to assign items to modules?**

In order to show the basic principle and effects of assigning items to different modules, we perform a small simulation study. We selected 15 items from the Dutch Crime and Victimization Survey (Q01-Q05, Q06,Q13, Q29, and Q31-Q38 from the survey. These questions comprise the central questions of sections 1,2 and 3 in the questionnaire. Following the example shown in Figure 1, we next create 5 modules: one core module that will be offered to all respondents, and 4 modules that will randomly be offered. In our simulation, we assign 3 items to each module.

The first step is to determine which of the 15 items should go into the core module. As there is no routing within this subset of items, all items can potentially go into the core module. There are many possible heuristics for selecting items for the core. In our example we used the heuristic of lowest shared variance. We used all 15 items in a Principal Component Analysis and looked at the 3 items with the lowest communality scores (i.e. the largest unique variance). In other words, we are selecting the items that are most difficult to predict using the other variables in the dataset.

The bottom row of Table 1 shows that the communalities are lowest for Q02, Q04 and Q05. Note that, had we chosen fifteen different items, we would have possibly come to a different decision, as the communalities are computed given the items in the current set.

Table 1: correlations and communalities of the fully observed dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Correlations |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Q01 | Q02 | Q03 | Q04 | Q05 | Q13 | Q29 | Q31 | Q32 | Q33 | Q34 | Q35 | Q36 | Q37 | Q38 |
| Q01 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Q02 | **.68** | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Q03 | .31 | .30 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |
| Q04 | .26 | .28 | .**27** | 1 |  |  |  |  |  |  |  |  |  |  |  |
| Q05 | .24 | .25 | .23 | .49 | 1 |  |  |  |  |  |  |  |  |  |  |
| Q13 | -.26 | -.23 | -.19 | -.17 | -.15 | 1 |  |  |  |  |  |  |  |  |  |
| Q29 | -.13 | -.11 | -.18 | -.09 | -.09 | .32 | 1 |  |  |  |  |  |  |  |  |
| Q31 | -.11 | -.09 | .14 | .07 | -.06 | .27 | .39 | 1 |  |  |  |  |  |  |  |
| Q32 | -.12 | -.10 | -.16 | -.07 | -.06 | .32 | **.48** | .42 | 1 |  |  |  |  |  |  |
| Q33 | -.13 | -.10 | -.21 | -.09 | -.08 | **.36** | **.61** | **.51** | .63 | 1 |  |  |  |  |  |
| Q34 | -.11 | -.09 | -.17 | -.10 | -.10 | .29 | .43 | **.48** | .41 | **.52** | 1 |  |  |  |  |
| Q35 | -.13 | -.11 | -.16 | -.08 | -.09 | .29 | .47 | .44 | .41 | **.51** | **.55** | 1 |  |  |  |
| Q36 | -.15 | -.13 | -.15 | -.08 | -.08 | .33 | .37 | .25 | .32 | .37 | .24 | .36 | 1 |  |  |
| Q37 | -.13 | -.13 | -.13 | -.06 | -.10 | .22 | .26 | .19 | .21 | .23 | .21 | .26 | .42 | 1 |  |
| Q38 | -.26 | -.23 | -.25 | -.17 | -.15 | .63 | .48 | .38 | **.45** | **.51** | .41 | **.46** | **.58** | **.39** | 1 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Communalities | .15 | **.13** | .15 | **.08** | **.07** | .37 | .49 | .38 | .45 | .58 | .43 | .47 | .36 | .21 | .63 |

Notes. N= 53605. We used a Principal Component analysis and forced the solution into 1 factor. Number in blue show lowest communalities, which were used as heuristic for assigning items to Core module. Items in red show high correlations, which were used to assign items to different modules (A-D).

The second step is determining how to assign the remaining 12 items into the modules that will be used in the planned missingness design. For this we use two approaches.

1. In the intuitive approach, we group items based on what would make a consistent questionnaire for respondents. In practice this means that respondents may skip entire sections of the questionnaire. In our simulation, it implies that items are grouped according to where they appear in the questionnaire. Module A will consist of Q01, Q03 and Q13. Module B will consist of Q29, Q31 and Q32. Module C of Q33-Q35 and Module D of Q36-Q38
2. In the statistical approach, we separate items with high correlations. We used a simple heuristic for this. For each item, we ranked the correlations from high to low. Then we tried to separate the items with the highest correlations, taking into account the fact that some items (such as Q38) is a strong predictor of many variables. Module A will consist of Q01, Q33 and Q36. Module B of QQ03, Q34 and Q37. Module C of Q13, Q35 and Q38 and Module D of Q29,Q31 and Q32.

Missing data patterns are created according to Figure 2. A total of 6 missing data patterns are created in both the intuitive and statistical approach. Each item in modules A-D will have 50% missing values. In order to speed up the imputation process, we selected at random 5.000 cases from the total sample for our statistical analyses.

In the next step, data are re-imputed under the 2 datasets (the intuitive and statistical planned missingness dataset). 40 datasets are imputed for each missing value, using the Mice algorithm in R (van Buuren, 2012). We use Predictive Mean Matching to ensure that the imputations matched the variable range of the original variables. The relatively large number of imputations is necessary to consistently estimate standard errors. After imputations, we look at both the standard error of the mean for both the complete dataset, and the two datasets where missingness was simulated. We compare the increase in the standard error between the three 3 datasets.

**Results**

Our findings confirm our expectations. The standard errors of the mean are about 0,21% higher in the intuitive planned missingness approach as compared to the full dataset. This is a negligible increase. The reason that our standard errors are not inflated because of the missingness design is probably because we have used Predictive Mean Matching to impute missing values. Standard errors for means are extremely small for the complete dataset, and the imputed values do not add to uncertainty in our estimations, because the means are consistently estimated. In short, when it comes to means, both the intuitive and statistical approaches to planned missingness result in a dataset with about the same precision for estimating means.

Table 2: Results for precision in means of the simulation study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Means |  |  |  |  |
|  | Complete data s.e. | Statistical missings s.e. | Intuitive missings s.e. | % increase statistical method | % increase intuitive method |
| Q01 | 13,93 | 13,96 | 14,00 | 0,23 | 0,52 |
| Q03 | 11,92 | 12,02 | 12,18 | 0,86 | 2,13 |
| Q13 | 19,03 | 19,41 | 19,74 | 1,99 | 3,75 |
| Q29 | 5,20 | 5,12 | 5,12 | -1,50 | -1,50 |
| Q31 | 8,92 | 8,74 | 8,95 | -2,07 | 0,35 |
| Q32 | 5,98 | 5,95 | 5,91 | -0,48 | -1,24 |
| Q33 | 7,04 | 7,09 | 7,09 | 0,70 | 0,70 |
| Q34 | 6,02 | 5,93 | 6,02 | -1,46 | 0,12 |
| Q35 | 7,40 | 7,50 | 7,48 | 1,43 | 1,12 |
| Q36 | 7,14 | 7,23 | 7,13 | 1,15 | -0,13 |
| Q37 | 12,44 | 12,32 | 12,55 | -1,02 | 0,84 |
| Q38 | 17,06 | 17,07 | 16,53 | 0,05 | -3,10 |
| Mean | 10,17 | 10,19 | 10,23 | 0,21 | 0,52 |

Notes: N=5000. s.e in means of variables multiplied by 1000. We use Barnard-Rubin pooling rules for pooling the standard error for the multiply imputed datasets.

For regression coefficients, our results are different. Here, we do find inflated standard errors because of the planned missingness design. After imputing values, we used Q04 (that was in the core module) as a predictor for each of the twelve variables assigned to one of the modules with missings.

Table 3 shows that on average, standard errors are 37% larger in the intuitive approach to creating missingness, and 33% in the statistical approach. It is important to remember that because of the fact that we induced 50% missing values in each variable, the standard errors were expected to increase by 41% (square root of 2) had there been no correlations between items.

The mean increase in standard errors obscures that there is large variation between variables. For some variables, we use that standard errors are only inflated by 16 or 21% (Q29), while for some other variables we observe increases in standard errors of 52 and 59% (Q03). Technically, it is possible that imputations add extra imprecision, so it is possible for standard errors to increase by more than 41%. This can happen when no good imputation model is available. Another possible reason for the large fluctuations in standard errors, is that we did not use a sufficient number of imputations. We re-ran the model with 5,10,30 and 60 imputations, and found only small differences between the models with 40 and 60 imputations. The variation in precision that we observe for each variable is due to the different patterns of missingness that were used in the intuitive and statistical approach

This means that if one wants to implement a planned missingness design in practice, one would need to enlarge the sample size by 0% in order to obtain the same precision for means and about 35% for the same precision in relations (regression coefficients). One striking conclusion is that for precision, it does not really matter how items are assigned to modules. The difference in standard errors between the intuitive and statistical method are only 5%.

Table 3: Results for precision in regression coefficients of the simulation study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Means |  |  |  |  |
|  | Complete data s.e. | Statistical missings s.e. | Intuitive missings s.e. | % increase statistical method | % increase intuitive method |
| Q01 | 33,98 | 44,94 | 43,78 | 32,27 | 28,84 |
| Q03 | 29,50 | 44,79 | 47,01 | 51,83 | 59,36 |
| Q13 | 93,99 | 114,41 | 143,92 | 21,72 | 53,12 |
| Q29 | 13,29 | 15,50 | 16,10 | 16,64 | 21,11 |
| Q31 | 22,85 | 30,46 | 29,53 | 33,28 | 29,22 |
| Q32 | 15,12 | 17,81 | 17,65 | 17,80 | 16,75 |
| Q33 | 17,80 | 21,30 | 21,34 | 19,65 | 19,88 |
| Q34 | 15,20 | 18,96 | 19,09 | 24,75 | 25,59 |
| Q35 | 18,92 | 24,28 | 23,96 | 28,33 | 26,66 |
| Q36 | 18,18 | 29,95 | 25,60 | 64,70 | 40,80 |
| Q37 | 32,31 | 48,71 | 42,42 | 50,78 | 31,31 |
| Q38 | 57,92 | 79,31 | 76,76 | 36,92 | 32,51 |
| **mean** | **30,76** | **40,87** | **42,26** | **32,88** | **37,42** |

Notes: N=5000. s.e. for regression coefficients multiplied by 1000.

12 separate regressions, each time with Q04 (complete variable) as predictor.

**References**

* Bakker, J. (2016) Unintentional mobile respondents in official statistics and their consequences for data quality. 71st AAPOR conference, 12-15 May, Austin, Tx.
* Barnard, J. & Rubin, D.B. (1999) Small-sample degrees of freedom with multiple imputation. Biometrika, 86. P. 948-955.
* Bogen, K. (1996) The effect of questionnaire lenth on response rates – a review of the literature. U.S. Census Bureau papers SM96/05, <https://www.census.gov/srd/www/abstract/sm9605.html>
* Couper, M.P., Antoun, C., Mavletova, A. (2017). Mobile Web Surveys: A Total Survey Error Perspective. In Biemer, P. et al. Total Survey Error in Practice. New York: Wiley, pp 133–54.
* Graham,.J.W., Taylor, B.J., Olchowski, A.E. & Cumsille, P.E. (2006) Planned missing data designs in psychological research. Psychological Methods, 11. 323-343. DOI:[10.1037/1082-989X.11.4.323](https://doi.org/10.1037/1082-989X.11.4.323)
* Ioannidis, E., Reis, F., Calizzani, C., Gras, F., Karlberg, M., Merkouris, T., Stavropulos, P, and Zhang, L. (2015) Optimum allocation of variables in a modular survey architecture. NTTS 2015 poster presentation.
* Johnson, A., Kelly, F., Stevens, S. (2012). Modular survey design for mobile devices. A research study on the potential uses of modular survey design for mobile and online surveys. Paper presented at the CASRO Online Research Conference, Las Vegas, March 1-2.
* Johnson, E. (2015) Mobile devices and modular survey design. Survey Sampling International course.
* Kelly, F., Johnson, A., and Stevens, S. (2013). Modular Survey Design: Bite sized chunks 2. Paper presented at the CASRO Online Research Conference, San Francisco, March 7-8.
* Mavletova, A. & Couper, M.P. (2015) A meta-analysis of breakoff rates in mobile web surveys. Mobile research methods. P.81-98 In Toninelli, D, Pinter, R & de Pedraza, P. (eds.) Mobile research methods: Opportunities and challenges of mobile research technology. London: Ubiquity Press.
* Mäenpää. C., (2001). Blockbildung und Methodenevaluation im Rahmen des Fragebogensplittings. Master Thesis, Nüremberg, Germany.
* Merkouris, T. (2010). An estimation method for matrix survey sampling. *ASA, Proceedings of the section on survey research methods*, 4880-4886.
* Mirza, I., and Gafur, M. (2014?) Survey chunking: higher data quality in mobile research.
* Raghunathan, T., and Grizzle, J. (1995). A split questionnaire survey design. *Journal of the American Statistical Association,* 90, 54-63.
* Rässler, S., Koller, F. and Mäenpää, C. (2002) A Split Questionnaire Survey Design applied to German Media and Consumer Surveys . Diskussionspapiere Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie 42b/2002

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.116.3161&rep=rep1&type=pdf>

* Rhemtulla, N, & Little, T. (2012) Tools oft he trade: Planned missingness designs for research in cognitive development. Jpourrnal of Cognitive Development, 13. Doi: 10.1080/15248372.2012.717340
* Smith, R., Kotzev, I., Miller, C., and Kachhi, D. (2012). Modularizing Surveys to meet today’s respondent challenges. Paper presented at the CASRO Online Conference, Las Vegas, March 1-2.
* Toepoel, V., & Lugtig, P. (2015) Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods, Data, Analysis*, 9(2). DOI:10.12758/mda.2015.009
* Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge, U.K: Cambridge University Press.
* Van Buuren S (2012). *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Press, Boca Raton, FL. 342 pages. ISBN 9781439868249
* Wagner, J. (2010) Fraction of Missing Information as a tool for monitoring the quality of survey data. Public Opinion Quarterly, 74, 223-243. Doi [10.1093/poq/nfq007](https://doi.org/10.1093/poq/nfq007)
* West, B., Ghimire, D., and Axinn, W. (2015). Evaluating a modular design approach to collecting survey data using text messages: evidence from Nepal. Population Studies Centre Research Reports.

1. [↑](#footnote-ref-1)