RISQ

Representativity Indicators for Survey Quality

RISQ – Fieldwork Monitoring

Work Package 6

Deliverable 5, version 2

Geert Loosveldt & Koen Beullens

Katholieke Universiteit Leuven

September 1st, 2009

SEVENTH FRAMEWORK PROGRAMME

The RISQ Project is financed by the 7th Framework Programme (FP7) of the European Union.

Cooperation Programme, Socio-economic Sciences and the Humanities, Provision for Underlying Statistics

1 Introduction

Development of Representativity Indicators for Survey Quality (RISQ) is a project funded by the European 7th Framework Programme (FP7). The NSI's of Norway, the Netherlands and Slovenia and the universities of Leuven and Southampton jointly seek to develop quality indicators for survey response. In this contribution, the focus is on the monitoring of R-indicators while the survey fieldwork is in motion and how different survey actors such as interviewers, management and (non)respondents influence the quality of the realized sample. We will try to systematically describe, underpin and demonstrate the possibility to use R-indicators for fieldwork monitoring. Therefore, a theoretical framework will be presented allowing to interpret and model the process of sample construction in terms of representativeness. Afterwards, the Belgian part of the European Social Survey will be used to demonstrate these monitoring procedures. As this deliverable deals with a particular issue where R-indicators are involved and as it builds on previous research efforts of the RISQ-project, it is appropriate to first revisit the basic concepts of the R-indicator.

2 R-indicators revisited

The motivation to start this project originates from the observation that whereas response rates are almost always provided they seldom indicate how balanced the composition of the sample is. Next to a response rate, an additional indicator is therefore needed in order to assess the quality of a survey sample. Hence, the R-indicator measures the degree to which the group of respondents resembles the entire sample or population. The basic idea behind R-indicators starts from the definition of a representative sample: all members of the gross sample have an equal probability (or at least a known probability) of being included in the net sample. Whenever this requirement is not met, the variance of the response probabilities, also called response propensities, grows and indicates the degree of non-correspondence between respondents and nonrespondents (Cobben & Schouten, 2007; Schouten & Cobben, 2007). A vital role with respect to the use of R-indicators lies in the auxiliary information. Auxiliary variables are available for both respondents and nonrespondents and permit to estimate the response propensities. In the case of a household survey, one typically thinks of age and gender information or other register information. Observable data can also be used such as the interviewers' appraisal for the households' dwelling or neighbourhood. If zip-codes or similar points of attachment are available, area level information such as population density or average income levels can also be linked to the data fund.

Previous RISQ-deliverables have paid attention to the statistical properties of R-indicators (deliverable 2 by Shlomo, Skinner, Schouten, Bethlehem & Zhang, 2009), how to use R-indicators (deliverable 3 by Schouten, Morren, Bethlehem, Shlomo & Skinner, 2009) and partial R-indicators (deliverable 4 by Shlomo, Skinner, Schouten, Carolina & Morren, 2009). These deliverables are available at www.r-indicator.eu. Some elementary definitions, properties and general guidelines from this existing material will briefly be presented here in order to improve the readability of the main sections of this deliverable. We will first introduce the R-indicator itself and relate it to nonresponse bias. Thereafter, the notion of partial R-indicators will be discussed.

The expression $r \subset s \subset U$ denotes that a sample s has been drawn from population U with probability $p(s)$ and that r is the realized response in gross sample s. Let R_i be the response indicator variable so that $R_i = 1$ if unit i responds and $R_i = 0$, otherwise. Ideally, the response propensities ρ_i are estimated by repeating (independent) survey requests and measure the proportion of successful attempts. In practice, however, (cross-sectional) surveys usually dispose of one replication so that a response propensity ρ_i is only a hypothetical construct. Instead, assume an auxiliary variable ℵ that is capable of fully explaining the response behaviour of the population units. Due to practical limitations, ℵ will in practice be replaced by a set of available auxiliary variables to approximate the response propensities ρ_i. We therefore assume the availability of auxiliary variables x_i for all elements in the sample s.

A definition of strong representativeness applies when $\rho_{\aleph}$ is a constant function or

$$\rho_i = P(r_i = 1 \mid s_i = 1) = \bar{\rho}, \quad \forall i .$$

Weak representativeness is obtained when the response to a survey is constant with respect to some auxiliary variable(s) x. It is evident that only weak representativeness can be evaluated, as the 'true' or 'super' vector ℵ is not available.

The dependence of ρ_i on x_i can be expressed in the form:

$$\rho_i = g^{-1}(x_i' \beta),$$

where $g(.)$ is a specified link function such as the logit, probit or identity (linear) link function. The choice of the link function only seems to have a minor influence, although further investigations on this topic are needed (see deliverable 3 by Schouten, Morren, Bethlehem, Shlomo & Skinner, 2009).

Now, the variance of the response propensities is obtained by

$$\hat{S}(\hat{\rho}_X) = \sqrt{\frac{1}{N-1} \sum_s d_i (\hat{\rho}_X(x_i) - \hat{\bar{\rho}}_X)^2},$$

where d_i is the design weight. Related to this variance, the R-indicator is expressed as:

$$\hat{R}(\hat{\rho}_X) = 1 - 2\hat{S}(\hat{\rho}_X).$$

It can be shown that the maximal variance is ½ (where $\hat{\bar{\rho}}_X = ½$), and as a consequence, $R(\rho_X) \in [0,1]$.

The variance of response propensities and its related R-indicator can directly be linked to the nonresponse bias. For any binary survey item y, with $\bar{y} = ½$, the maximal possible bias can be defined as:

$$Bias(y_{dr}) \leq \frac{1 - R(\rho)}{4\bar{\rho}}.$$

So-called partial R-indicators can be used to determine the impact of a particular auxiliary variable on the variance of the response propensities. For fieldwork monitoring these partial indicators are very useful since they suggest under- or overrepresented groups in the sample that may need more (less) effort in order to improve the representativeness of the obtained response set. Partial R-indicators basically deal with the allocation of the total variance of the response propensities to the between and within structures of the auxiliary variables, conditional or unconditional on each other. Four variants of partial R-indicators have been developed, of which the first two (P_1 & P_2) allocate the total variance unconditional on other auxiliary variables, whereas P_3 & P_4 treat the variance allocation conditionally.

Suppose X and Z are auxiliary variables with respectively k and l strata. Both variables are used to model response propensities ρ_i, then

$$P_1(Z,k,\rho_{X,Z}) = \sqrt{S_b^2(\rho_{X,Z} \mid Z=k)} = S_b(\rho_{X,Z} \mid Z=k)$$

$$P_2(Z,k,\rho_{X,Z}) = S_b(\rho_{X,Z} \mid Z=k)\frac{(\bar{\rho}_{X,Z,k} - \bar{\rho}_X)}{\mid \bar{\rho}_{X,Z,k} - \bar{\rho}_X \mid} = \sqrt{\frac{N_k}{N}}(\bar{\rho}_{X,Z,k} - \bar{\rho}_{X,Z})$$

Note that $P_1 = \mid P_2 \mid$. P_2 clearly shows that the distance between the mean propensity in stratum k of Z and the overall mean propensity is measured. The partial indicator P_1 can also be computed on the variable level:

$$P_1(Z,\rho_{X,Z}) = S_b(\rho_{X,Z} \mid Z)$$

Note that it is also possible to determine the between variance S_b^2 with regard to the strata of Z when Z is not used to estimate the response propensities. We will, however, not apply this possibility for fieldwork monitoring.

Conditional partial indicators allocate variance that exclusively belongs to a particular auxiliary variable. This can clearly be seen in the next expressions:

$$P_3(Z,\rho_{X,Z}) = \sqrt{S_W^2(\rho_{X,Z} \mid X)} = S_W(\rho_{X,Z} \mid X)$$

$$P_3(Z,k,\rho_{X,Z}) = \sqrt{\frac{1}{N-1}\sum_{l=1}^{L}\sum_{U_l}\delta_{k,i}(\rho_{X,Z}(x_i,z_i) - \bar{\rho}_{X,Z,l})^2}.$$

P_3 deals with the within variance S_w^2 with respect to the categories of Z after the effects of X have been removed.

The final conditional partial R-indicator P_4 has a similar meaning as P_3, as it measures the added value of auxiliary Z to the overall R-indicator, provided that X is already included as an auxiliary variable:

$$P_4(Z,\rho_{X,Z}) = R(\rho_X) - R(\rho_{X,Z}) = 2(S(\rho_{X,Z}) - S(\rho_X))$$

In general, the R-indicator and its partial variants can be considered as valuable measurement instruments to assess the representativity of the sample. Moreover, the direct relation with nonresponse bias is an interesting feature from a practical point of view. However, some comments may be appropriate in order to interpret its usefulness and limits. First, and this is the most common critical issue, the quality of the R-indicator very strongly depends on the availability of auxiliary variables. Hence, one can only claim weak representativity. Strong representativity claims are only valid when key auxiliary variables can be disposed of and this is obviously unapproachable in practice. Therefore it is recommended to present R-indicators together with the auxiliary variables. Also, when comparing different survey or different fieldwork configurations within the same survey, the same set of auxiliary information should be used, including the same classification of the variables and interactions between the variables. Second, the R-indicators are estimated based on a random sample, so that they can be considered as random variables themselves. This means they have a standard error and can be set within the bounds of a confidence interval. Third, comparing surveys of different sample sizes is only allowed if the sampling variation in the response propensities ρ_i has been taken into account. In small samples, this sample variation is considerable and has the tendency the increase the variance of the response propensities such that the estimated $\hat{R}(\hat{\rho})$ is biased downwards. RISQ-deliverable 2 provides the procedure to separate the sampling variance from the *real* or population variation, so that the comparison of differently sized samples is legitimate.

3 Monitoring as a process-oriented approach

There is a general belief that process quality is the key to the final product quality. This belief is perhaps strongly institutionalized in many manufacturing disciplines; survey researchers however have only recently acknowledged the usefulness of process data to improve survey quality. Process quality is different from output quality. Under the output approach, one typically assesses the differences between respondents and nonrespondents (after refusal conversions, etc.) or the difference between respondents and population with regard to some variable(s) (see e.g. Groves and Couper, 1998). On the other hand, the focal point of the process approach is the construction or realization of the sample, including the preparation of the sample frame, the sampling procedure and the fieldwork. The process approach focuses on the selection of sample units, the timing and sequence of contacts, the efforts made to make someone participate, etc. Moreover, the process precedes the output and therefore the process quality determines the output quality or 'If the process of gathering data is good, there is no need to worry about the quality of the final product' (Lyberg and Biemer, 2008). Also, whereas the output assessment is usually respondent-oriented, the process approach focuses on a broader range of survey agents, such as the interviewers or the fieldwork management, as these have an important impact on the selection and treatment of sample units. Furthermore, the assessment of the process quality requires more data and documentation about the selection, the timing, the contact attempts, etc. That is why the availability of good paradata is critical for fieldwork monitoring and fieldwork improvement.

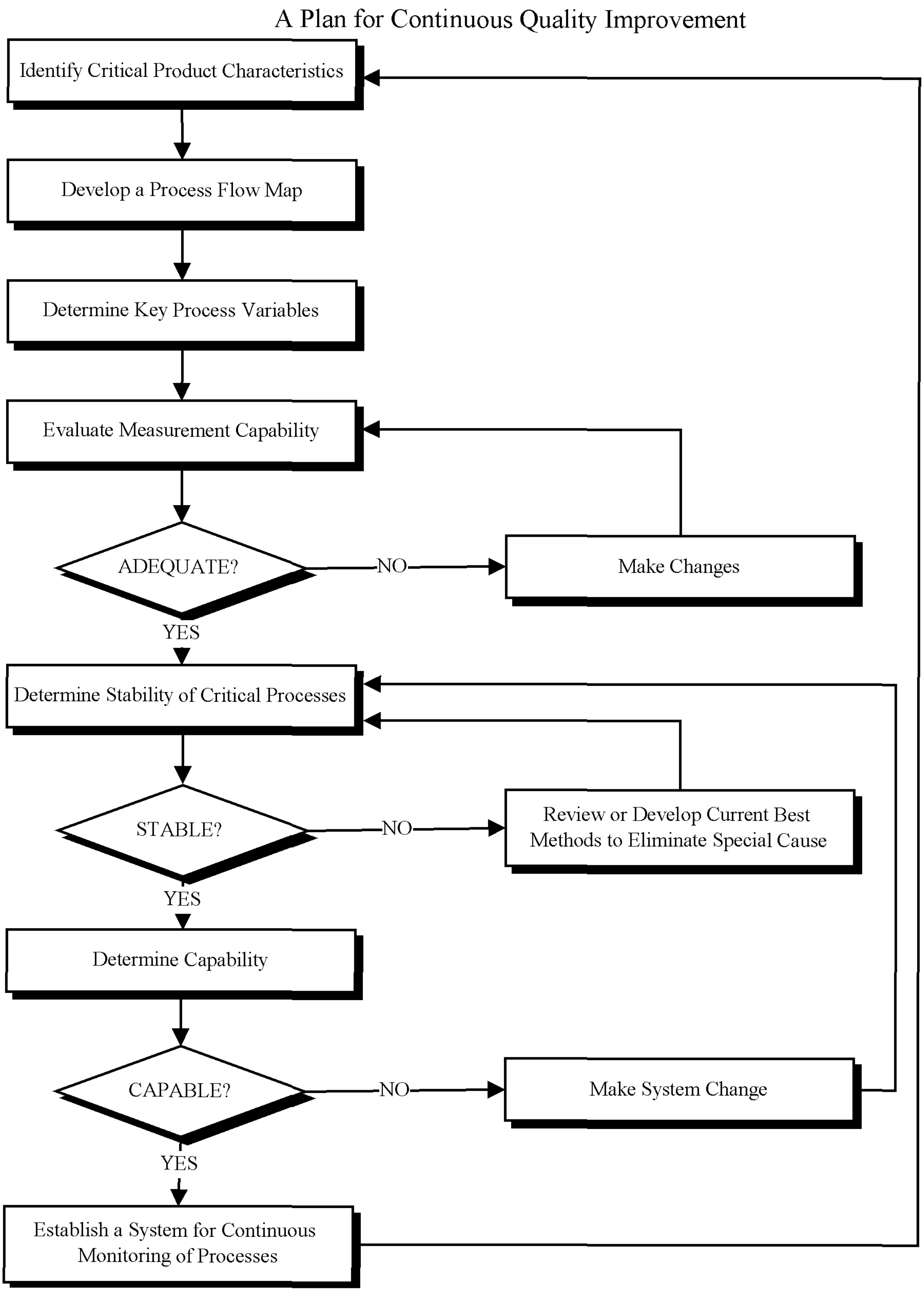
Process monitoring is relatively new in survey practice. The term paradata (and the ambition to use it) for example first appeared in 1998 (Couper). Dippo (1997) recognizes that the integration of continuous quality improvement in statistical service requires a wider approach than in manufacturing. Their processes to be addressed are typically not physical products, but human or machine action, decisions and the path decisions take. Aitken et. al. (2003) also argue that literature on quality improvement and process monitoring is relatively scarce in survey research, notwithstanding the unmistakable benefits suggested by their examples.

A general theoretical framework to inspire fieldwork monitoring with respect to representativeness is adopted from Morganstein and Marker (1997) and is presented in Figure 1. Their framework builds on the achievements of the TQM-paradigm.

The first step is to identify the critical quality characteristics. It is rather obvious that the RISQ-project focusses on representativity issues and defines this characteristic as the variance of response propensities. Note that the critical characteristics not only address to the final response representativity, but also to specific parts of the process that involves the (non)selection of sample units (e.g. noncontacts, refusals, other nonresponse).

The next activity in improving statistical products is to develop a flow chart map, yielding a better understanding of the related (sub)processes. Figure 3 on p.15 shows such an example of the process flow related to the sample construction activities of the Belgian ESS, 3rd round. Three components should be in the flow chart. First, the sequence of processes is delineated, indicating all decisive points. Second, the owners (agents or stakeholders) are identified. Interviewers, (non)respondents and fieldwork management take the most prominent positions in the flow. Third, the key process variables are listed. These factors can vary with every repetition of the process and affect critical product characteristics. In RISQ-terms, the key process variables represent the quality of the sub-processes such as the assignment of interviewers, hold periods between contact attempts, etc. Later on, they will be referred to as treatment variables, since they can be considered as controllable variables and serve as an input to improve the survey quality.

Figure 1: A plan for continuous quality improvement



The evaluation of the measurement capability entails the quality of the process data. It refers to a wide range of information about the process. Morganstein and Marker (1997) argue that in their experience, the measurement error with respect to process data is often one of the least appreciated aspects of quality improvement procedures: ‘Researchers often select a process because it is easy to measure, rather than choosing a more important but harder-to-measure process.’ We tend to endorse this claim as many fieldwork organization or interviewers do not have a long tradition in documenting or registering their contact activities, as well as with respect to the organization or archiving of such data. Unfortunately, the auditing of the quality of paradata is beyond the scope of the RISQ-project.

The next step involves the actual monitoring of the fieldwork. Here, the sources of process and output variations are identified. The R-indicator may show sudden jumps or a gradual deterioration of representativeness. This may be accomplished by using control charts and other statistical tools or methodologies. The empirical section will deal with these tools.

The following steps in the quality improvement plan deal with methods and procedure to further develop and institutionalize the monitoring activities using the R-indicator. We will restrict ourselves however to the earlier steps as mentioned above.

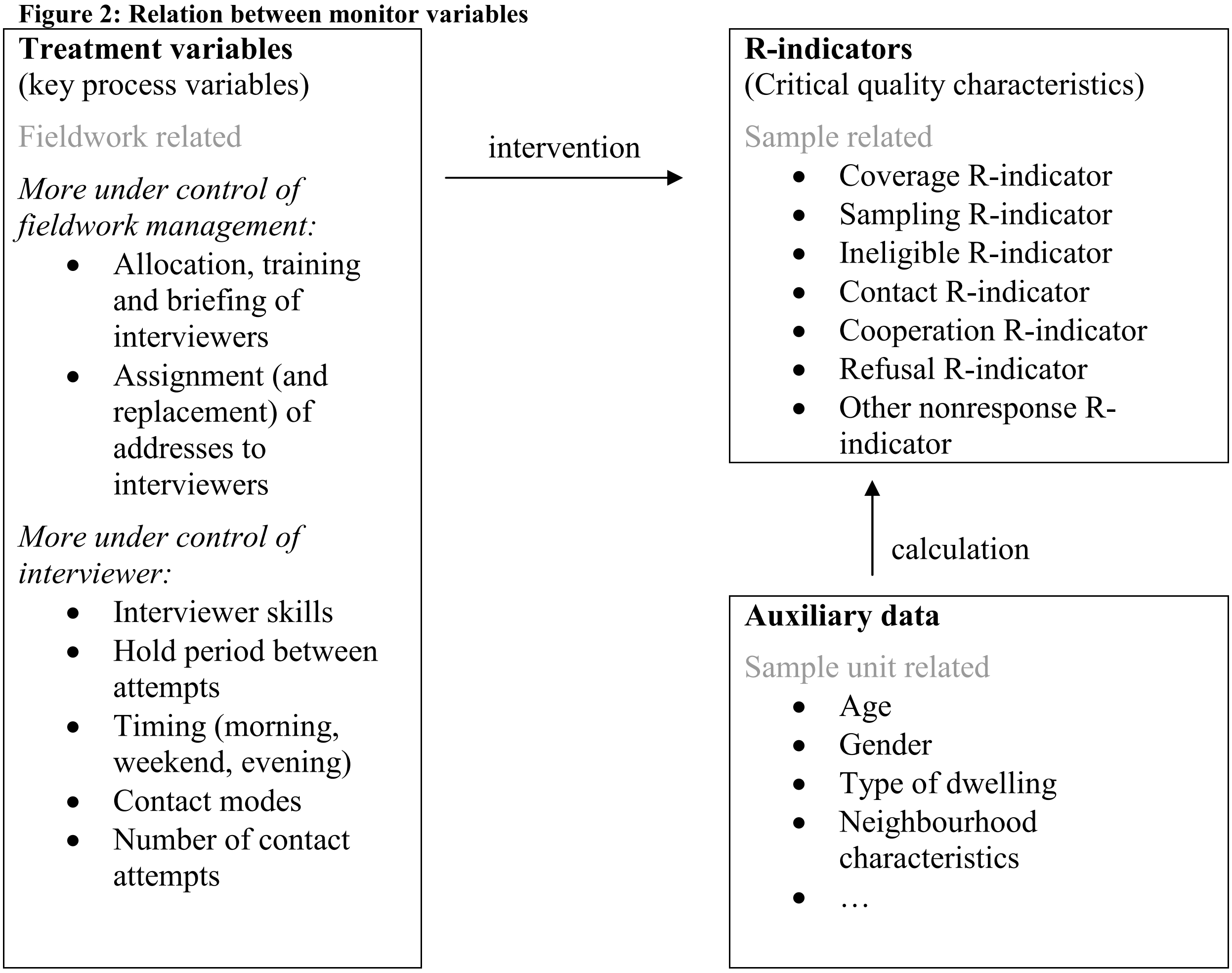
4 Analytical elements

As Montgomery (2005) posits, product quality improvement basically tries to repulse product variability and this is exactly what R-indicators seek to trace in the composition of survey samples. From a practical point of view, the monitoring observer or fieldwork manager may be interested in the current quality of a survey and likes to be informed about potential levers or forces to improve representativeness, that is, to reduce the variance of the estimated response propensities. Hence, the above-mentioned theoretical framework needs to be further elaborated and translated into useful survey terminology in order to tackle the issues regarding fieldwork monitoring. In this section it will be argued that there is a need to distinguish between treatment variables (key process variables) and critical quality characteristics (R-indicators). So-called auxiliary variables can be put on the side of the critical quality characteristics, as they permit to measure the variability of the response propensities, necessary to obtain R-indicators. The interrelations between these analytical elements are depicted in Figure 2.

First we will focus on treatment variables, including the agents who control these variables. Here we assume, in agreement with Groves and Couper (1998), that survey participation is subject to two main bundles of factors: influences that are under the control of the researchers and influences that go beyond their control. The surveys’ design, including the topic, the mode of administration and selection of units, the selection and training of interviewers are believed to be, to a reasonable extent, under the governance of the researcher. The survey climate on the other hand, as well as the respondents’ background, psychological predispositions or household structure cannot be controlled by the researcher. Monitoring routines typically address the impact of the factors that are controllable in order to evaluate the sample composition and this controllability is closely related to the concept of survey agency.

The various survey agents, also called process owners in the terminology of Morganstein and Marker (1997), play specific roles in the production process and have ditto contributions or competences regarding representativeness. One may distinguish between (non)respondents, interviewers, fieldwork organization (occasionally a subcontractor), the survey sponsor, government, etc. In this perspective, controllability of the survey variables is relative to the role of the agent in question. The survey sponsor may have little control over the interviewers’ behaviour in the field, but has a considerable decisive input concerning the allocation of the survey budget. Interviewers have more control over sample cases than the fieldwork managers. Governments may have no control over the cooperativeness of a particular respondent but are empowered to decide upon privacy legislation or the use of administrative data. For this application, we will take the perspective of the fieldwork management. Both the management and the interviewers dispose of the 'key process variables' that potentially alter the quality of the survey during the fieldwork process. One may think in this respect of the number of contact attempts, contact modes, interviewer skills, advance letters, hold period between two contact attempts, etc. Key process variables (can also be interpreted as treatment variables) may influence the response success and the related variance among the response propensities.

Figure 2: Relation between monitor variables



We propose to distinguish between two kinds of fieldwork variables. Of the first kind, it is a priori clear what the effect is on the response propensities, e.g. the more contact attempts, the higher the contact success. The second kind of process variables may affect the response success, but the effects may depend on the sample unit, e.g. a weekend or evening attempt may be more successful among working people, whereas attempts during daytime are relatively more fruitful among students or unemployed people. This distinction is relevant in that it informs the fieldwork manager how to intervene in the process. Another relevant distinction with regard to treatment variables pertains to the selection for further treatment on the one hand and the implementation of a particular treatment once the sample unit has been selected. This distinction will be elaborated further on.

The critical quality characteristics are the focal point of the RISQ-project. In fact, R-indicators are the critical quality characteristics. They indicate the extent to which the quality of the sample is contaminated by response propensity variation. Ideally, this variation should be very low. It should be mentioned that the R-indicator can be evaluated at the different stages in the realization of the final sample, resulting in a multi-layered concept. Each of these representativity measures can be estimated with respect to the entire sample or with respect to one of the previous steps in the survey process: e.g. measuring response representativity amongst the successfully contacted subsample. In Table 1, an illustrative overview is given concerning all possible (or relevant) types of representativity related to the process of sample construction.

Table 1: Types of representativity

	Type of representativity							
(subsample)	Sample frame	Gross sample	Attempt	Ineligibility / eligibility	Contact / noncontact	Other nonresponse	Refusals	Cooperation
Population	X (1)	X	X	X	X	X	X	X
Sample frame		X (2)	X	X	X	X	X	X
Gross sample			X	X	X	X	X	X
Attempt				X	X	X	X	X
Eligibles					X (3)	X	X (4)	X
Contact				X		X	X	X
Refusals + cooperation							X	X (5)

The columns in the table represents the steps or moments in the process at which the representativity indicators are measured with respect to the (sub)sample, indicated by the rows. E.g. the column labelled as 'contact / noncontact' measures the representativity of the successfully contacted in the population, the sample frame, the selected sample or the sample found eligible. We will only select a few of these combinations, of which the relevance is inspired by the flow chart. Note that some of these combinations are well-known sources of survey error. Coverage error (1) indicates the dissimilarity between sample frame and population, sampling error (2) refers to difference between gross sample frame. We also mention noncontact error (3), refusal error (4), noncooperation error (5). The column 'Attempt' reflect the possibility that some sample members may not be attempted.

Conceptually, we represent the relation between the diverse types of variables as shown in Figure 2. This figure integrates the different kinds of variables (critical quality characteristics and key process variables) and defines their relation. Critical quality characteristics are so-called output quality indicators and are determined by the process quality. The auxiliary variables do not seem to play an important role in this respect, although they determine the measurability of the quality indicators $R(\rho)$.

Deliverable 3 already distinguishes between the different kinds of auxiliary variables (Schouten et. al., 2009). Auxiliary variables are observed for all sample units outside the survey questionnaire. Moreover, they exclusively refer to characteristics of the sample units. We arrive at the following taxonomy (slightly modified with regard to deliverable 3):

1. Auxiliary variables that become available from a source other than the survey or the survey data collection, and that are constant during the fieldwork (e.g. register data).
2. Auxiliary variables that are collected by the interviewer or survey organisation during the fieldwork but that are constant during the fieldwork (e.g. neighbourhood characteristics as observed by the interviewer).

3. Auxiliary variables collected by the interviewer or survey organisation that are allowed to change during the fieldwork but are independent of the interaction between the survey organisation and the respondent (e.g. mood of respondents).

A final remark with regard to the building stones of Figure 2. is the fact that they generally (but not exclusively) coincide with different levels of observation. On the aggregate or sample level, the R-indicators can be found. These indicators always summarize the propensities of the sample members. On the intermediary level, the auxiliary variables can be situated, as they refer to the individual sample units. The third class of auxiliary variables in the above-mentioned taxonomy is however an exception. On the lowest level, the attempts or possible interactions between interviewers and sample units can be found. Each contact attempt can be described in terms of an additional fieldwork effort and each effort can be specified in terms of timing, modes, the skills of the interviewer conducting the attempt, etc. In the following section, this levelled approach will be further elaborated, as it is relevant for the conceptualisation and estimation of fieldwork-conditional response propensities.

5 Modelling response propensities

Considering Figure 3, it is obvious that many decision have already been taken before a sample unit is successfully contacted and/or participates. All these prior decisions are made by survey agents other than the sample units and may lead to a situation where the actual contexts to which the sample unit is presented to are rather diverse. Sample units are assigned to different interviewers with different skills or routines, some clusters of addresses are only assigned during the course of the fieldwork, some cases are attempted on a weekday, others during the weekend, etc. As already specified in deliverable 2, it is hard to argue that a 'true' response propensity ρ_i is uniquely defined, particularly because the survey process presents stimuli or incentives to the sample cases that encourage/discourage them to respond. In short, it seems that a proportion of the variability of responses depends on fieldwork decisions, particularly when fieldwork variables coincide with auxiliary variables. Such correlations can be advantageous in the case where hard-to-contacts or reluctant sample units are assigned to favourable fieldwork contexts. However, it may seem more adequate to expect that most fieldwork organisations are still response rate-oriented, probably prioritising the more promising cases. Therefore, we propose to simulate a situation in which all sample cases have been treated equally and compare this simulated R-indicator to the observed one. An equal treatment R-indicator can serve as a normative reference. Furthermore, this simulation generates response propensities that are *process free* and this improves the monitoring capabilities as they locate strata of profiles that are *really* hard to contact or hard to persuade to participate. Moreover, comparing raw propensities and context free propensities allows diagnosing which profiles have been prioritized by the fieldwork management and/or interviewers, which may be guidance to corrective measures.

We propose to design three kinds of response propensities. Raw propensities ρ_i are determined only by considering auxiliary information and by ignoring any kind of treatment information. The second kind of response propensities $\rho_i^{'}$ assume that in case of a unfavourable outcome (e.g. noncontact or refusal) the probability of being reselected is equal amongst all profiles. Weighting procedures will be used to obtain such propensities. The third kind of propensities $\rho_i^{''}$ do not only assume equal selection probabilities, but also equal treatments once the units have been reselected.

For the modelling of the propensities, we choose to use the discrete-time hazard model as discussed by Singer and Willett (2003, pp. 357-467). This model is a more specified or elaborate application of the logistic regression. Typically, the model involves a variable that reflects an indication of time in a discrete manner; in our case this variable would be the successive number of contact attempts. Table 2 portrays a typical dataset that can be dealt with by the proposed model.

Table 2: Example data set

$person_i$	$visit_j$	gender	mode	response	(re)visit	v_1	v_2	v_3	v_4	w_{ij}	w_j	ρ_{ij}	ρ'_{ij}	ρ''_{ij}
1	1	M	F2F	0	1	1	0	0	0	w_{11}	w_1	ρ_{11}	ρ'_{11}	ρ''_{11}
1	2	M	TEL	1	1	0	1	0	0	w_{21}	w_2	ρ_{12}	ρ'_{12}	ρ''_{12}
1	3	M				0	0	1	0	w_{31}	w_3	ρ_{13}	ρ'_{13}	ρ''_{13}
1	4	M				0	0	0	1	w_{41}	w_4	ρ_{14}	ρ'_{14}	ρ''_{14}
2	1	F	F2F	1	1	1	0	0	0	w_{12}	w_1	ρ_{21}	ρ'_{21}	ρ''_{21}
2	2	F				0	1	0	0	w_{22}	w_2	ρ_{22}	ρ'_{22}	ρ''_{22}
2	3	F				0	0	1	0	w_{32}	w_3	ρ_{23}	ρ'_{23}	ρ''_{23}
2	4	F				0	0	0	1	w_{42}	w_4	ρ_{24}	ρ'_{24}	ρ''_{24}
3	1	M	F2F	0	1	1	0	0	0	w_{11}	w_1	ρ_{31}	ρ'_{31}	ρ''_{31}
3	2	M	F2F	0	1	0	1	0	0	w_{21}	w_2	ρ_{32}	ρ'_{32}	ρ''_{32}
3	3	M			0	0	0	1	0	w_{31}	w_3	ρ_{33}	ρ'_{33}	ρ''_{33}
3	4	M			0	0	0	0	1	w_{41}	w_4	ρ_{34}	ρ'_{34}	ρ''_{34}
4	1	F	F2F	0	1	1	0	0	0	w_{12}	w_1	ρ_{41}	ρ'_{41}	ρ''_{41}
4	2	F	TEL	1	1	0	1	0	0	w_{22}	w_2	ρ_{42}	ρ'_{42}	ρ''_{42}
4	3	F				0	0	1	0	w_{32}	w_3	ρ_{43}	ρ'_{43}	ρ''_{43}
4	4	F				0	0	0	1	w_{42}	w_4	ρ_{44}	ρ'_{44}	ρ''_{44}
5	1	M	F2F	0	1	1	0	0	0	w_{11}	w_1	ρ_{51}	ρ'_{51}	ρ''_{51}
5	2	M	F2F	0	1	0	1	0	0	w_{21}	w_2	ρ_{52}	ρ'_{52}	ρ''_{52}
5	3	M	TEL	0	1	0	0	1	0	w_{31}	w_3	ρ_{53}	ρ'_{53}	ρ''_{53}
5	4	M	F2F	1	1	0	0	0	1	w_{41}	w_4	ρ_{54}	ρ'_{54}	ρ''_{54}

The table illustrates the levelled structure of the data as the i-index refers the to intermediate level of individuals and the j-index pertains the micro-level of visits (interviewer-individual interactions). If one would not be interested in the monitoring of the R-indicator while the survey is still in motion, but rather seeks to focus on the final representativeness of the sample, the dataset counts as many data lines as there are (non)respondents. For each element a specific propensity is obtained, using a set of auxiliary (sample unit related) variable(s), with which an R-indicator is calculated. Conversely, for monitoring routines it is more interesting to disaggregate the data from sample unit level to the level of visits. Now, for each $person_i$-$visit_j$ data line a specific response propensity can by determined, using the following model:

$$g(h(\rho_{ij})) = [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}]$$

The V-variables represent the discrete time events or visits and are dummy-coded as indicated by the example data set. The A-variables are the auxiliary information and are assumed to be constant at any time j (e.g. gender). The link function g may be the logit, probit or identity link, as discussed in previous RISQ deliverables. The obtained hazard propensities $h(\rho_{ij})$ can be transformed to survival propensities using $s(\rho_{ij}) = s(\rho_{i(j-1)})[1-h(\rho_{ij})]$. This transformation is relevant since one is not interested in the response propensities *at* a certain time j, but rather *until* a certain visit j. Also note that censoring can occur during the data collection process: some individuals may be considered as final nonrespondents after a few attempts (e.g. person 3), while others receive more survey efforts. This means that the inclusion (or exclusion) probability π_{ij} during the fieldwork process should be taken into account. π_{ij}^{-1} can be used as a weight to correct the censoring. In the table above the w_{ij}'s represent these weights conditional on the available auxiliary variables, w_j's can also be obtained, irrespective of auxiliary information. The censoring adjusted survival propensities become $s(\rho_{ij}) = s(\rho_{i(j-1)})*[1-h(\rho_{ij})w_{ij}]$. The model as shown above and combined with w_{ij}'s to obtain the survival propensities provides raw response propensities ρ_i in the sense that they disregard possible differences with respect to selection or treatments. Note that the w_{ij}'s can be estimated using a model for binary response 'selected for renewed attempts' versus 'not selected for renewed attempt'. The weights w_{ij}'s are conditional on auxiliary information. In case where the weights w_j's are unconditional, one assumes equal selection probabilities for additional contact attempts. The resulting propensities $\rho_i^{'}$ and associated R-indicator can then be compared to the raw version. If the situation improves $(R'(\rho') > R(\rho))$, fieldwork selections may have prioritised the convenient cases and bypassed the less attractive sample units.

Occasionally, T-variables (treatments) can be introduced on the visit-level in order to control for systematic assignments of particular profiles to particular fieldwork treatments (e.g. more promising sample cases may have been systematically less approached by telephone). In this case, the treatment information is introduced in the model, but is not used to determine the propensities $\rho_i^{''}$. Only the parameter estimates related to the V-variables and A-variables are used for this purpose. It is advisable to first mean-centre the treatment variables. Doing so, the hypothetical situation is created that all sample members have been assigned to similar treatments, without changing the marginal distributions of the treatments; that is, the fieldwork offers the same amount and nature of the efforts on the aggregate level, only the efforts are reallocated such that equal treatment conditions apply. Dramatic shifts in mean response propensities (=response rate) are therefore not expected. Minor changes may occur. A simple example of how T-variables may affect R-indicators can be found in Appendix B.

The Table 3 gives an overview of the three types of response propensities and how to obtain them.

Four assumptions about these models should be discussed. First, the models as presented in Table 3 assume that the effects of the auxiliary variables are independent of the number of attempts. This assumption may be relaxed by allowing interactions between the V-variables and the A-variables, such that at every visit j all auxiliary variables have j-specific estimates. This provides more flexibility to the propensities and also accommodates the possibility that some profiles are more (or less) sensitive to additional visits such in the case of noncontact or refusal conversion. In the empirical section, such interaction will be applied.

Another assumption relates to the stability of the treatment effects over the attempts. As stipulated in the final hazard model of the above table, each treatment effect is estimated only once and this assumes a constant effect, irrespective of the number of visits. This constraint may also be relaxed by allowing interaction terms between the V-variables and the T-variables. It may be possible that e.g. the effects of interviewer skills mitigate as the number of visits increase. In the empirical section, these interaction will be applied.

Table 3: Procedures to estimate different types of response propensities

Raw propensities ρ_{ij}	Discrete-Time hazard model: $g(h(\rho_{ij})) = [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}]$ Estimation of the hazard propensities $h(\rho_{ij})$: $h(\rho_{ij}) = g^{-1}([\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}])$ Survival (cumulative) propensities: $s(\rho_{ij}) = s(\rho_{i(j-1)})*[1-h(\rho_{ij})w_{ij}]$ where $w_{ij} = \pi_{renewed\ attempt} = g^{-1}([\alpha_{w1} V_{1ij} + \alpha_{w2} V_{2ij} + \ldots + \alpha_{wJ} V_{Jij}] + [\beta_{w1} A_{1i} + \beta_{w2} A_{2i} + \ldots + \beta_{wP} A_{Pi}])$
Equal selection propensities ρ'_{ij}	Discrete-Time hazard model: $g(h(\rho_{ij})) = [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}]$ Estimation of the hazard propensities $h(\rho_{ij})$: $h(\rho_{ij}) = g^{-1}([\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}])$ Survival (cumulative) propensities: $s(\rho_{ij}) = s(\rho_{i(j-1)})*[1-h(\rho_{ij})w_j]$ where $w_j = \pi_{renewed\ attempt} = g^{-1}([\alpha_{w1} V_{1ij} + \alpha_{w2} V_{2ij} + \ldots + \alpha_{wJ} V_{Jij}])$
Equal selection and equal treatment propensities ρ''_{ij}	Discrete-Time hazard model: $g(h(\rho_{ij})) = [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}] + [\beta_{T1} T_{1i} + \beta_{T2} T_{2i} + \ldots + \beta_{TP} T_{Pi}]$ Estimation of the hazard propensities $h(\rho_{ij})$: $h(\rho_{ij}) = g^{-1}([\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}])$ Survival (cumulative) propensities: $s(\rho_{ij}) = s(\rho_{i(j-1)})*[1-h(\rho_{ij})w_j]$ where $w_j = \pi_{renewed\ attempt} = g^{-1}([\alpha_{w1} V_{1ij} + \alpha_{w2} V_{2ij} + \ldots + \alpha_{wJ} V_{Jij}])$

Third, it is presupposed that the treatment variables have similar effects among all (non)respondent profiles. However, one may expect that e.g. morning attempts are more successful among older people or that apartment dwelling are better approached by telephone.

Introducing interactions between A-variables and T-variables facilitates the option to obtain different treatment effect among the different groups in the sample. This assumption has been tested for the upcoming analyses. Such interactions were not convincingly observed, so that the interactions have been omitted.

Fourth, we assume that the auxiliary information or (non)respondent characteristics do not change during the course of the fieldwork. As the fieldwork process does usually not take more than a few months, it seems reasonable the consider age, dwelling or area information as relatively stable. Conversely, more volatile data such as individuals' mood swings or other attitudinal variable are much harder to incorporate in the modelling strategy. In the example further on, we do not dispose of such volatile auxiliary variables so that this problem is simply put aside. In the case where such data is available, different models should be applied. We will however not further develop this particular idea.

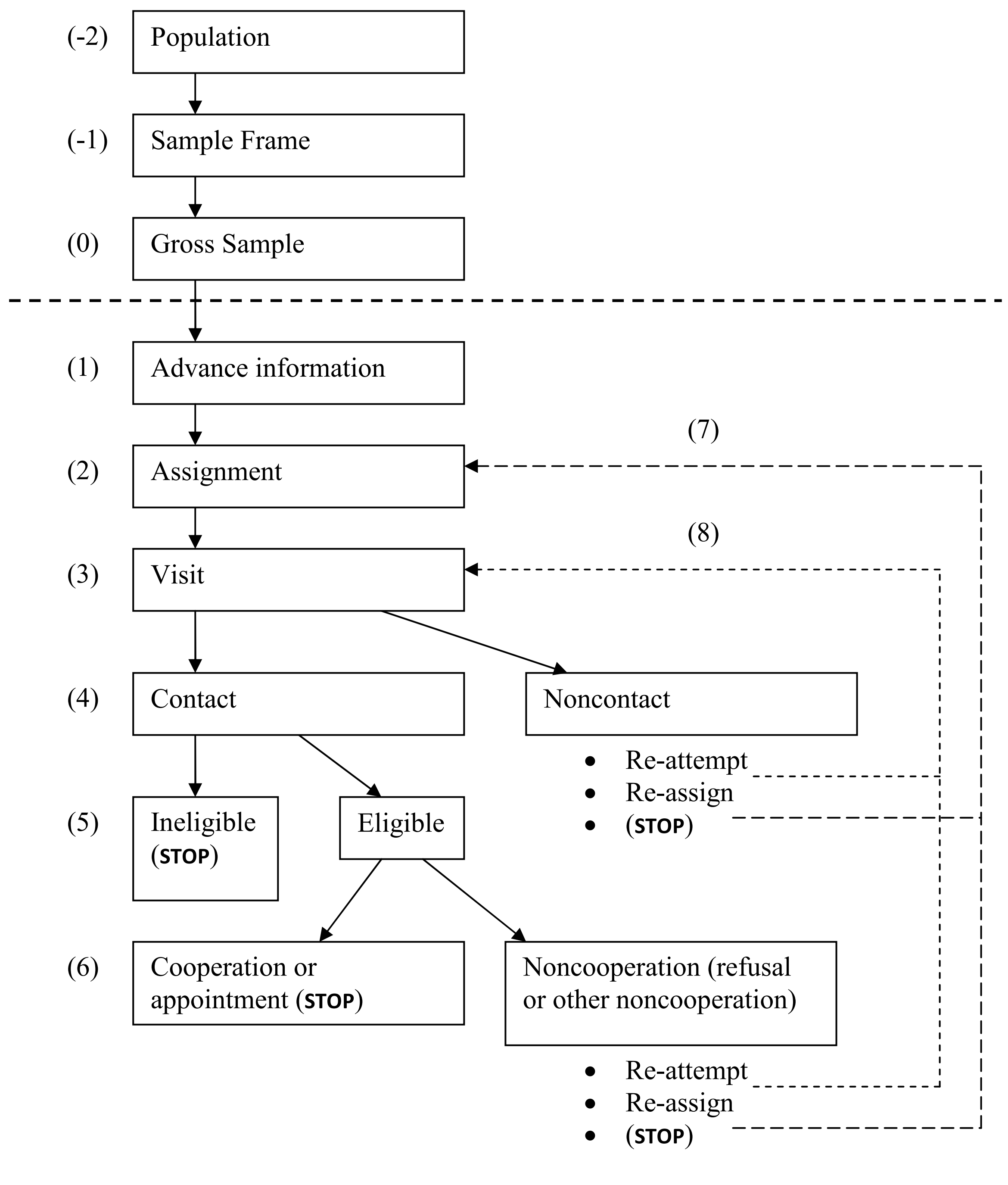
6 Empirical example: the European Social Survey – Belgium, 3rd round

To illustrate the specified monitoring techniques, the case of the Belgian part of the third round of the European Social Survey will be dealt with. First, some general background information will be provided. Then the auxiliary and treatment variables are presented. The results of the monitoring exercise will be provided following the logical sequence of the contact process. First, the sample cases need to be contacted. Once contact is established, the eligibility of the case is evaluated, after which the interviewer requests survey participation. In the case of a noncontact, refusal or other types of noncooperation (e.g. illness, language barrier) the subprocesses can be iterated, by the same interviewer (re-attempt) or by another interviewer (re-assign). As a guide for this contact process, a possible representation of the operational flow is portrayed in the following flow chart. We will predominantly focus on the process below the dashed line in Figure 3. The above-the-line activities can also be monitored, but since no frame data is available with respect to this survey, nor do we dispose of a large fund of population-based auxiliary information, this part of the monitoring will be skipped. Moreover, estimation procedures based on population totals are not (yet) available.

6.1 European Social Survey – Belgium, 3rd round: background information and sampling design

The ESS is a biennial and multi-country survey covering over 20 nations. The first round was fielded in 2002/2003, the second in 2004/2005, the third in 2006/2007. The main goal of the ESS is to screen and explain Europe' s changing institutions, its political and economic structures, the populations' beliefs, attitudes and behaviour. It is funded via the European Commission's 6th Framework Programme, the European Science Foundation, and national funding bodies in each country. It involves strict random probability sampling, a minimum target response rate of 70% and rigorous translation protocols. The noncontact rate should not exceed 3%. The hour-long face-to-face interview includes questions on a variety of core topics repeated from previous rounds of the survey and also two modules developed for Round Three covering personal and social well-being and the organisation of the life course in Europe. Other topics include: media, social trust, political interest and participation, socio-political orientations, social exclusion, national, ethnic and religious allegiances, timing of key life events and the life course, personal and social well-being and satisfaction with work and life, demographics and socio economics. All participants have to be 15 years or older and resident within private households, regardless of their nationality, citizenship, language or legal status, in the country of residence.

Figure 3: Flow diagram of the contact process – ESS3, Belgium



Process owners
(-1) government, statistical agency, private company, …
(0) government, statistical agency, private company, fieldwork management, …
(1) fieldwork management, interviewer
(2) fieldwork management
(3) interviewer
(4) interviewer, (non)respondent, family member, neighbour, …
(5) interviewer, (non)respondent, family member, neighbour, …
(6) interviewer, (non)respondent, family member, neighbour, …
(7) fieldwork management
(8) fieldwork management, interviewer

The Belgian 3rd survey of ESS was fielded from 23.10.06 until 19.02.07, using TNS Dimarso as a subcontractor. All 118 experienced interviewers worked on a free-lance basis and were personally briefed about the ESS for ½ day or less. They are paid per completed interview and all received some refusal conversion training. In Belgium, the basis is the commercial database of 'Orgassim'. Using the National Rgister, Orgassim has developed a database with 'Statistics of inhabitants per building'. With this database it is possible to make an individual database with age, gender and address for each person. The names of the persons are not available in this database. Then, the individual database is linked with another commercial database and 'enriched' with names (65% matches). A person is identified by his or her name or the combination of gender and age. The database is updated annually. Unfortunately, the sample frame data are not at disposal so that its representativity check cannot be carried out. Next, the gross sample is drawn from the frame. The Belgian sample is a result of a stratified two stage probability sampling design. The ten provinces and Brussels are used for regional stratification. At stage 1 the primary sampling units (PSU's) are 'virtual' clusters located in municipalities, which means that the clusters within the municipalities are not further defined regionally. The number of clusters for each province is proportional to the size of the population in each province. For that a list of municipalities with a population distribution (+15 years) for each province is used. The number of clusters in a municipality is proportional to the size of its population. The total number of clusters equals 338. At stage 2 in each of the 338 clusters, 9 persons are selected for the gross sample by simple random sampling, implying that the number of contacted persons in each municipality equals the number of clusters in the municipality times nine.

6.2 Auxiliary variables

From the register, some information about the sample cases is available such as **age** and **gender**. We divided the age variable into four classes: 14-20, 21-40, 41-60 and 60+. This classification is used since these categories coincide with the age classes on the contact form (in some cases, interviewers needed to estimate the age of the (non)respondent). From the identification numbers, it can be derived which Belgian province the sample case lives in. The eleven provinces are than reduced to the three constitutional **regions** (Flanders, Wallonia and Brussels). The register data (though not publicly available) also contains the postal codes of the sample units. It should be mentioned that this information provides access to a relatively large fund of external data. Postal codes can be linked to administrative or census data such as the **population density** or the **percentage of non-Belgians** living of the municipality. We also managed to link the **average income** of the municipalities to the postal codes. These latter three variables are categorized as follows:

Table 4: Categorization of three of the auxiliary variables

Population density in municipality		Percentage of non-Belgians in municipality		Average annual per capita income in municipality	
	inh./km²		%		€
1	≤200	1	<2	1	<12.000
2	201-400	2	2-5	2	12.000-14.000
3	401-700	3	5-15	3	14.000-16.000
4	701-2500	4	>15	4	>16.000
5	>2501				

Interviewer-observed data is also available in the ESS. Whether or not the sample units lives in an **apartment** is dichotomized in an auxiliary variable. Furthermore, a composite index is constructed reflecting the **quality of the neighbourhood**. This latter variable indicates to what extent the interviewer felt the neighbourhood shows traces of vandalism, graffiti, litter, rubbish, deliberate damage or in what state the building and dwellings in the area are. Unfortunately, this information about the apartments and the quality of the neighbourhood is not available for the entire gross sample (about 5% missing). Therefore, it was decided to impute these values, conditional on the other available auxiliary variables. The variable reflecting the quality of the neighbourhood counts three categories (‘poor neighbourhood conditions’, 22%; ‘good neighbourhood conditions’, 31%; ‘excellent neighbourhood conditions’, 47%).

6.3 Treatment variables

The treatment variables (also called key process variables) are summarized in the following overview and discussed in more detail afterwards. Note that the fieldwork supervisor or monitor strongly depends on the availability of process data. In this exercise, we need to restrict ourselves to the data that is collected by means of the standard ESS contact forms. Information about the weather, the age and gender or other information about the interviewers are not at our disposal. Also note that of some decisions, the contact form does not indicate which specific survey agent is responsible, e.g. some cases have been attempted more than the mandatory four contact attempts. In case there was no replacement of interviewers, it is not clear who decided to carry out additional contact attempts (interviewer of survey management). We hope however that thanks to this RISQ-project, a growing awareness of the usefulness of paradata emerges so that a meaningful contribution can be made to better organize process data or to do recommendations for variable selection or their registration.

Note that some of these treatment variables may be slightly modified according to the specific type of representativity, e.g. when assessing contact representativity the contact skills of the interviewer will be used as a treatment variable, whereas in the case where cooperation representativity (cooperation versus refusal) the persuasive skills of the interviewers will be deployed as a specific treatment variable. Also the hold period (elapsed time for renewed attempt after unsuccessful tentative) is adapted to the specific context of the subprocess to be evaluated: with respect to contact representativeness, the hold period is measured between any two subsequent contact attempts, whereas with respect to participation representativeness (refusal versus cooperation) the hold period spans between two visits that decide upon refusal or cooperation (noncontact visits in-between are ignored).

Advance information	• Did the (non)respondent receive / read the information?
Assignment	• Assignment of interviewer to individuals / skills of the interviewers • Assignment of sample units during the fieldwork
Attempt	• Contact mode (telephone versus personal) • Timing (weekend, evening, morning, afternoon)
Re-assignment	• Re-assignment to other interviewer • Skills of the new interviewer
Re-attempt	• Total number of attempts • Elapsed time (hold period) after unsuccessful tentative

(1) Advance information

In ESS3-Belgium, an advance letter and a brochure were sent to each prospect respondent. It is however not clear whether all units received or even read this information. As it is believed that advance information improves the responsiveness, selective (non-)receptivity or (un)willingness to read may induce systematic (non)contactability or reluctance to participate. Nevertheless, there is no information available to find out whether the advance information arrived at the right place and was actually read. In other words, this part of the process lacks measurement capability and can consequently not be used as a treatment variable.

(2) Assignment

(2a) Not all sample clusters were assigned at the same moment. About two thirds of the gross sample were distributed at the beginning of the fieldwork, the remaining cases were assigned after two months. This may affect the allocation of contact efforts, e.g. shorter hold periods or less efforts until sample cases are considered as final noncontacts.

(2a) Furthermore, as a result of two-stage probability sampling, clusters of 9 sample units are assigned to interviewers who live in the neighbourhood of the clusters. In order to reduce travel costs, all members of the same cluster live in the same municipality. Given the variability among interviewers with regard to e.g. contact skills or persuasiveness and given the fact of local implementation of interviewers, there is a potential risk that not all sample units are treated equally, suggesting that the assignment of sample units to interviewers may affect representativity issues. In order to assess this, we first need to develop an indicator reflecting the skills of all interviewers. One possibility would be to use historical interviewer performance data. Unfortunately, these data are not available in ESS3 – Belgium so that we choose to use the ESS3 data themselves. The following model creates a variable reflecting the contact skills of the interviewer. Similar models are used to obtain interviewer skills that relate to persuasiveness of avoidance of other nonresponse.

$$\ln\left(\frac{P(contact='yes')}{P(contact='no')}\right)=\alpha_0+\beta_1x_1+\beta_2x_2+\ldots+\beta_px_p+\begin{bmatrix}\alpha_1\\ \alpha_2\\ \vdots\\ \alpha_k\end{bmatrix}$$

Parameter α_0 reflects the overall contact rate among all interviewers and the β's represent the fixed effects of sample unit information (age and gender) or area information (population density, proportion of immigrants, income). Also a fixed effect has been introduced with respect to the wave to which the sample case was assigned. The vector at the end of the expression is most important in this context. For each interviewer a separate α_{int} (with index from 1 to k) is generated reflecting how much the particular interviewer deviates from the overall contact rate, controlling for relevant fixed effects such as age, gender or neighbourhood characteristics. We will use the vector containing the α_{int} information as a treatment variable. A problem, however, in this context is that the interviewers' performances serve both to estimate their skills as well as to estimate the selectiveness with regard to sample case assignment. Both estimations may of course not be completely independent of one another. A possible solution is to split the sample into two equivalent parts (within each interviewer) and use one part to estimate the interviewer skills, leaving the other part to estimate the selectiveness. We have also assumed that the fixed effects are not interviewer specific. We consequently ignore the possibility that some interviewers perform better in certain segments of the sample (e.g. exclusively women or elder people).

(3) Contact mode

Here, we will distinguish between telephone calls and personal visits. Of the four mandatory contact attempts, only the first one had to be a personal visit. Subsequent tentatives could be done by telephone, although this opportunity is not often used by the interviewers (about 5% of the contact efforts). High population density areas (Brussels in particular) are systematically less attempted by telephone.

(4) Daily period and day of the week

The ESS protocol advises fieldwork operators to spread their contact attempts such that a variety of contact occasions are offered to the sample unit. Doing so, the interviewers maximize the probability of finding their target at home as they mix afternoon visits, morning visits, evening visits and weekend attempts. It appears to be that most tentatives occurred during daytime (between 12.00am and 6.00pm). Only 11% of visits took place before 12.00am and 20% took place after 6.00pm. 22% of the contact attempts were during the weekend. The treatment variable referring to the daily period counts three categories (before 12.00am, between 12.00am and 6.00pm and after 6.00pm). The seven days of the week are obviously counted as the second treatment variables that pertains to the timing of the visits.

(5) Re-assignment of interviewers

After the first four mandatory contact tries, interviewers could be replaced. Often, better interviewers were deployed. This treatment variable reflects whether the original interviewer is in charge of the follow-up of the particular sample case.

(6) Hold periods

The ESS central coordination team highly recommends that the elapsed time between the first and the last contact attempt should be longer than one month before a sample case can be registered as a final noncontact. Evidence however suggests that this prescription is violated in about 50% of the noncontacted units. It is expected that a prolonged contact period results in a higher contact success, as long absences can be bridged this way. Longer hold periods may also result in better refusal conversion rates. It is possible that deadline constraints, transportation constraints or interviewer specific conditions treat sample cases differently with regard to this advice.

(7) Non-treatment

As was already discussed in section 5, fieldwork decision may lead to censoring: some cases are not visited any more, even if they have not been contacted or interviewed. It was already discussed that a system of weight variables can take this selectivity into account.

7 Results

As the gross sample of 3249 cases is distributed among the interviewers, the inclusion/exclusion-process towards the final or net sample starts. The different obstacles that have to be taken are successively:

1. As the interviewers visit the address contact has to be made.
2. As soon as contact is established, it has to be determined whether the sample case is eligible.
3. Among all contacted and eligible cases, other nonresponses (then refusals) are met such as illness or language barriers.

4. Among all contacted, eligible and available individuals, some will participate, others will refuse to.

At each of these steps, representativeness will be assessed, taking into account the sometimes subtle differences with regard to the treatment variables. At the end of the process, the final representativity will be presented. In each of these cases, we will deal with the different fieldwork conditions: raw propensities ignoring any process variable, equal selection propensities and finally equal selection and equal treatment propensities. Graphical representations will be provided as a function of increasing contact attempts or visits. Partial indicators will serve to assess the contribution of the auxiliary variables to the (non)representativeness. We will also suggest which actions or treatments may be considered to improve representativeness.

Notice that all available auxiliary and treatment variables have been used for the estimation of the propensities. The reason is that there is no a priori knowledge about their impact and that their influence can change during the fieldwork.

7.1 Making contact

The ESS central coordination team expects all participating countries to achieve contact rates of 97% and more. It also advises, in order to succeed, to have all sample cases visited at least four times before the case can be considered as a final noncontact. Also a good mixture of evening, daytime and weekend attempts is proposed. At least the first contact attempt should be made personal (no telephone). The time span between the first and the last attempt should be at least one month. Of course, in some cases these requirement were not met and in other cases interviewer did more efforts than was expected. Differences in reselection for noncontact conversion and differences in contact treatments may lead to an increase or a decrease of the propensity variance. Propensities have been obtained by the procedures that are presented in section 5. Treatment variables are: contact skills of the interviewers, the day of the week as well as the timing of the attempt, the elapsed time with regard to the previous attempt, the contact mode (personal or not), whether or not a new interviewer was deployed and finally whether or not the sample case was assigned at the beginning of the fieldwork or only later after two months.

The R-indicators, contact rates and noncontact bias are depicted in the two graphs shown below. The number of contact attempts at the horizontal axis should be interpreted cumulatively: it refers to the representativity situation until the j^{th} contact attempt. Note that the propensities are modelled such that the auxiliary as well as the treatment variables are allowed to have different effects at every new contact attempt. All treatment variables are believed to have similar effects with respect to the different subgroups in the sample.

Figure 5 shows that the contact rate (raw propensities) increases from about 60% after the first attempt to 97% after 8 contact attempts. The R-indicator also improves, as depicted in Figure 4: the starting value of the contact R-indicator is about 0.80 at the first attempt and it increases to a value of about than 0.90 after 8 attempts. These findings at least apply to the raw propensities, leaving aside the differences in re-selection and treatment. Because the first four contact attempts are mandatory, the raw R-indicator and the equal selection R-indicator coincide. During the first three contact attempts, the equal treatment (and equal selection) R-indicator is below the raw equivalent. This suggests that the fieldwork efforts have had a favourable effect on the representativeness of the contacted sample. This is also reflected by the nonresponse bias curve in Figure 5.

Figure 4: Evolution of the contact R-indicator as the number of contact attempts increases

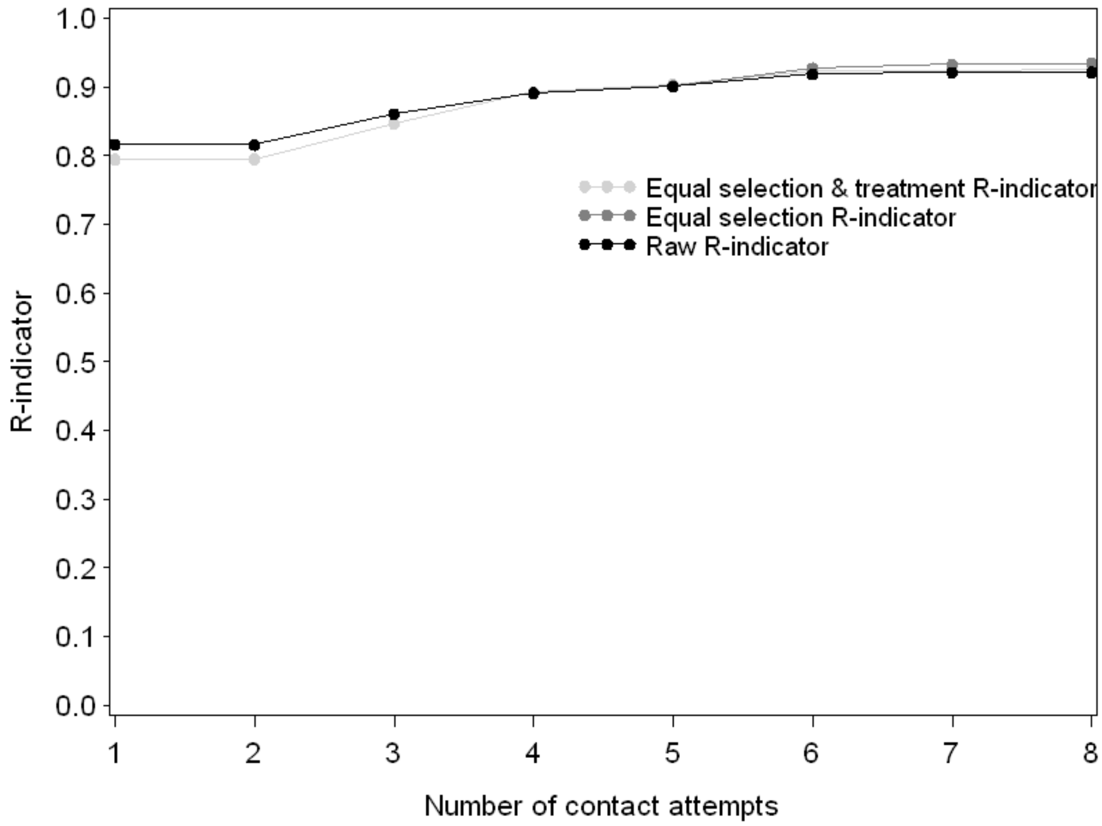
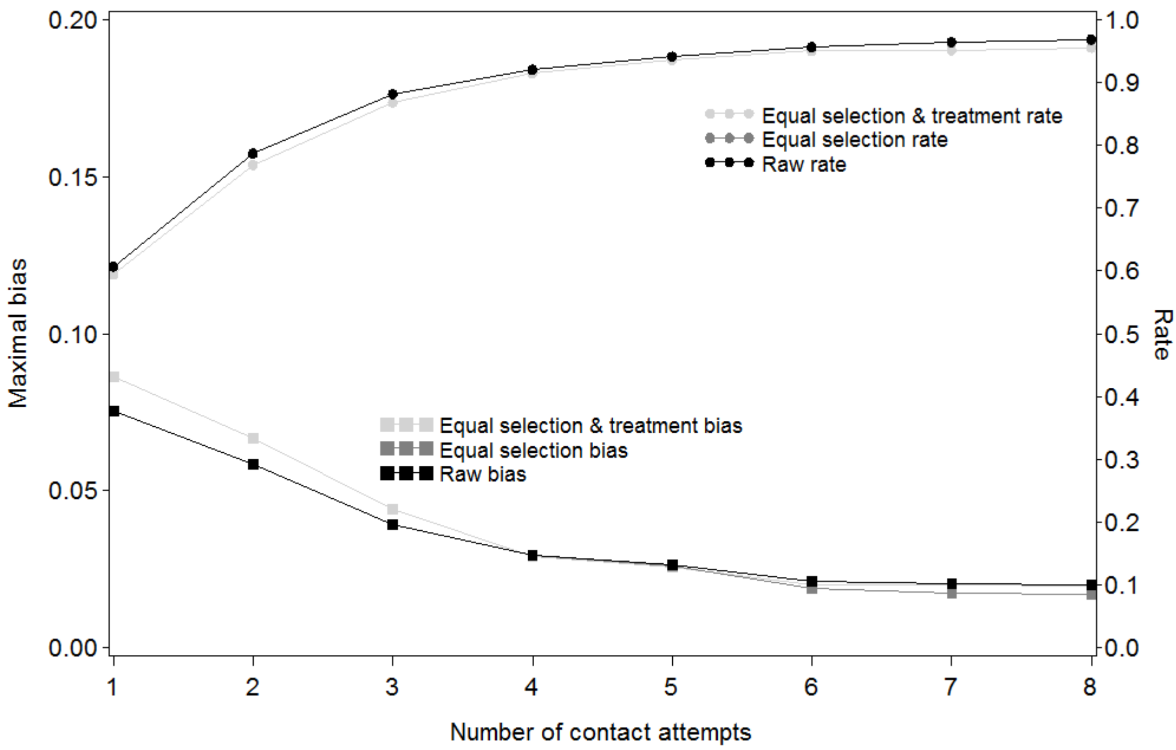


Figure 5: Evolution of the contact rate and noncontact bias as the number of contact attempts increases



After the fourth contact attempt, traces of a slight selectivity bias emerge: if all noncontact units would have had equal (re-)selection probability, the R-indicator would slightly improve. The treatments however seem to compensate this bias as the raw R-indicator and the equal treatment and selection R-indicator coincide after the fourth contact attempt. Overall, it seems that particularly the (re-)selection of the sample units has an unfavourable impact on the representativity of the contacted sample. The treatment efforts toward the selected cases on the other hand realize a recovery of this selectivity bias. Consequently, if the fieldwork management seeks to reduce the variance of the contact propensities, it may consider the selection of less promising profiles for noncontact conversion.

Table 5 shows where to look for the sources of representativity bias. It shows the unconditional partial R-indicators (P_2 – see section 2 for the definition) after the 8th contact attempt. The partial indicators do not only refer to the categories of the variables but are also calculated on the variable-level. The results in the table suggest that Brussels, high population density areas, areas counting much foreigners, apartment dwellers and people living in poor neighbourhood conditions are harder to contact. Also males seem to be slightly more difficult to contact. The conditional partial indicators (P_3) suggest that the type of dwelling is perhaps the most dominant variable determining the variance in the contact propensities.

Regarding possible remedies to overcome a lack of representativity is a disproportional selection of hard-to-contact profiles. A second possibility lead to the scrutiny of the treatment variables. As suggested by the table in the appendix, the skills of the interviewer are by far the most influential variables for contact success. Assigning the more difficult groups the better interviewer may improve contact representativity. An overview of the effects of the treatment variables can be found in the appendix. It is obvious that the reduction of the noncontact rate decreases the maximal bias.

7.2 Determining eligibility

Of the 3249 original sample cases, 3169 were contacted. Eligibility of these cases has to be determined now. Of the 80 cases that were not contacted, there is no information available on which the eligibility can be evaluated. Of the remaining 3169 units, more than 300 units did not fit the definition of the target population (derelict homes, person moved out of the country, deceased, etc.), probably due to an inferior quality of the sample frame. This considerable amount of cases may have a serious effect on the composition of the sample, all the more because no specific replacements were searched for with the same characteristics (age, gender or place of residence). We would also like to emphasize that the assessment of eligibility is an integral part of the contact process. As can be observed from the Belgian contact files, some cases have first been found ineligible, re-approached and eventually interviewed. Other cases initially refused cooperation, but were found ineligible during refusal conversion activities. On the other hand, it is expected that (in)eligibility is a rather stable state: additional efforts to convert ineligibles into eligibles are not common survey practices. All these elements suggest that the measurability of eligibility is not straightforward during data collection and that prudence is called for not to ignore this possible source of vagueness.

Therefore we will simply evaluate eligibility at the individual level leaving aside the possible effects of fieldwork operations and we will stick to the decisions of the fieldwork administration with regard to the assessment of eligibility.

Table 5: Overview of unconditional and conditional partial R-indicators

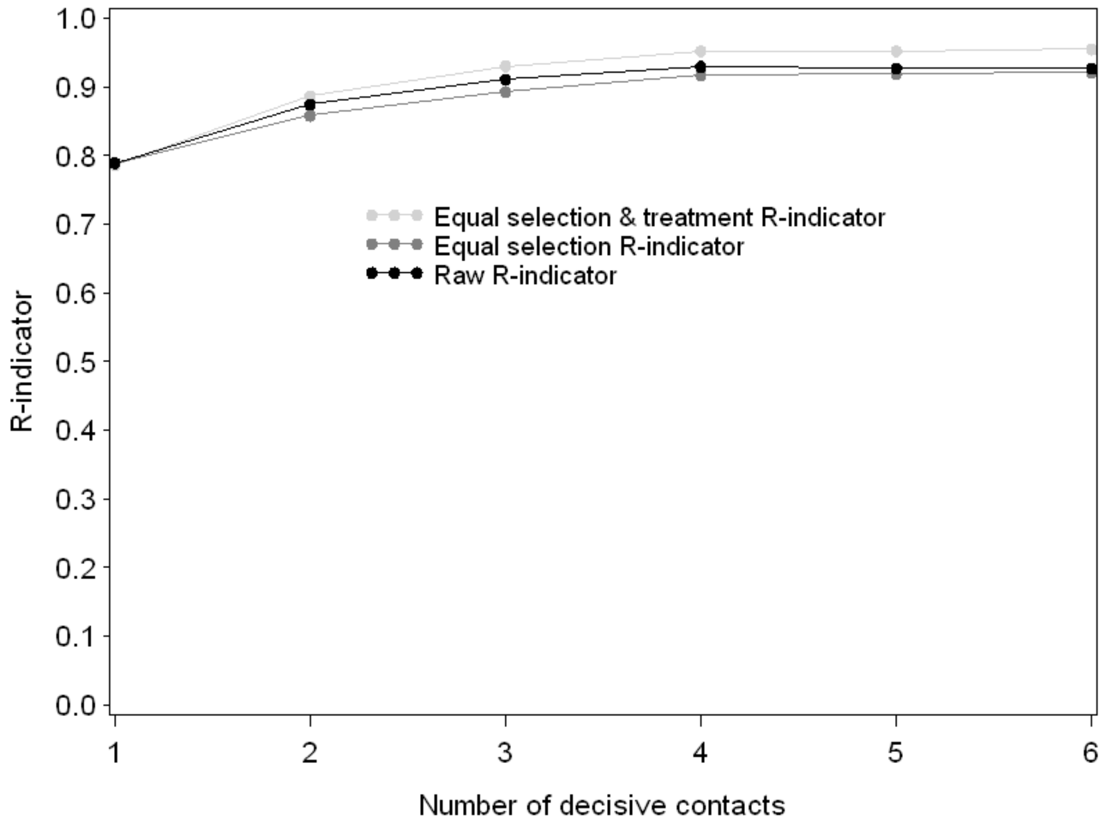
	Class size	Contact		Eligible	Available		Cooperation	
		Raw	Equal S & T		Raw	Equal S & T	Raw	Equal S & T
n		3249	3249	3169	2847	2847	2578	2578
R-indicator		0,92	0,92	0,86	0,93	0,96	0,89	0,86
Rate		0,97	0,96	0,90	0,91	0,93	0,71	0,70
Bias		0,02	0,02	0,04	0,02	0,01	0,04	0,05
Unconditional partial R-indicator (P2); italically printed indicators refer to the variable level								
Age classes		*0,0017*	*0,0038*	*0,0414*	*0,0131*	*0,0169*	*0,0247*	*0,0336*
Age <20	272	-0,0012	0,0031	0,0036	0,0033	-0,0104	0,0247	0,0294
Age 21-40	1013	0,0000	-0,0008	-0,0360	0,0000	-0,0169	0,0000	0,0000
Age 41-60	1137	0,0000	-0,0023	0,0126	0,0098	0,0065	0,0000	0,0084
Age >60	827	0,0012	0,0023	0,0234	-0,0115	0,0169	-0,0165	-0,0210
Gender		*0,0052*	*0,0054*	*0,0180*	*0,0098*	*0,0104*	*0,0000*	*0,0042*
Female	1666	0,0041	0,0038	0,0126	0,0082	0,0104	0,0000	0,0042
Male	1583	-0,0041	-0,0046	-0,0144	-0,0098	-0,0117	0,0000	-0,0042
Region		*0,0274*	*0,0131*	*0,0252*	*0,0115*	*0,0091*	*0,0329*	*0,0420*
Flanders	1908	0,0099	0,0084	0,0126	0,0049	0,0052	0,0124	0,0252
Brussels	306	-0,0285	-0,0115	-0,0234	-0,0131	-0,0130	-0,0412	-0,0420
Wallonia	1035	0,0023	-0,0046	-0,0036	0,0000	-0,0013	0,0000	-0,0126
Population density		*0,0279*	*0,0154*	*0,0288*	*0,0180*	*0,0104*	*0,0329*	*0,0420*
≤200 inh./km²	378	0,0058	0,0046	0,0108	0,0000	-0,0013	0,0041	0,0000
201-400 inh./km²	846	0,0076	0,0054	0,0090	0,0131	0,0065	0,0165	0,0168
401-700 inh./km²	585	0,0099	0,0077	0,0090	0,0082	0,0039	0,0165	0,0168
701-2500 inh./km²	1053	-0,0006	-0,0061	-0,0072	-0,0115	0,0000	-0,0124	-0,0042
>2501 inh./km²	387	-0,0279	-0,0123	-0,0252	-0,0115	-0,0130	-0,0329	-0,0420
Non-Belgians in area		*0,0215*	*0,0146*	*0,0324*	*0,0180*	*0,0091*	*0,0288*	*0,0378*
<2%	639	0,0087	0,0092	0,0180	0,0131	0,0052	0,0206	0,0252
2-5%	981	0,0076	0,0061	0,0144	0,0082	0,0039	0,0082	0,0126
5-15%	1062	0,0012	-0,0038	-0,0108	-0,0131	0,0013	-0,0041	-0,0168
>15%	567	-0,0210	-0,0115	-0,0234	-0,0066	-0,0117	-0,0288	-0,0252
Anual Income in area		*0,0198*	*0,0123*	*0,0198*	*0,0131*	*0,0065*	*0,0124*	*0,0252*
<12.000 €	576	-0,0186	-0,0061	-0,0180	-0,0082	-0,0065	-0,0124	-0,0210
12.000-14.000 €	1233	0,0041	-0,0008	0,0018	-0,0016	-0,0039	0,0000	-0,0084
14.000-16.000 €	1062	0,0099	0,0092	0,0108	0,0115	0,0052	0,0082	0,0210
>16.000 €	378	-0,0012	-0,0069	0,0000	-0,0066	0,0052	0,0041	0,0000
Dwelling		*0,0256*	*0,0238*	*0,0432*	*0,0180*	*0,0065*	*0,0412*	*0,0462*
No apartment	2662	0,0122	0,0115	0,0198	0,0082	0,0039	0,0165	0,0210
Apartment	587	-0,0256	-0,0238	-0,0414	-0,0213	-0,0091	-0,0453	-0,0462
Neighbourhood quality		*0,0192*	*0,0138*	*0,0306*	*0,0147*	*0,0065*	*0,0165*	*0,0252*
Poor	706	-0,0192	-0,0138	-0,0288	-0,0147	-0,0078	-0,0165	-0,0252
Good	997	0,0070	0,0069	0,0144	0,0115	0,0039	0,0082	0,0126
Excellent	1546	0,0070	0,0031	0,0072	0,0000	0,0026	0,0041	0,0042
Conditional partial indicator (P3), at variable level								
Age classes		*0,0052*	*0,0069*	*0,0360*	*0,0131*	*0,0195*	*0,0206*	*0,0252*
Gender		*0,0052*	*0,0054*	*0,0162*	*0,0098*	*0,0117*	*0,0041*	*0,0042*
Region		*0,0058*	*0,0092*	*0,0054*	*0,0033*	*0,0013*	*0,0082*	*0,0084*
Population density		*0,0058*	*0,0084*	*0,0036*	*0,0082*	*0,0039*	*0,0124*	*0,0084*
Non-Belgians in area		*0,0023*	*0,0054*	*0,0108*	*0,0082*	*0,0039*	*0,0082*	*0,0126*
Anual Income in area		*0,0076*	*0,0123*	*0,0072*	*0,0082*	*0,0052*	*0,0082*	*0,0126*
Dwelling		*0,0151*	*0,0184*	*0,0288*	*0,0098*	*0,0039*	*0,0206*	*0,0210*
Neighbourhood quality		*0,0087*	*0,0100*	*0,0126*	*0,0098*	*0,0026*	*0,0041*	*0,0084*

Table 5 suggests a considerable representativity issues with respect to eligibility, as the R-indicator equals only 0,86, resulting in a rather unfavourable risk of bias of 0,04. The age class 21-40, males, residents of Brussels, high density and low income areas as well as area with high proportions of non-Belgians are more likely to be excluded because of ineligibility. Also apartments and less attractive neighbourhoods are more associated to ineligibility. Age, gender and the type of dwelling seem to be the most dominant variables in this regard, as indicated by the conditional partial indicators (P_3).

7.3 Other nonresponse

So far, the initial gross sample has been reduced such that noncontacts and ineligibles have been removed. The current subsample counts 2847 elements. Other nonresponse than refusal can occur since target respondents can be ill or have left for a few days or weeks. Also language barriers can cause that individuals are not available for an interview. Such sources of nonresponse can however be converted by simply trying another time or sending another interviewer. The following figures show how this subprocess evolves with respect to representativity during the course of the fieldwork.

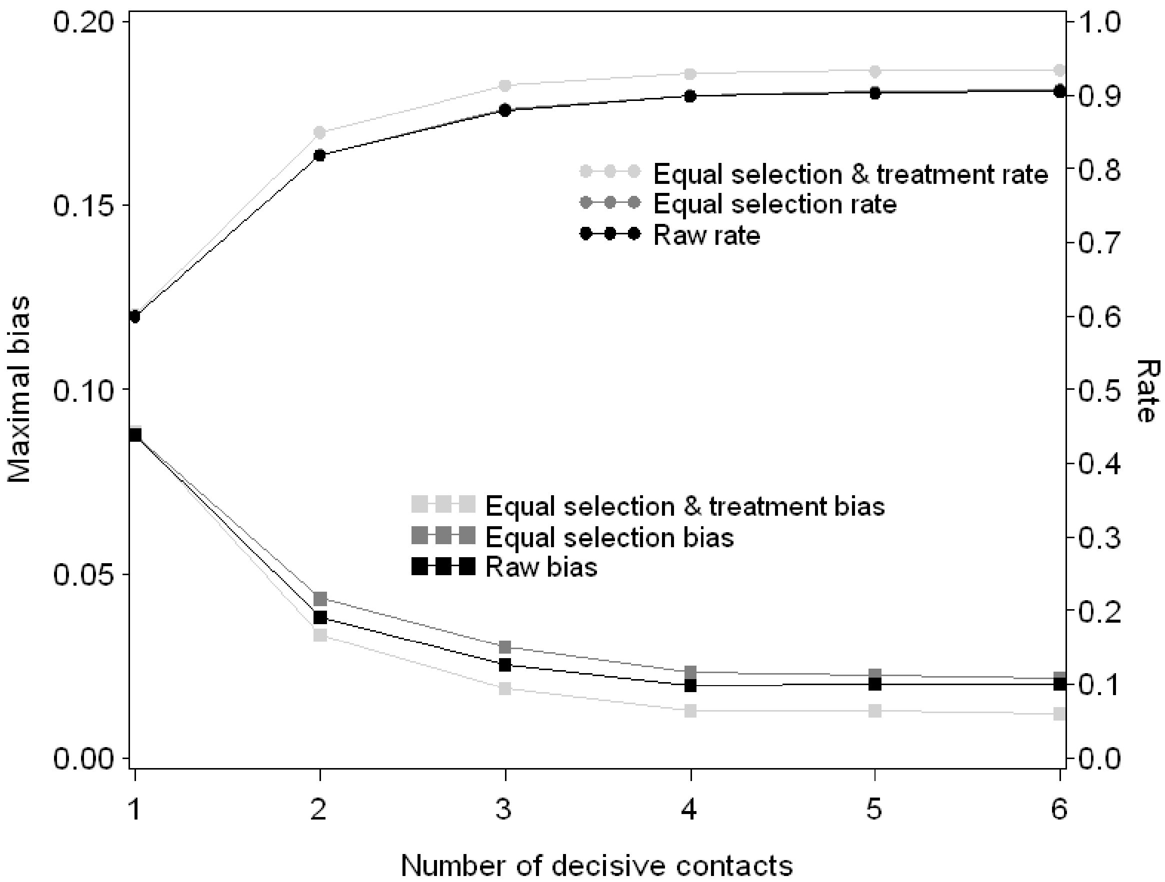
Figure 6: Evolution of the availability R-indicator as the number of contacts increases



The representativeness of this subprocess improves as the fielding advances. Note that visits where no contact was established are not taken into consideration for calculation of these R-indicators. In other words, only visits that could possibly determine the availability of the target individual are taken into account. Also note that some of the treatment variables have slightly been changed so that they are more relevant to this analysis. The skills of the interviewer do no longer refer to the contact competences, but rather relate to the competences of interviewers finding someone available. Also the hold period is now the time span between two successive visits were contact is made. The graph above suggest that the selection of the cases for conversion has slightly improved the representativeness with regard to availability. On the other hand, if all selected cases would have been treated equally, the R-indicator will probably have improved. Consequently, the fieldwork management is advised to concentrate on the treatment of the selected cases, if it wants to improve availability representativeness.

Excessive deviations of variable categories cannot be found in Table 5 with respect to P_2 or P_3 partial indicators. On the other hand, it becomes clear that sparse profiles with respect to previous subprocesses such as making contact or ineligibility (males, Brussels residents, high density areas, apartment dwellers, …) are also hard to include in this particular subprocess. Also in the next process, similar observations will be made. This implies correlational tendencies between the different subprocesses.

Figure 7: Evolution of the availability rate and bias as the number of contacts increases



7.4 Refusal versus cooperation

The last part of the contact process is to persuade the target person to participate. Reluctance however is a considerable reason for exclusion from the sample, as about ¼ of the gross Belgian sample refuses to cooperate. The following graphs portray the initial representativity situation at the first request to cooperate and at the occasional refusal conversion attempts.

Figure 8: Evolution of the cooperation R-indicator as the number of decisive contacts increases

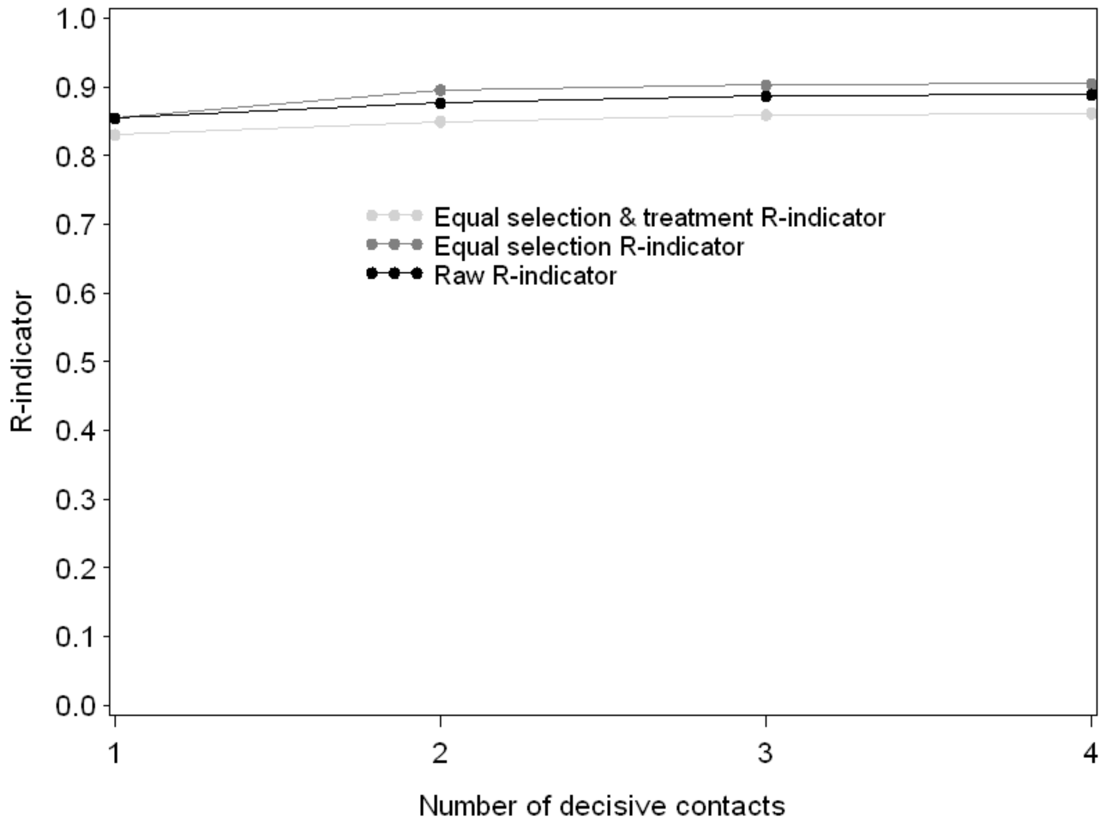
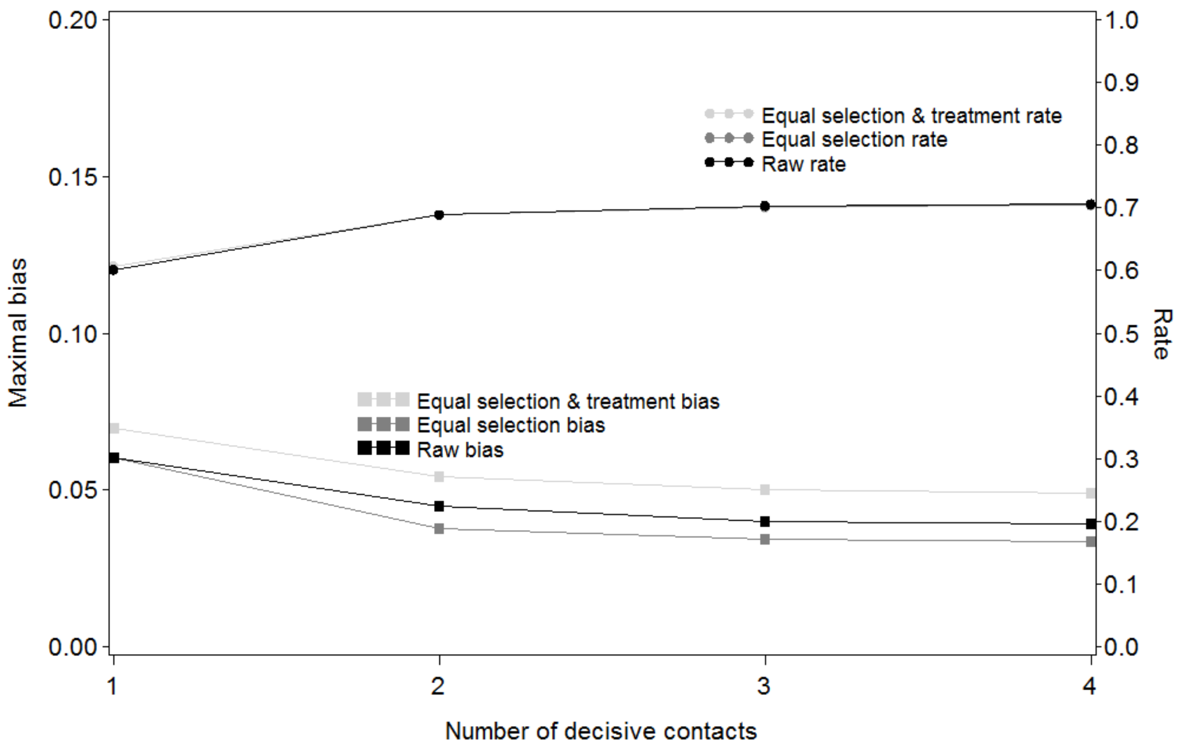


Figure 9: Evolution of the cooperation rate and bias as the number of decisive contacts increases



The results suggest that the selection of sample units has slightly prioritised the more promising cases, although the equal treatment condition largely compensates this selection bias. This may be the result of keeping only the better interviewers for the refusal conversion activities. It is again striking that the same profiles (males, apartment dwellers,…) are the critical subgroups in this last subprocess. This means that all subprocesses reinforce the same exclusion mechanism.

Also notice that refusal conversion attempts on have a weak effect on the response rate. Only an additional 8% cooperation rate is realized. The effects on bias reduction are therefore also rather modest. Also notice that the cooperation rate hardly changes when moving from raw propensities to equal selection and treatment scenarios.

7.5 Final representativeness

Since we dispose now of all propensities related the all parts of the contact process, for all combinations of the categories of the contributing variables, we are able to estimate the general or final propensities of the sample units. Moreover, the propensities of the different conditions (raw, equal selection, equal selection and treatment) are available so that final estimates can also be applied to these different conditions. Each time the propensities at the last attempt of each subprocess (e.g. $\rho_{contact}$ at the 8th contact attempt) are multiplied:

$$\rho_{final} = \rho_{contact} \times \rho_{eligbible} \times \rho_{available} \times \rho_{cooperative}$$

The final rates, R-indicators and bias estimates and their respective standard errors are given in Table 6. The standard errors are obtained by recomputing the estimated propensities for 200 replicate samples. This bootstrap method resampled with replacement of the elements.

Table 6: Rates, R-indicators and maximal biases and respective standard errors for three propensity conditions, final propensities

	(Response) rate	R-indicator	Maximal bias
Raw	0,5586 (0,0090)	0,7557 (0,0164)	0,1078 (0,0076)
Equal selection	0,5557 (0,0092)	0,7839 (0,0153)	0,0961 (0,0072)
Equal selection and treatment	0,5725 (0,0114)	0,7509 (0,0189)	0,1065 (0,0089)

In order to assess that the R-indicators significantly change we will not construct the confidence interval of the raw R-indicator and check whether the two other R-indicators fall inside or outside the interval. After all, it can be expected that the three R-indicators (raw, equal selection, equal selection and treatment) are strongly correlated. Therefore an estimate was obtained of the differences between these R-indicators at every replication. The difference in R-indicators between the raw and the equal selection version (0,7557 versus 0,7839) is significant at p = 0,0052. The related maximal biases is also significantly different at p = 0,0069. Also note that the difference in final response rate between the raw scenario and the equal selection and treatment scenario (0,5586 and 0,5725) is significantly different at 0,0362. Other significant differences have not been found.

The results suggest that fieldwork operations have failed to care for a balanced (re)selection of cases during the fieldwork process. The introduction of the treatment variables on the other hand, has amply restored this deficit. This means that the less promising sample cases (once they are reselected) may have been treated with more care in order to have them included in the sample.

Fieldwork managers may be interested in ways to improve the sample's representativity. It was already suggested by Table 6 that an equal (re)selection condition can significantly improve representativeness. Further improvement may of course be achieved when selecting the less promising profiles even more often. By means of simulation studies the effects of different (re)selection regimes can be estimated. Another lever to improve representativeness can be found in the effects of the treatment variables. These are documented in the appendix and suggest that the effect of the different kinds of interviewer skills (e.g. contact skills, persuasive skills) is by far the most important determinant of the outcomes during the contact process. Here again, situations in which various assignment systems are simulated can be deployed to estimate the margins for survey sample improvement.

8 Discussion

The practice of fieldwork monitoring seems to benefit from the insights of Total Quality Management or statistical quality control. It can be used as a steppingstone to better understand the process of sample construction in survey research. Particularly the development of a process flow chart that deconstructs the contact process into the different subprocesses, the identification of critical product characteristics and the determination of key process variables are perceived as a helpful guide during the monitoring activities. The quality improvement framework also cautions for the lack of measurement capabilities. This means that the dependence on the quality, availability and accessibility of both auxiliary (respondent related) variables as well as process variables cannot be underestimated in this respect.

A distinction was made between mere auxiliary variables (respondent characteristics) and fieldwork characteristics or treatment variables (including selection information). This distinction allows to evaluate the quality of the fieldwork, because treatment variables document the decisions and operations during the contact process. These treatment variables also inform the fieldwork managers or other relevant survey agents to improve the contact process. In this regard, the Belgian examples suggested that the (re)selection of sample units for noncontact, refusal of other nonresponse conversion leaves interesting opportunities to improve the fieldwork with regard to representativeness. Also the assignment of specific treatments such as the skills of the interviewers are an important lever for quality improvement.

Note however that this example only relates to the Belgian variant of the third round of the European Social Survey. This survey may consequently not be representative for other surveys. More applications of fieldwork monitoring may therefore be needed.

9 References

Aitken, A., Hörngren, J., Jones, N., Lewis, D., & M.J., Zilhão (2004) Handbook on improving quality by analysis of process variables. Eurostat

Cobben, F. and Schouten, B. (2007), An empirical validation of R-indicators, Discussion paper 08006, CBS, Voorburg, available at www.cbs.nl/nl-NL/menu/methoden/research/discussionpapers/archief/2008

Couper, M. (1998). Measuring survey quality in a CASIC environment. Proceedings of the survey research methods section, American Statistical Association (http://www.amstat.org/Sections/Srms/Proceedings/).

Dippo, C. (1997) Survey measurement and process improvement: Concepts and integration. In L Lyberg et al. (Eds.), Survey Measurement and Process Quality (pp. 457-474). New York: John Wiley Sons.

Groves, R, & M., Couper (1998) Nonresponse in Household Interview Surveys. New York: Wiley.

Lyberg, L. & P. Biemer (2008). Quality assurance and quality control in surveys. In E. de Leeuw, J. Hox & D. A. Dillman (eds.) International Handbook of Survey Methodology 421-441. New York: Lawrence Erlbaum Associates.

Montgomery, D. (2005) Introduction to statistical quality control. New York: Wiley

Morganstein, D.R. and Marker, D.A. (1997). Continuous quality improvement in statistical agencies. In L Lyberg et al. (Eds.), Survey Measurement and Process Quality (pp. 475-500). New York: John Wiley Sons.

Schouten, B., Cobben, F. (2007), R-indexes for the comparison of different fieldwork strategies and data collection modes, Discussion paper 07002, CBS, Voorburg, available at www.cbs.nl/nl-NL/menu/methoden/research/discussionpapers/archief/2007

Schouten, B., Morren, M., Bethlehem, J., Shlomo, N., Skinner, C. (2009) How to use R-indicators? Work Package 4, Deliverable 3, RISQ Project, 7th Framework Programme (FP7) of the European Union, available at www.r-indicator.eu.

Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J., Zhang, L. (2008) Statistical Proporties of R-indicators, Work Package 3, Deliverable 2.1, RISQ Project, 7th Framework Programme (FP7) of the European Union, available at www.r-indicator.eu.

Shlomo, N., Skinner, C., Schouten, B., Carolina, T., Morren, M. (2009) Partial Indicators for Representative Response, Work Package 5, Deliverable 4, RISQ Project, 7th Framework Programme (FP7) of the European Union, available at www.r-indicator.eu.

Singer, J., & J., Willett (2003) Applied longitudinal data analysis. Modelling change and event occurrence. London: Oxford Univeristy.

Appendix

The impact of the treatment variables have not been documented yet. This may however be interesting since these treatment variables are the means to improve the sample composition. Therefore a brief overview is given here. Table 7 provides the standardized parameter estimates of the of the treatment variables in the following model:

$$
\begin{aligned}
g(h(\rho_{ij})) = {} & [\alpha_1 V_{1ij} + \alpha_2 V_{2ij} + \ldots + \alpha_J V_{Jij}] \\
& + [\beta_1 A_{1i} + \beta_2 A_{2i} + \ldots + \beta_P A_{Pi}] \\
& + [\beta_{T1} T_{1i} + \beta_{T2} T_{2i} + \ldots + \beta_{TP} T_{Pi}]
\end{aligned}
$$

The model has been fit for (1) contact success, (2) finding someone available and (3) persuading someone for an interview (refusal versus cooperation). The effect of the auxiliary variables and the number of visit have been omitted in the table. The parameters are estimated under the assumption that the treatment variables have similar effect on all profiles.

Table 7: Effect of treatment variables of subprocesses of the fieldwork, ESS3 - Belgium

	Contact	**Available**	**Participation**
Sample units in the subprocess	3249	2847	2578
Successes	3169	2577	1815
Failures	3004	1636	1485
Telephone (=yes)	0,13****	-0,05*	0,04*
Interviewer skills			
Contact	0,30****		
Available		0,11****	
Persuasion			0,28****
Initial wave (=yes)	-0,02	-0,03	0,01
Elapsed time	0,05*	-0,20****	-0,03
Weekday (ref: Sunday)			
Monday	0,03	-0,03	0,13**
Tuesday	0,04	-0,04	0,08
Wednesday	0,01	-0,06	0,09*
Thursday	0,02	-0,08	0,14**
Friday	-0,03	-0,07	0,08*
Saturday	0,00	-0,10*	0,05
Daily period (ref: after 6pm)			
before 12am	-0,07***	-0,03**	0,08**
between 12am and 6pm	-0,03	-0,07	0,08**
New interviewer (=yes)	0,01	-0,04	-0,08*

*: p< 0,05; **:p < 0,01; ***: p < 0,001; ****: p < 0,0001

Apparently, the interviewer skills have the strongest effect on the outcomes, particularly with respect to contactability and cooperativeness. Concerning availability, the elapsed time with regard to the previous attempt seems to be most dominant. The negative sign in the parameter indicates that short hold period result in more successful attempts. This can probably be explained by the fact that when the target respondent is not available, the interviewer is informed by a proxy to come back a little later to find the target at home. When no such information is provided by the proxy, the interviewer has no specific clue when to return and so will his/her success rate be, assuming that interviewers will not will not return quickly when no additional information is available about the target person.

The timing (weekday and daily period) have only a modest effect on the successes. Mondays and Thursday are the best occasions the convince individuals to cooperate, as opposed to Sundays. Evening attempts are better to make contact, although they are less favourable to succeed in gaining cooperation. Telephone attempts seem to be more successful in contacting the target unit, although telephone attempts are believed to be underreported (unsuccessful attempts are systematically less documented in the contact records). Finally, assigning a new interviewer apparently has a slight negative impact on the cooperation success.