Research & Methods
ISSN 1234-9224 Vol. 20 (1, 2011): 27–58
Institute of Philosophy and Sociology
Polish Academy of Sciences, Warsaw
www.ifispan.waw.pl
e-mail: publish@ifispan.waw.pl

Do More Contact-attempts Reduce Non-response Bias in Representative Face-to-Face Interviews? Findings from a PAPI Survey with a Low Response Rate

Piotr Jabkowski
Adam Mickiewicz University Poznań

The aim of this article is to examine whether providing pollsters with multiple opportunities to carry out interviews with a sample of randomly selected respondents reduces non-response bias. First, I present the procedures for assessing non-response bias in sociological surveys, opting for a method based on post-stratification weights. Next, using a unique dataset, I address four basic questions related to the research problem: (1) Do more contact attempts reduce the dissimilarity between the sample and population distributions? (2) Do more contact attempts reduce the need for post-stratification weighting? (3) Which categories of respondents are underrepresented and/or overrepresented in the sample in the early and late contact attempts? (4) Do more contact attempts reduce non-response bias? While more contact attempts increased the response rate and closed the gap between sample and population, they do not always reduce the level of non-response bias. Indeed, for some types of variables, the more contact attempts, the larger the non-response bias.

Key words: non-response bias; contact-attempts; response rate; dissimilarity between sample and population

INTRODUCTION

Analyses of survey quality are mainly focused on relations between response rate and non-response bias. Low survey response rates may lead to non-response bias, i.e. significant dissimilarities between the sample and population structure (and thus lesser representativeness of the sample). While researchers focus on devising effective methods for reducing non-response bias[1] (Billiet at al. 2009: 7–9), we have few firm guidelines for reconciling response rate and non-response bias[2]

(Groves 2006: 649–650). Some reasons for low response rate have no bearing on the extent of non-response bias, while others have considerable influence (Groves 2006: 650–651). Moreover, within a single survey, non-response bias might be significant with regard to some variables and not others[3] (Sztabiński et al. 2007: 32–37), which makes it even more difficult to make any conclusions about the relations between response rate and non-response bias.

When conducting quantitative research, researchers try to reduce the level of non-response bias by attempting to obtain the highest possible survey response percentage[4] (Stoop 2005: 23–25). The strategy of increasing contact attempts relies on the pollsters' returns to a particular address at different times and on different days of the week (Stoop 2005: 275–279), which seems to be a relatively easy way of improving the survey response (Sztabiński at al. 2009: 72–74). It is effective because it increases the chances of conducting interviews with individuals who are difficult to reach or possess specific and desired social and demographic features (Stoop 2004: 40–43). Such a strategy, however, can influence the quality of a survey, as prolonged fieldwork periods may have as negative an influence as low response rate (Domański 2006: 43–46). Some researchers find that efforts to reduce non-response rate have little effect on non-response bias (Fricker et al. 2010: 935).

The situation gets even more complicated when the survey is conducted on populations and samples with historically low response rates (e.g. representative surveys in large cities) regardless of the efforts undertaken by the researchers[5] (Grzeszkiewicz-Radulska 2009: 167–175) or when the techniques used are burdened *a priori* with high levels of non-response, e.g. telephone surveys (Curtin et al. 2005: 90–95), or website surveys (Vehovar et al. 2002: 230–232). In such cases, even with a long fieldwork period and multiple attempts to make contact, non-response level may still remain very high (Keeter et al. 2006: 762–763). It seems reasonable to ask whether the strategy based on multiple attempts to reach a respondent is effective for attaining a satisfactory level of response rate and reducing non-response bias. With regard to surveys with low expected response rate, perhaps we should direct efforts to some other actions designed to reduce non-response bias (e.g. short surveys with few clue questions in situations, in which the pollster finds that the interview with the respondent will most likely not take place)[6] (Lynn 2003: 241–252).

The aim of this article is to examine whether providing pollsters with multiple opportunities to carry out interviews with a sample of randomly selected respondents reduces non-response bias. It is important to establish if a higher number of visits always reduce the negative effects of non-response. On the basis of research done with the Quality of Life Indicators data in Poznan research project, conducted on samples of residents randomly selected by name, I will try to address four basic issues related to this subject:

(Q1) Do more contact attempts reduce the dissimilarity between the sample and population distributions?

(Q2) Do more contact attempts reduce the need for post-stratification weighting?

(Q3) Which categories of respondents are underrepresented and/or overrepresented in the sample in the early and late contact attempts?

(Q4) What is the level of non-response bias in point estimation? Do more contact attempts reduce the differences between unweighted and weighted estimators?

THEORETICAL AND METHODOLOGICAL FRAMEWORK

Non-response bias can be classified as systematic errors. Random errors – characteristic of research conducted on samples – is less of a problem, since they can be accurately established, predicted and, as a result, minimized (Jabkowski 2007: 67). Systematic errors, as they fall outside the routine statistical analyses and require additional, expensive and methodologically advanced empirical analyses, are a much more serious challenge. Non-response is just one of many sources of systematic errors, but at the same time it is also a fairly general source of total survey error (Biemer 2010: 821–825)[7].

Non-respondents do not form a homogeneous group. There are many reasons for unavailability: some may be beyond the researcher's influence (e.g. incorrect address data, incorrect assignment of the selected person to the surveyed population, change of address excluding the respondent from the surveyed population, going away of the respondent during the research, as well as the death or ill-health of the respondent), others remain within his or her control (e.g. refusal to participate in a survey, breaking off the interview without chances for continuation, and the respondent's avoidance of meeting with the pollster despite numerous attempts to make contact). The presence of unavailable individuals, i.e. unit non-response (as well as the respondent's lack of answer to specific question(s), i.e. item non-response) influences the value of non-response bias (Grzeszkiewicz-Radulska 2009: 22–24). I focus on unit non-response.

When the fieldwork period ends, the research sample is divided into two separate sets – the persons who were interviewed (respondents) and the persons who were not (non-respondents). In other words, in a sample containing N elements, one will always have n_1 respondents and n_2 non-respondents (while $N=n_1+n_2$). These two values allow one to establish response rate[8] $w_1=n_1/N$, as well as non-response rate $w_2=n_2/N$. Non-response bias is defined as the difference between the estimator value in the respondent layer and the unknown estimator value, which could be established if the survey had been conducted with all the persons selected for the sample (Billiet et al. 2009: 6). Since this difference determines the non-random and systematic deviation of the observed results from the real results, it is a measure of systematic

error. Non-response bias can also be defined as the function of non-response rate and the difference between respondents and non-respondents. For the percentage estimator, non-response bias then takes the form of a well-known expression:

$$\hat{b} = w_2(\hat{p}_1 - \hat{p}_2) \tag{1}$$

where $\hat{p}_1$ is the value of the percentage estimator in the respondent layer, while $\hat{p}_2$ denotes the unknown percentage estimator value in the non-respondent layer (Jabkowski 2007: 73). For the arithmetic mean estimator, the error rate can be noted as:

$$\hat{b} = w_2(\hat{\mu}_1 - \hat{\mu}_2) \tag{2}$$

where $\hat{\mu}_1$ indicates the respondent mean, and $\hat{\mu}_2$ is the nonrespondent mean (Groves 2006: 648; Lisowski 1971: 10).

As visible from formulas (1) and (2), the value of non-response bias depends on two factors. First is the presence of unavailable units in the research sample; there is no minimum response rate below which the survey estimates are necessarily subject to bias (Groves 2006: 650). Second is the difference between the values of the statistics for respondents and non-respondents (Jabkowski 2007: 93). If non-respondents were randomly distributed throughout a population, bias would not be a problem even in high non-response rates. If that was the case, we would not have to worry about the response rate level at all, as this value would only influence random errors (which are not a serious problem in surveys based on random sampling), not being actually related to systematic errors. Most research suggests that the division of a sample into available and unavailable is strongly correlated with the known characteristics of the persons drawn; the unavailable units differ from the persons not surveyed in terms of their opinion distribution – substantive variables, as well as their socio-demographic characteristics (Domański 1999: 72–78, Groves 2006: 657–662, Sztabiński et al. 2007: 31–37).

Defining non-response bias as the difference between the estimators in the available and unavailable unit layers leads to difficulties is establishing the *extent* of this error. The main difficulty consists in collecting adequate data. The main and most obvious limitation is the difficult access to the unavailable individuals; when the exact values of the unavailable unit layer are unknown, the extent of the error remains undetermined. While it is easy to establish response rate, not much is known about the non-respondents (Domański 1999: 67). There were many procedures to increase room for manoeuvre in this area, but they are not free from flaw, including the disputable and problematic assumptions on which they are based (Grzeszkiewicz-Radulska 2009: 47).

Groves (2006: 654–657) collated and described five basic procedures for assessing non-response bias in representative surveys. The characteristics, limitations and possible uses of these methods were described in detail elsewhere, and I will discuss only very general premises. (1) The first method is based on response rate comparison across subgroups – this assumes that there is no evidence for "non-response bias" if response rates are similar across certain subgroups of a population (e.g. sex, age, education level, etc.); (2) the second procedure is based on using rich sampling frame data or supplemental and matched data; for example, by using variables on an external data set, a researcher compares respondent and non-respondent values; we can use this method when we can match each person in the sample with individual records taken from an external database, (3) the third method consists of comparing response based estimates with similar estimates from other more accurate sources (e.g. official statistical data); (4) the fourth is based on studying variation within the existing survey: nonresponse follow-up studies; (5) the last one is based on contrasting alternative post survey adjustments for non-response. Groves (2006: 657) points out that "because of the diverse properties of the techniques above, it is wise to study nonresponse biases using multiple methods simultaneously."

In this study I used one method located in the fifth group of procedures for assessing non-response bias, based on post-stratification weights, where the bias is defined as a deviation between sample and population distribution. This procedure was described in details by Billet et al. (2009: 9–21) and used by them to compare non-response bias in the European Social Survey (ESS). In view of the Groves' advice to study of nonresponse bias by recourse to multiple methods at the same time, my decision to use only one procedure seems problematic. This obviously affects the theoretical and methodological limitations of the conducted analyses, but due to the lack of adequate data I could not use the procedures (2), (3) as well as (4), while the procedure (1) is in fact useless when it comes to estimating the level of non-response bias for substantive variables in survey.

So in the used method, we have to establish the expected population distributions (a.k.a. "golden standards") for several variables and compare them with the obtained sample distributions. – the source of the data must be reliable, and we can use official statistics. In my study, the data was weighted by three stratification variables (1) gender, (2) age (16–24, 25–34, 35–44, 45–54, 55–64, 64+) and (3) education (Primary or no education, Vocational, Secondary schools or post-secondary, Bachelor's Degree or Higher). In fact, these three variables are commonly used in post-stratification weighting in many surveys (for example in ESS). Weights are computed by dividing the cell proportion in the multivariate 3D table in the population (two categories of gender are multiplied by six categories of age and four categories of education, so we have forty-eight different cells) by the corresponding cell proportion in the obtained sample (Billet et al. 2009: 11).

Weighting of a research sample influences not only the values of estimators, but also their variations, which increases the statistical error. The evaluation measure of the influence of such effect is the Variance Inflation Factor (VIF), which shows the degree to which the variance value in the obtained sample is multiplied in the weighted sample, i.e. $Var(\overline{y}_w) = Var(\overline{y}_{SRS}) \times VIF$, where $VIF = 1 + \frac{s_w^2}{\overline{w}^2}$, $\overline{w}^2$ is a square of the arithmetic mean and s_w^2 is the variance of all weights values (Billet et al. 2009: 11).

Non-response bias is here defined as the difference between the estimator value calculated in the unweighted sample and weighted sample. For the percentage estimator, non-response bias can be noted as follows:

$$\hat{b} = \hat{p}_{SRS} - \hat{p}_w \tag{3}$$

where $\hat{p}_{SRS}$ is the value of the percentage estimator in the unweighted sample and $\hat{p}_w$ is the value of the same estimator in the weighted sample. As for the arithmetic mean estimator, we note non-response bias as:

$$\hat{b} = \overline{y}_{SRS} - \overline{y}_w \tag{4}$$

where $\overline{y}_{SRS}$ is the value of the arithmetic mean estimator in the unweighted ample, while $\overline{y}_w$ is the respective value in the weighted sample.

If non-response bias takes a positive value, it means that the estimator value in the obtained sample has been overestimated in relation to its "actual" value. Negative values indicate the underestimation of the estimator value in the obtained sample. Expressions (3) and (4) can be standardized, so that non-response bias percentage is presented as follows:

$$Sbias(\hat{b}) = \frac{\hat{p}_{SRS} - \hat{p}_w}{se(\hat{p}_{SRS})} \tag{5}$$

where $se(\hat{p}_{SRS})$ is the standard error of $\hat{p}_{SRS}$ estimator. Similarly for arithmetic mean:

$$Sbias(\hat{b}) = \frac{\overline{y}_{SRS} - \overline{y}_w}{se(\overline{y}_{SRS})} \tag{6}$$

where $se(\overline{y}_{SRS})$ is the standard error of $\overline{y}_{SRS}$ estimator.

This non-response bias measure modification (named standardized non-response bias or standardized bias) is advantageous: in comparison to the values calculated from formulas (3) and (4), we can, in each survey, compare the

non-response bias between different variables, as well as compare, for selected variables, the non-response bias between two or more surveys. Since both statistics have asymptotic normal distribution 0–1, it lets us denote the statistically significant biases at the assumed significance level. Of course, at the 0.10 level of significance, the critical values are the range $(-\infty;-1.65]\cup[1.65;+\infty)$ and for 0.05 level, they are the range $(-\infty;-1.96]\cup[1.96;+\infty)$. This defines non-response bias as the difference between the estimator value in unweighted and weighted sample.

Though widely used, this method of estimating the influence of non-respondents on non-response bias has disputable assumptions. Limitations are discussed in detail in Billet et al. (2009: 14, 20–21). First, Billet et al. (2009) point to the problematic nature of the assumption that the differences between weighted and unweighted estimates are merely attributed to non-response. Such an assumption is particularly risky when there is no knowledge about other potential sources of errors that influence the total non-response error. Second, one needs to make sure if the deviations from the "golden standards" in the distribution of variables, which constitute the basis for post-stratification weighting in the obtained sample, are the result of unit non-response or the effect of errors in sampling. Third, weights may be biased in themselves when the source of "golden standards" does not accurately reflect population distribution. Fourth, this procedure assumes that for non-random non-respondents, the bias may be in fact seriously underestimated[9]. Fifth, the bias may be underestimated also when the variable for which this bias is determined is not sufficiently correlated with post-stratification variables. As Billet et al. (2009: 20) summarize, "there is strictly no guarantee that the adjusted sample reflects better the distribution of the target variable in the population".

DATA DESCRIPTION AND SCHEME OF ANALYSIS

The analyses presented are based on the results of the Quality of Life Indicators in Poznan (QoL 2008) research project. The population included the residents of Poznan aged between 16 and 70 and, registered in the data repository of the Poznan City Council for permanent or temporary residence at least three months before the selection of the sample. The interviews were conducted on residents randomly selected by name, without the possibility of surveying the respondents on the reserve list. Using a substitute sample has no theoretical justification and does not increase the quality of statistical conclusions (Jabkowski 2007: 68–84). The number of individuals drawn for the sample was appropriately higher than the target size of the sample so that, having subtracted the non-responders from the target sample, the remaining part would correspond to the intended sample size (Sawiński 2005: 84–85).

Events of each visit (result and reason for not conducting the interview, primarily) were written down by the pollsters on special research return rate sheets[10] based on contact forms used in the European Social Survey[11] (Domański 2006: 31). Each pollster was required to undertake at least three attempts to make contact with the respondent assigned to him or her; the attempts were to take place at different times of the day, with one of them during the weekend. If after these three attempts the interview still did not take place or it was still not possible to contact the respondent, but there was no clear refusal to participate in the survey nor information that the interview cannot take place during the planned survey period, the address was then given to a different pollster. The maximum number of attempts the pollsters could make to conduct the interview was seven visits.

Table 1 presents the values of response rates and the distribution of the number of visits necessary for conducting the interview in QoL 2008, as well as in the 4th round of ESS (ESS_R4_2008).

Table 1 The distribution of the number of visits ending in an interview – QoL_2008 vs. ESS_R4_2008

Number of visits	Poznań QoL 2008		ESS R4 2008[1]	
	Percentage distribution	Cumulative Frequency	Percentage distribution	Cumulative Frequency
1 visit	37,1	37,1	41,0	41,0
2 visits	28,4	65,5	20,0	61,0
3 visits	17,8	83,2	15,2	76,2
4 visits	9,3	92,5	9,5	85,7
5 visits	4,7	97,2	4,8	90,5
6 visits	1,4	98,6	2,9	93,3
7 visits	1,4	100,0	6,7	100,0
RR[2]	47,1		50,5	

[1] The ESS data analysis for cities with more than 500 thousand inhabitants (with the exception of Warsaw) – own calculations based on data located in the ESS repository.

[2] Response rate presented in a tabulated summary is the easiest method of estimating response rate. Other more precise methods include, for example, the percentage of successfully completed interviews within the whole randomly selected sample, diminished by events beyond the control of the researcher, such as death or a long-lasting illness of the respondent, incorrect address data, or a change of address of the respondent resulting in disqualifying them from the studied population (e.g. if they move to an area which is not covered by the study). With the exception of the aforementioned instances, all other reasons for failing to realize the interview have to be monitored as they diminish the representativeness of the sample. Such reasons may include: total lack of contact with the respondent, inaccessibility of the studied person within the time period of the study ,,exhausting the limit for contact attempts or the respondent's refusal to participate in the study. A brief review of several ways of calculating the value of response rate can be found in the publication "*Standard Definitions Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th edition*", published by The American Association for Public Opinion Research in 2011.

While analysing these results, it is important to first notice, that in the QoL 2008 project the response rate reached 47.1 % and was similar to the rate (registered in a comparable period of time) in the 4th round of ESS[12]. In the QoL_2008 program, the time devoted to fieldwork was shorter than in the case of ESS_R4_2008 – hence, most likely, the lower response rate value. No substantive differences are visible in the distribution of the number of visits that the pollsters needed to conduct the interviews, except for the 6th and 7th visits. Longer fieldwork period enables better access to "late respondents" (Stoop 2004: 38–48), and in ESS_R4_2008 it was possible to conduct more interviews during the 6th and 7th visit.

The main goal of this article is to examine the thesis that the more contact attempts reduces non-response bias: To verify the thesis, it would be best to conduct an experiment consisting in performing several parallel measurements with the same research tool: (a) based on equally numerous, (b) independent, (c) random samples of the same population, (d) differing only in the research strategy chosen, i.e. the maximum number of visits, set by the researcher, that a pollster can make to a particular respondent's address. The first research would only allow for one contact attempt. The next, respectively, two, three, four, five, six and seven contact attempts, depending on the maximum value accepted. Appropriate as it might seem, this strategy is too costly.

I propose a much more simple analysis model using information from just one research project. For the sake of the analysis, I divide the result set into seven inseparable subsets, each of them containing the answers of the same respondents, numbered according to the number of visits during which the interview took place. The first set contains the answers of the respondents of the first contact attempt – this set corresponds to the data that could be acquired from the strategy allowing one contact attempt only. The second subset additionally includes the measurement values of respondents of the second contact attempt – this set represents the data that would be acquired if the strategy allowing two attempts was applied. Up to the seventh set, which contains information from all the interviews – it is equivalent to the strategy giving the pollsters seven attempts to contact the respondents. This is an ad hoc research with no experimental design and the juxtaposition of data sets from the respective consecutive interview attempts leads to, *de facto*, dependent measurements. While I cannot say anything about the effectiveness of additional contact attempts, I can describe the result of what happened after each contact attempt.

For evaluating the effect of weighting on substantive findings, I used the variables listed in tables A1 and A2 in the Appendix. The first table contains the dichotomic nominal variables used in the QoL_2008 project – they will make it possible to establish non-response bias for percentages. The second table provides a list of the variables in the Likert's scale format. Each of the Likert's scales

(presented in Table A2 in Appendix) was measured by reference to a number of basic question items.

RESULTS

(Q1) Do more contact-attempts reduce dissimilarity between the sample and population distributions?
(Q2) Do more contact-attempts reduce the need for sample weighting?

Answers to the questions above are closely related. The smaller the disproportions between the distributions of selected socio-demographic features in the sample and the distributions of the "golden standards" in the population, the smaller the need for post-stratification weighting. This confirms and conforms to the idea of post-stratification weighting, the values of which increase together with the rising disproportions of the sample and population. The ability to make more contact attempts by the pollsters serves this very purpose: to give them a chance to reach the least accessible categories of respondents to increase the response rate and reduce non-response bias. Then, it might be legitimate to formulate the above questions in a slightly different way: (Q1a) To what extent do contact-attempts reduce the dissimilarity between the sample and population distributions? and (Q2a) To what extent do contact attempts reduce the need for sample weighting?

In order to answer the first question, I use the Index of Dissimilarity (ID), which is an index well known in methodological literature used to calculate the total difference between two distributions of the same variable. Originally, this index was introduced in the mid-fifties of the twentieth century in sociology and human geography as a measure of spatial and residential segregation of the racial populations in the USA. Since then, it has been widely used to measure inter-group income inequality, but it can also be used to summarize the differences between the pairs of marginal distributions (Mulekar et al. 2008: 2009). The value of ID is relatively simple to compute and is defined as:

$$ID = \frac{1}{2}\sum_{j=1}^{k}\left|p_j - \pi_j\right| \text{ (Kuha et al. 2011: 376)} \quad (7)$$

where:

$j = 1,..., k$ denotes the groups created after combining all the categories of sex, age and education post-stratification variables;

π_j for $j = 1,..., k$ denotes the distribution of „golden standards", that is the population proportions in every j-th category known in the whole population (of course $\sum_{j=1}^{k}\pi_j = 1$);

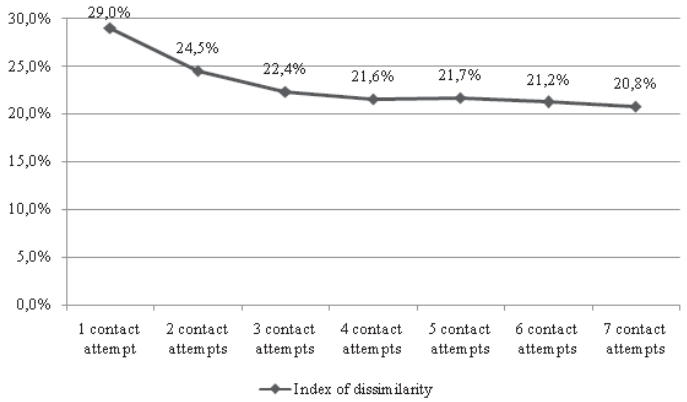
p_j for $j = 1,..., k$ denotes the respondent proportions in the obtained sample corresponding to every j-th category in the population, and $\sum_{j=1}^{k} p_j = 1$;

for each j = 1,..., k, post-stratification weights are calculated from the formula $w_j = \frac{\pi_j}{p_j}$.

The index of dissimilarity is intuitively straightforward to understand and has concrete meaning because it shows the percentage by which the distributions of two sets differ. The range of ID is [0,1]. Interpretation is as follows: The higher the value, the greater the disproportion, and if the structure of the sample corresponds to the structure of the population, the ID = 0.

Figure 1 shows how the index of dissimilarity, i.e., the differences between the population and the obtained sample, by consecutive pollster visits. The first value shows the deviation of the sample structure from the population structure for the interviews conducted during the first visit, the second value shows the same deviation for the interviews conducted during the first and second visits and so on, up to the final value, which characterizes the deviation between the sample and population for the whole set of results.

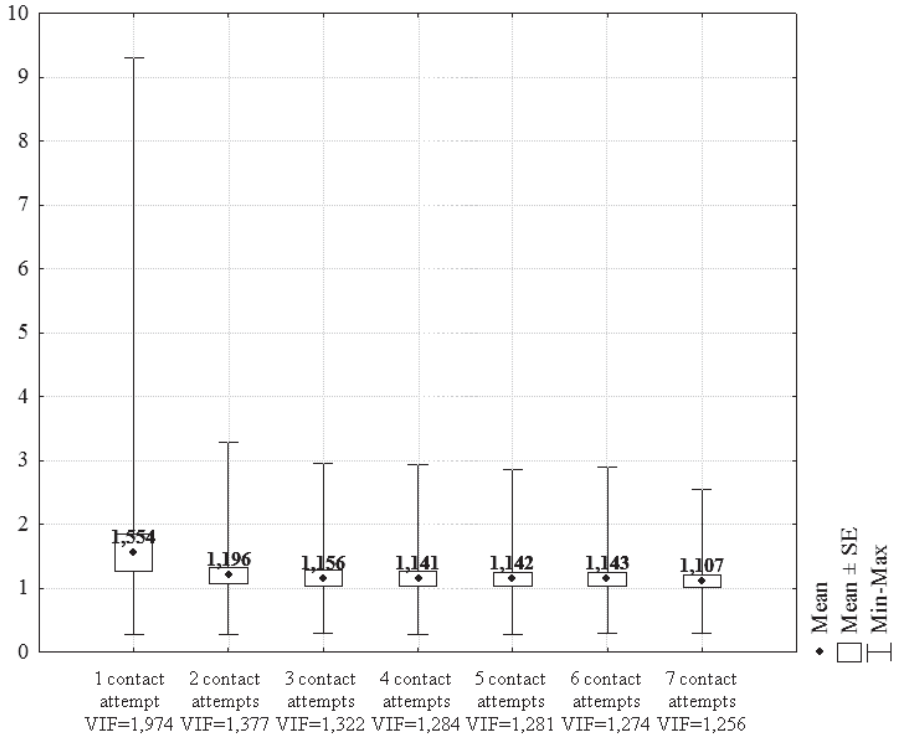
Figure 1 Index of dissimilarity – the differences between the population and the obtained sample



As confirmed by many other methodological analyses, (e.g. Stoop 2004: 38), with each additional visit the index value gets lower, i.e. the more the contact attempts, the more the obtained sample was similar to population. In terms of effectiveness, does the strategy of more contact attempts meaningfully reduce the gap between sample and population? For the interviews conducted during the first and second visit together, the ID level was reduced by 4.5% in relation to the ID level for the sample obtained after the first visit only. For those obtained in the

third visit it went down further by 2.1%. In total, the possibility of making seven attempts to conduct an interview gave an ID value reduction of less than 9% – from the level of 29% for those interviews conducted in the first visit to 20.8% for all the conducted interviews realized up to the seventh visit. In the sample obtained with two and three contact attempts, the adjustment of the socio-demographic respondent feature distribution in the sample to the corresponding distribution in the population was improved by more than 6.5%. However, for four, five, six and seven contact attempts, the ID level was reduced by only about 1.5%. Therefore, a significant reduction of the disproportions between the "golden standard" distribution in the population and the distribution of socio-demographic variables in the obtained sample can be observed *only during the first three visits*. Later on, the effect was hardly discernible.

Figure 2 The descriptive statistics for the post-stratification weights



Similar conclusions can be reached by analysing the post-stratification weights' values in the subsequent sets of results. Figure 2 presents Box Plots with descriptive statistics (minimum value, maximum value, arithmetic mean and the standard error

of the mean) for post-stratification weights. Additionally, information was given about the VIF value, which shows to what degree the post-stratification value increases the estimator variation and reduces the accuracy of the estimation.

Increasing the number of visits the pollsters can make in order to conduct an interview actually reduces the necessity for post-stratification weighting. Socio-demographic feature distributions in the samples obtained after each subsequent visit were more and more adjusted to the model distributions in the population, i.e. the ID values became lower and lower. In each subsequent set including cumulative measurement results, the influence of post-stratification weighting on the estimator variation became lower.

For the interviews conducted during the first visit, the most underspecified category consisted of men with secondary education from adjacent age cohorts aged 24–45. The disproportions in the distributions of the sample and population were so large that the first age category of men with secondary education, i.e. aged 24–35, the weight value of more than 9 had to be assigned, while the second one – aged 34–44 – the weight value of slightly less than 7. In each subsequent set there is a decrease in both the maximum post-stratification weight value, as well as their dispersion, calculated with the standard error of the arithmetic mean. While more contact attempts reduce the need for sample weighting, I am less firm with the answer to the question of the extent to which contact attempts reduce the need for sample weighting.

Making more contact attempts is, in many cases, "beneficial" to the quality of the research: with each subsequent visit the adjustment of the sample structure to the population improved and the necessity of post-stratification weighting was reduced. However, it is disputable whether these effects justify the costs – financial and otherwise – related to the extension of the fieldwork period. This dispute cannot be definitively made with these data, but the question is certainly raised.

The issue that I consider most important is whether more contact attempts significantly reduces non-response bias. Before I proceed to those issues, I examine more carefully the categories of respondents which are underrepresented and/or overrepresented in the sample in early and late contacts attempts.

(Q3) Which categories of respondents are underrepresented and/or overrepresented in the sample in early and late contacts-attempts?

I first define the variable which I will use in further analyses. The consecutive values will represent the number of visits during which the interview was successfully conducted with the respondent. In other words, its value will be 1 if the interview took place during the first visit, 2 if it happened during the second visit and so on, up to visit number 7. Many methodological studies use different information about the number of visits a pollster had to make to contact a particular

respondent as indexes of availability of the respondents or their readiness for cooperation with the pollster (Stoop 2004; Domański 2006). For example, the number of visits to a particular address that ended in an interview or its definitive non-realization i.e. the number of visits before the first contact with the respondent was made, could serve as the measure of availability. The number of visits from the first contact with the respondent to a successful interview is also used as a "cooperativeness" index.

While conducting regressive analyses, I wish to establish which socio-demographic categories of the respondents can be noticed during the first and later attempts to conduct an interview. I also want to learn which socio-demographic features characterize non-respondents and if they are similar to late-respondents. If the possibility of making multiple contact attempts serves any concrete purpose, it is to get hold of the opinions of the people who significantly differ from those who provide answers during the early contact attempts. If the distributions of the respondents' socio-demographic features and the distributions of answers to substantive questions were similar in the early and late contact attempts, the multiple returns of the pollsters to the selected individuals would not serve anything other than increasing the percentage of research sample realization, without having any influence on the reduction of non-response bias.

I expect that the socio-demographic profile of those respondents with whom the interviews were conducted during early contact attempts will be significantly different from the profiles of both (a) those who did not give the interview at all and (b) those who gave them after multiple returns of the pollsters to the drawn address. Stoop (2004: 23–54), in a follow-up survey done in 2000 one year after the Dutch wave of the Amenities and Services Utilization Survey, compares different types of respondents and non-respondents. In relation to the analyses I have conducted, there are two meaningful conclusions. In the first three contact attempts, young people, single people and those who have full-time jobs are more difficult to interview (Stoop 2004: 41). Next, the increased number of calls made the net sample more similar to the population (Stoop 2004: 38), which was confirmed also in my own research (see Figure 1). Very interesting research on Poland was conducted by Domański (2006: 29–49) and based on the data from the first round of the ESS research. While attempting to answer the question about the factors determining the availability of respondents, the author reaches the conclusion that it is easier to reach the people without a job. Yet, the factors reducing availability of the people drawn for a survey are high levels of education and income. It is more difficult to reach men and working people; however, in these two cases, the relations were not statistically relevant and it was not possible to generalize them to the whole population (Domański 2006: 38). The analyses in which the author tries to determine the factors increasing / decreasing the chances of conducting an

interview during the first visit and visits from 3 to 7 turn out to be very interesting as well. Regression analyses show that the chances of conducting an interview during the first visit depend only on the level of education (the lower the education, the higher the chance of getting the interview) and socio-occupational affiliations of the respondents (lower rank white-collar workers were more easily available). From the regression model that describes the chances of conducting an interview during visits 3–7, young people and the residents of cities were less available (i.e. more seldom reached during visits 1 and 2) (Domański 2006: 40–42).

Table 2 presents the results of logistic regression for three dichotomic dependent variables that were formed by the transformation of the "visit number" variable. In the first model, the dependent variable was established by means of assigning value 1 to all those respondents with whom the interview was conducted during the first visit. All the other respondents were assigned value 0. In the second model, the variable was established in which code 1 was assigned to all the respondents with whom the interview was done during visits 4 to 7. In the third model, the variable was established in which non-respondents were assigned value 1, respondents 0, no matter during which visit the interview took place. The predicator set is composed of socio-demographic characteristics of the respondents and the description of the neighbourhood in which they live. Of all the basic socio-demographic features, I took the following into consideration in my analyses: sex (men 1, women 0), age – in the form of six dichotomic categorical variables (24 and younger, 24–34, 35–44, 45–54, 55–64, more than 64), education defined by subsequent stages of education (dichotomic variables referring to elementary education, vocational secondary education, secondary education and higher education), income (divided into four dichotomic quartile groups, as well as a category for those refusing to answer the question about their income) and occupational status (five dichotomous categories distinguishing between working people, the unemployed, students, pensioners and housekeepers). The last six positions on the list of predicators are occupied by the characteristics of the neighbourhood, in which the individuals drawn for the sample live. During the first visit, the pollsters' noted (1) type of the residential building, coded into the form of three dichotomic variables (detached and semi-detached house, apartment in a block, apartment in a tenement building), (2) presence of an intercom (yes = 1), (3) presence of an alarm system (yes = 1), (4) condition of the building (bad = 1), (5) condition of the neighbourhood (bad = 1) and visible traces of vandalism (yes = 1).

Reference categories are (a) people older than 64, (b) with secondary education, (c) with income range from Q1 (First Quartile) to Me (Median), (d) working and (e) living in apartment buildings. Within the regression equation rate values for reference categories are not calculated, yet they are, of course, the basis for the interpretation of the relationship direction for the remaining complementary respondent categories[13].

Table 2 Logistic regression results: (1st model) early respondents, (2nd model) late respondents and (3rd model) non-respondents

Independent variables	Interview completed during the first visits – early respondents	Interview completed from 4th to the 7th visit – late respondents	Interview not completed – non-respondents
	1st model	2nd model	3rd model
Constants	0,197	-1,417	-0,595
Sex (Man = 1)	**-0,184***	0,001	**0,303****
Age:			
- 24 and less	-0,302	0,075	0,468
- 25–34	**-1,216****	**1,588****	**0,846****
- 35–44	**-1,045****	**1,473***	**0,888****
- 45–54	**-0,835***	**1,320***	**0,507***
- 55–64	-0,117	0,185	0,256
- more than 64 (*ref.*)	-	-	-
Education:			
- elementary	**0,531***	-0,166	
- vocational secondary	**0,477***	**-0,943****	
- secondary (*ref.*)	-	-	
- higher	-0,080	**0,498***	
Income:			
- Q1 and less	0,324	**-0,486***	
- Q1 – Me (*ref.*)	-	-	
- Me – Q3	-0,150	0,357	
- more than Q3	**-0,497***	**0,628***	
- refused to answer	-0,226	0,278	
Occupational status:			
- working (*ref.*)	-	-	
- unemployed	**0,908***	**-1,896****	
- students	0,458	-0,489	
- pensioners	0,028	-0,036	
- housekeepers	**0,736***	**-0,503***	
Type of building:			
- (semi-)detached house	**-0,560***	**1,225****	-0,001
- apartment building (*ref.*)	-	-	-
- tenement building	-0,501	-0,077	-0,255
Intercom (Yes = 1)	-0,228	0,361	0,112
Alarm system (Yes =1)	-0,022	**0,569***	**0,562****
Cond. of building (Bad = 1)	0,329	**-0,781***	**-0,372****
Neighborhood (Bad = 1)	-0,195	0,337	0,112
Traces of vandalism (Yes =1)	0,077	-0,366	0,143
R^2 Index	13,0%	18,3%	4,3%

* correlation significant at the 0,1 level

** correlation significant at the 0,05 level

The results in the table above show that the chances of conducting an interview during the first visit increase when the respondents are from the oldest age group, i.e. more than 64 years of age, from one of the two lowest education groups, or when they are unemployed or housewives. The chances are lower, when the person drawn is a man, a person aged between 25 and 54, working and classified to the category with the highest income. Little importance can be attributed to the characteristics of the neighbourhood; however, it is much more difficult to conduct an interview during the first visit with people living in detached houses. Results here are confirmed by other methodological studies. Taking a look at the data containing the information about the day and time of the first contact attempt, that the vast majority (almost 80%) of all the attempts took place from Monday to Friday, 50% of which before noon or in the early afternoon. Attempts were made during the usual working hours increases chances of getting the interview with women (including housewives), the unemployed and oldest respondents because they were home more often (Domański 2006: 38).

In later stages of interview realization those respondent categories that had been inaccessible during the first contact attempt become dominant (column 1 and column 2 in table 2). When persons of ages 25 to 54 were randomly selected for the sample, the chance for realizing the interview during the first visit was low, but it was these very categories that appeared much more frequently in interviews conducted in the period between the 4th and the 7th visit. Similarly, at later stages of interview realization, there was an increased chance of conducting the study with highest income persons, those with higher education, and occupationally active. Interestingly, while there was a greater chance of realizing the interviews with women during the first visit, gender was no longer a significant factor differentiating those chances during visits 4 through 7. These later respondents inhabited single and multi-family buildings and had alarm systems installed in their buildings. A different contact strategy was adopted by the pollsters for the interviews conducted during visits 4 through 7. Although most contact-attempts occurred in days from Monday to Friday, just as in case of the first attempt, an increased number of them was made in the evening (over 70%), unlike the first attempt, which was more often made either in the forenoon or early in the afternoon. Interestingly, similar relations between time and day of contact-attempts and the subsequent number of contact-attempts can be observed in works by other authors (Tancreto et al. 2005: 3630–3631).

Multiple returns of the pollsters to the respondent's address can be advantageous, as they allow to reach and obtain opinions from less accessible respondent categories that could not have been interviewed had the number of acceptable contact-attempts been limited. Even though it is consistent with the general consensus among methodologists dealing with the subject of non-respondents, this is an equivocal

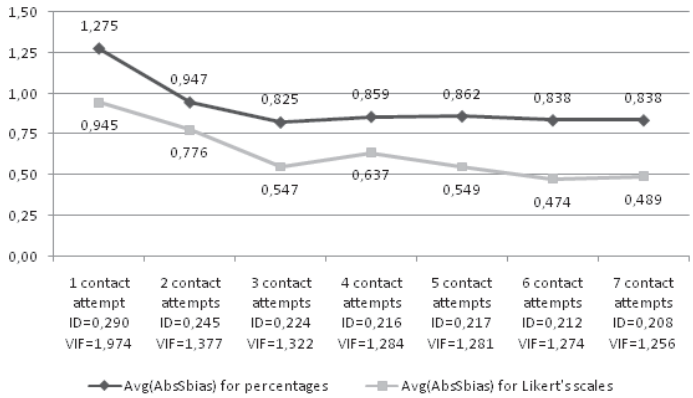
formulation. A considerable reduction of the index of dissimilarity level (and what follows a significant reduction in the need for post-stratification weighting) could be observed solely during the first three visits. The interviews conducted during visits 4 through 7 did not significantly facilitate the adjustment of socio-demographic characteristics distribution in the sample to the "golden standards" distribution in population. The option of making further contact-attempts was therefore advantageous, but not as much as it could be inferred from the results presented in table 2.

Coefficients in the regression formula for model 3, where the factors influencing chances of remaining in the non-respondent category have been specified, is interesting. Although the set of explanatory variables used here was narrower in comparison with the set of predicators employed in the two previous models (education, income and socio-occupational status could not have been determined for non-respondents), there are visible similarities between the structure of late respondents and non-respondents. In both cases persons of ages 25–54, living in single-family houses and in buildings protected with alarm systems, are dominant. The only difference is that there was no difference in gender for late-respondents, while men were prevalent among non-respondents. It could therefore be argued that even though certain less accessible respondents were interviewed during visits 4 through 7, the same categories remained underrepresented in the obtained sample. Multiple returns of the pollsters to the given addresses did not change this situation.

(Q4) What is the level of non-response bias in point estimation? Do more contact-attempts reduce differences between unweighted and weighted estimators?

Although sociologists are much more interested in relations between variables than the values of descriptive statistics in themselves, the question above is crucial for assessing the effectiveness of making multiple interview-attempts as a strategy meant to reduce the negative effects of non-response. Previous analyses demonstrated that in each subsequent set containing cumulative measurement results from interviews conducted during previous visits the adjustment between the realized sample and population structure was improved; therefore it should be expected that the level of non-response bias will be reduced. The data compiled in Figure 3 confirm this, as they present the average absolute standardized values of non-response bias (hereafter "Avg(AbsSbias)") for all percentages and the Likert scales used in the data analysis.

Figure 3 Avg(AbsSbias) for percentages and Likert's scales



If only the results presented above were to be considered, they show that more contact-attempts reduce differences between unweighted and weighted estimators and reduce non-response bias. A particularly significant reduction in non-response bias can be observed in interviews realized during the second and third visits, while those conducted during visits 4 through 7 did not bring any considerable changes. In order to precisely answer the question whether an increased number of contact-attempts limits non-response bias, one must closely examine the individual indicators, abandoning this homogeneous view of a heterogeneous situation. In other words, the average absolute standardized non-response bias coefficients set generally for arithmetic average estimators and structure indicators do not fully reflect the phenomenon under investigation, as the Avg(AbsSbias) value is a resultant of individual non-response bias levels set for a dozen or so indicators (in many cases unrelated to each other).

The best candidates for verifying the thesis stated here will be those indicators for which the non-response bias level exceeds critical values. There is no point in dealing in detail with variables for which there are no significant differences between estimators in the weighted and in the unweighted sample. For the purpose of this analysis, I selected those variables that were characterized by the highest levels of non-response bias. Detailed non-response bias values for the dozen or so variables are listed in the appendix in tables A.3. and A.4. The coefficients with the highest values of non-response bias among the structure indicators were: (1) CIA6.1 – percentage of persons declaring participation in Polish presidential elections (Avg(AbsSbias) value = 2.481), (2) EDU1.3 – percentage of households with children attending high school (Avg(AbsSbias) = 2.138),

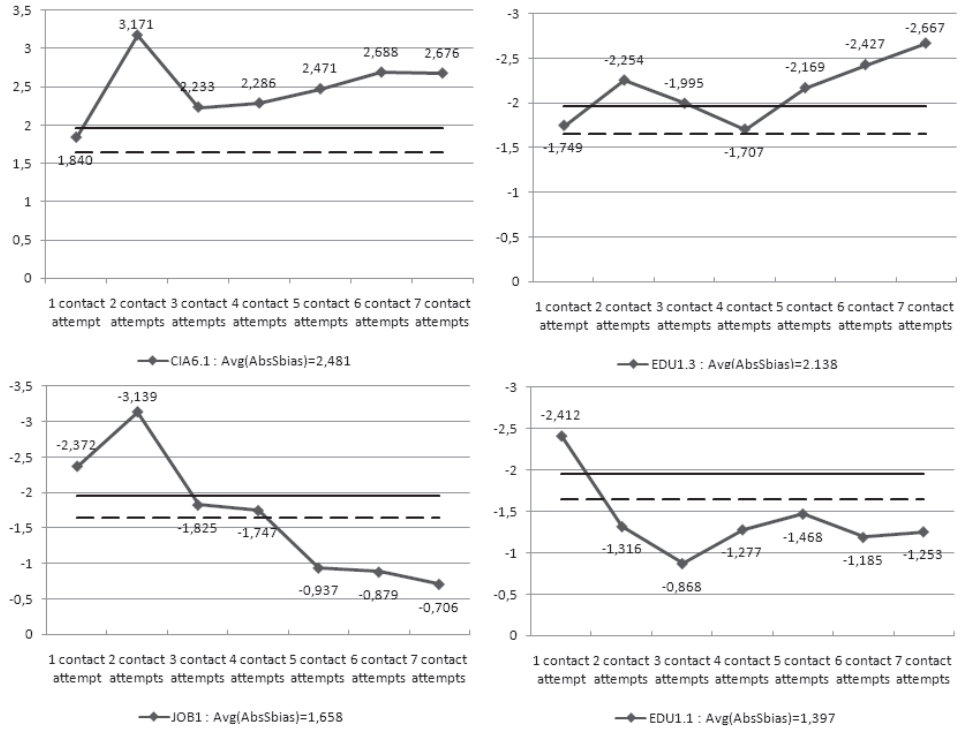
(3) JOB1 – percentage of persons occupationally active (Avg(AbsSbias) = 1.658) and (4) EDU1.1 – percentage of households with children attending elementary school (Avg(AbsSbias) = 1.397). Among Likert scales, the highest average non-response bias values could be attributed to variables: (1) WBI2 – life situation improvement prognosis (Avg(AbsSbias) = 1.772), (2) HEL4 – self-assessment of health satisfaction, (Avg(AbsSbias) = 1.312), (3) WBI5 – self-assessment of satisfaction with one's neighborhood (Avg(AbsSbias) = 1.202), as well as (4) CRI2 – self-assessment of one's sense of security in the evening and at night (Avg(AbsSbias) = 1.010).

First are some general observations. Only in two cases critical value was exceeded (with the confidence interval at 95%); after reducing the confidence interval to 90%, the number of indicators for which non-response bias was statistically significant rose to 3. In spite of the low response rate level and despite considerable disproportions noted in the distributions of post-stratification variables in the obtained sample as well as in population, response-bias did not pose any significant problem for the vast majority of indicators. Inaccessible individuals seem to have relatively little effect on the results of the study. However, a more systematic analysis of this phenomenon is necessary, with focus on a few specific variables.

The two graphs below illustrate standardized non-bias values (hereafter, "Sbias") – determined in accordance with formulas (5) and (6) – for subsequent measurement sets matching the results from interviews realized cumulatively from the first visit to the visit labelled with the ordinal number of a given set. Additionally, two horizontal lines represent the critical values for confidence intervals set at 95% (continuous line) and 90% (dotted line).

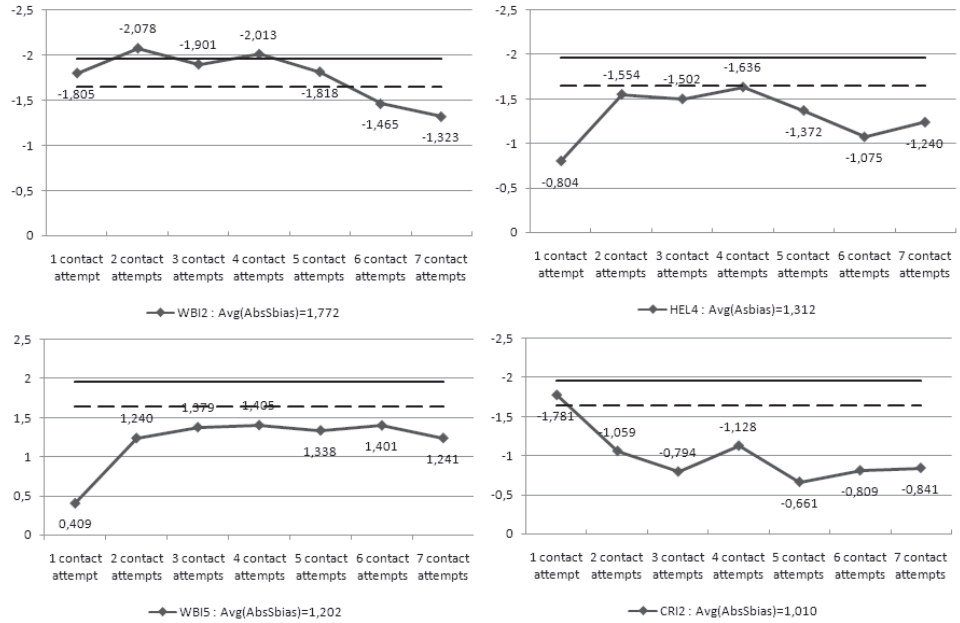
Persons participating in Polish presidential elections (CIA6.1) were overrepresented, while respondents occupationally active (JOB1) and those from households with children attending elementary school (EDU1.1) or high school (EDU1.3) were underrepresented. Participation in the Polish presidential elections was the variable with the highest non-response bias level among all analyzed variables. The overrepresentation of persons participating in elections is a result of the fact that one of the indicators of the respondent's readiness to cooperate with the pollster is his or her interest in public matters (Billiet et al. 2009: 17) and a sense of civic duty (Domański 2006: 34), which are also determining factors in election participation (CBOS 2010: 9). The underestimation of persons occupationally active was also predictable as this category of respondents is much more difficult to access for study. The underestimation of persons from households with children, attending elementary school or high school, can be attributed to the fact that persons of ages 33 to 54 (the ones that more typically have children at school age) form a less accessible respondent category.

Figure 4 Non-response bias (Sbias) for percentages



Independently from the number of visits intended to conduct the interviews the non-response bias level remained statistically significant. During subsequent visits, not only could the differences between the estimators values not be eliminated, but were greater in visits 4 through 7. At the same time, three contact-attempts were necessary to reduce non-response bias to levels below the critical value for the third presented variable (JOB.1), while the level of non-response bias for the EDU1.1 variable became statistically insignificant in the set of results of the interviews realized during the first two attempts at conducting them. For the remaining indicators, presented in Table A3 in the appendix of this study, non-response bias was not statistically significant, not even for a set limited to the first interview, or was reduced below the critical value after the second visit. Similar results can be observed for the Likert scales (Figure 5 & Table A4). The only problem was posed by the variable describing the life situation improvement prognosis (WBI2), for which the non-response bias level could be reduced only after adding measurement results from interviews conducted during the sixth contact-attempt.

Figure 5 Non-response bias (Sbias) for Likert's scales



The results for Likert scales presented on Figure 5 also demonstrate that values of the indicator for satisfaction with one's neighbourhood were overestimated in the obtained sample, while the three remaining variables were underestimated. The oldest persons were overrepresented in the obtained sample, while those from younger age categories prevailed among non-respondents, and thus the main explanation is the respondent age structure, specifically significant disproportions in respondent age distribution between the sample and population. The age of studied persons remained strongly correlated with all the above variables; there is a directly proportional relationship between age and the assessment of one's neighbourhood and an inversely proportional relationship between age and life situation improvement prognosis, assessment of one's health and assessment of one's sense of security. Thus, the values of non-response bias can be attributed to these correlations.

CONCLUSION

My primary goal has been to verify the thesis that enabling the pollsters to make multiple interview-attempts reduces negative consequences stemming from inaccessibility of some of the study's randomly selected persons . I addressed four questions related to this topic.

1. Do more contact-attempts reduce dissimilarity between the sample and population distributions?

The answer is affirmative. Specifically, in each subsequent set containing cumulative measurement results from interviews conducted from the first visit to the visit designated with the number matching the given subset, the distributions of post-stratification variables in the obtained samples were increasingly more similar to the distribution of "golden standards" in population. Yet, the analyses have demonstrated that the significant reduction of disproportions between these distributions could only be observed during the first three visits, as the effect was imperceptible later.

2. Do more contact-attempts reduce the need for post-stratification weighting?
The answer is yes, taking into account how I defined and set post-stratification weights. The assertions presented in point 1 apply here as well, i.e. during the visits 1 through 3 the maximal values of post-stratification weights and their dispersion are clearly reduced. The effect of weighting on the precision of estimating descriptive statistics is also increasingly diminished. Concurrently, interviews conducted during visits 4 through 7 do not bring any significant changes to the values of post-stratification weights.

3. Which categories of respondents are underrepresented and/or overrepresented in the sample in early and late contact-attempts?

Multiple interview-attempts allow us to reach less accessible respondents and obtain their opinions. These respondent categories would not have been available for study had the number of possible contact-attempts been limited. The same categories which were dominant in the late respondents group remained unavailable, as they were clearly dominant in the non-respondent category. Multiple returns of the pollsters to the given addresses did not change this situation.

4. What is the level of non-response bias in point estimation? Do more contact-attempts reduce differences between unweighted and weighted estimators?

The main advantage of increasing the number of allowed interview attempts is the growth of the response rate. Concurrently, this strategy is not as clearly and unambiguously advantageous when it comes to reducing the non-response bias level. However, the effect of the non-response bias on the values of descriptive statistics is only a problem for certain types of variables and in most cases it does not exert influence on point estimator values. The limited effect of non-response bias on the descriptive statistics values occurs even at low response rate levels. Interestingly, it is also independent from the number of interview-attempts. On the other hand, there occur such types of variables for which the researcher is not able to reduce the level of non-response bias below critical values, even if pollsters

are allowed to make multiple attempts at contacting the selected respondents. Moreover, in certain situations, the level of error may be increased.

The conclusion is therefore quite surprising, as the increased number of visits does not unambiguously affect reducing non-response bias, although it remains a relatively simple method of improving the response rate.

NOTES

1 A review of several most frequently used procedures for assessing non-response bias can be found in Billiet at. al (2009) who describes Groves's (2006) approaches to non-response bias. One of these methods, based on comparing response bias estimates with similar estimates from other more accurate sources (e.g. official population statistics), in which bias is understood as the amount of deviation between the 'true' population distributions and the distributions in the obtained sample (Billietat al. 2009: 7), has been used for measuring non-response bias in the analyses presented in this article.

2 In his famous article *Nonresponse Rates and Nonresponse Bias in Household Surveys,* Groves (2006) describes five models explaining the relationships between response rate and non-response bias: (1) *Separate Causes Model*, (2) *Common Cause Model*, (3) *Survey Variable Cause Model*, (4) *Nonresponse Measurement Error Model,* and (5) *Nonresponse Error Attenuation Model*. He concludes his deliberation with the following statement: „*To summarize […], there are three contrasting features of the causal models. First, they differ in the nature of the causal relationships involving* [response propensity (*p*)] *and* [the reported survey variable (*y*)] *that are producing the covariance between p and y*.[…] *Second, the last two models incorporate possibilities of measurement errors affecting estimated nonresponse biases. Third, the models pose very different challenges for reducing nonresponse bias in postsurvey adjustment.*"

3 Sztabinski et. al. (2007) performed comparative analyses for (a) respondents participating in the Pilot Study of ESS and the Main Study of ESS Round 2 as well as for (b) non-respondents who returned the follow-up questionnaire. The results of these analyses are ambiguous, both in terms of socio-demographic descriptions of the studied persons and in terms of the opinions expressed by respondents from both comparative categories. In terms of socio-demographic characteristics of the studied persons there were significant diversities related to variables such as education level, family income per capita, domicile and main activity; they were not, however, significant in terms of variables describing sex, age, and household size. In relation to changing attitudes and opinions, significant diversity was noted in terms of trust in others, satisfaction with democracy and mood assessment, while interests in politics and the role of women in family exhibited insignificant diversity. Similarly ambiguous conclusions were reached by Domański (1999), who noted that non-respondents differed from respondents – the differences were reflected in the comparison of several dozen variable distributions, although they had no influence on the obtained interrelations between variables.

4 Stoop (2005), in the second chapter of his book *The Hunt for the Last Respondent*, demonstrates that one of the main reasons why researchers aim at achieving high response rate levels is that this value is often treated as a criterion of the research team's professionalism. In many cases pursuing low non-response rate levels is motivated by business issues (the survey response rate is viewed as an indicator of the entire survey process) rather than methodological ones (reducing non-response bias).

5 The problems faced by researchers conducting studies in large cities are well exemplified by analyses performed by Grześkiewicz-Radulska (2009), on the basis of CBOS data from nationwide Polish public opinion surveys in years 1993–2004. The author compares four return rates measures: response rate (RR), contacts rate (CON), refuse rate (REF), and cooperation rate (COOP), depending on the type and size of the locality. From the comparisons performed one can infer that during the whole analyzed time period (1993–2004) the values of RR and COOP coefficients for cities with 500 thousand inhabitants were clearly at the lowest levels (the highest values for both measures were observed in rural areas). Moreover, even though RR and COOP values were being systematically reduced irrespective of the types and sizes of localities, it was in cities with more than 500 thousand inhabitants that this reduction was most significant. Values related to the REF coefficient were similar, but they differed in the respect that the largest refuse rates could be observed in the largest Polish cities (the smallest occurred in rural areas) with the highest dynamics of refusal increase in time. Although a slightly lower diversity level in particular locality classes were observed within the scope of the CON coefficient, the obtained values were still lowest for cities with more than 500 thousand inhabitants. The described relations between return rates measures and locality types are universal in character and have been well documented in methodological literature. For instance, Goyder at al. (1992) tested urbanization effects on survey non-response in three Canadian cities. The analyses they performed demonstrated that the main variation of non-response rate was in the contrast between all large cities and intermediate, small or rural areas.

6 A particularly promising approach to this matter was presented by Lynn (2003), who described the PEDAKSI methodology (Pre-Emptive Doorstep Administration of Key Survey Items) for the collection of key survey data from non-respondents to personal interview surveys and subsequent assessment of non-response bias. The PEDAKSI methodology combines data obtained directly from the respondent with supplementary data from the sampling frame and public statistical resources. Implementation of the PEDAKSI methodology is conducted in the following manner: after contacting the respondent, the pollster subjectively assesses the likelihood of conducting the interview with the given respondent. If, according to the pollster, there is little chance of further cooperation, they should ask the selected person a few key questions from the so-called Key Items Form Questionnaire (KIF) in the preliminary stage of arranging the interview. Lynn's experiment demonstrates that employing the KIF questionnaire even in situations in which the primary study was realized does not significantly raise costs. Neither has it any negative consequences on realizing the main part of the study even though some of the respondents might consider answering the primary questions unjustified, since they have already answered them briefly in the KIF questionnaire. At the same time, the questionnaire enables the pollster to obtain information from inaccessible individuals, which might have been lost otherwise. It should be noted, however, that the PEDAKSI methodology is based on several key assumptions, without which its usage would be groundless. Firstly, that the KIF questionnaire is be successfully completed by a significant number of inaccessible individuals, which will make it possible to estimate the results for the whole group of individuals unavailable for the study. Secondly, that the inaccessible individuals answering the KIF questionnaire are representative of all inaccessible individuals in the sample (PEDAKSI provides procedures which allow one to compare both groups of inaccessible individuals based on auxiliary variables of public statistics and pollster observation), and finally, that the measurement performed with the use of KIF questionnaire remains accurate and reliable. Lynn's implementation of the

PEDAKSI methodology as part of the British Crime Survey yielded promising results and appears to be more useful than traditional *follow-up* studies.

7 Statistical errors, at least theoretically, may not occur at all. However, exactly as in the case of random errors, such an assumption would be naive and their occurrence should be accepted.

8 The methodologically accepted methods of defining non-response rate are discussed in a further section of this article.

9 A distinction into three missing data mechanisms, which concern the relation between the missingness process and the outcome variable, can also be found in literature. The distinction is made between missing completely at random (MCAR), missing at random (MAR), and the missing not at random (MNAR). For details see For details see Akacha et al. (2010: 1–2), Wood et al. (2004: 525–526).

10 Research return rate sheets were used primarily to monitor the pollsters' work and the interaction between the pollster and the randomly selected person, from the interview arrangement phase up to the definitive end of respondent contact. Each pollster was obliged to present a report from each visit, containing: 1) a description of the interview arrangement phase, 2) a description of the encountered situation in case of incorrect address data, 3) the result of the visit (whether the interview was completed, interrupted, or unrealized, or whether there was any contact with persons other than the respondent), 4) reasons for failing to conduct the interview, 5) a description of the reasons for refusal, 6) a description of the respondent's neighborhood. All this information was meant to facilitate the coordination of the study, particularly by helping to assess the chance of conducting interviews with persons with whom the interviews had previously failed to be realized. Indications that the respondents were only temporarily unavailable and information concerning the reasons for their refusal were particularly important. If the refusal was not categorical (i.e. eliminating the possibility of further contact), the task of conducting the interview was passed on to another pollster with more experience.

11 http://ess.nsd.uib.no/streamer/?module=main&year=2009&country=&download=%5CFieldwork+documents %5C2009%5C07%23ESS4+-+ Contact+forms%5C.%5CESS4Source_Contactforms.pdf

12 ESS studies indicate a continuous decrease in response rate in large cities. For instance, in 2001, the response rate was approximately 55% (Sztabiński 2001: 117–121).

13 For example, in the first of the presented models, negative values for the variable describing income above the third quartile (persons with the highest income) mean that the chance of successfully completing interviews with persons from this category during the first visit is significantly lower than in case of the reference category (i.e. persons with income in between the lower quartile and the median). Positive values for regression coefficients are analogically interpreted as indication of a greater chance of successfully completing interviews in a given respondent category than in the reference category.

REFERENCES

Akacha, Mouna and Jane L. Hutton. 2010. 'Modelling the Rate of Change in a Longitudinal Study With Missing Data, Adjusting for Contact Attempts.' *CRiSM Working Paper* No. 10–08. University of Warwick. Centre for Research in Statistical Methodology: Coventry.

Biemer, Paul P. 2010. 'Total Survey Error. Design, Implementation, and Evaluation.' *Public Opinion Quarterly* 74(5): 817–848.

Billiet, Jaak, Hideko Matsuo, Koen Beullens and VasjaVehovar. 2009. 'Nonresponse Bias in Cross National Surveys: Designs for Detection and Adjustment in the ESS.' *ASK. Research & Methods* Vol. 18(1): 3–43.

CBOS – Public Opinion Research Centre. Report no. BS/110/2010. 'After the Presidential Election.' Warszawa.

Curtin, Richard, Stanley Presser and Eleanor Singer. 2005. 'Changes in Telephone Survey Nonresponse Over The Past Quarter Century.' *Public Opinion Quarterly* 69(1): 87–98.

Domański, Henryk. 1999. 'Non-available Persons. How it Affects Results of Survey Research.' *ASK. Research & Methods* Vol. 8: 67–92.

Domański, Henryk. 2006. 'Number of Visits and Duration of Surveys.' *ASK. Research & Methods* Vol. 15: 29–49.

Fricker, Scott and Roger Tourangeau. 2010. 'Examining the Relationship between Nonresponse Propensity and Data Quality in Two National Household Surveys.' *Public Opinion Quarterly* 74(5): 934–955.

Goyder, John, Jean Lock and Trish McNair. 1992. 'Urbanization Effects on Survey Nonresponse: A Test Within and Across Cities.' *Quality & Quantity* 26: 39–48.

Groves, Robert M. 2006. 'Nonresponse Rates and Nonresponse Bias in Household Surveys.' *Public Opinion Quarterly* 70(5): 646–675.

Grzeszkiewicz-Radulska, Katarzyna. 2009. 'Respondenci niedostępni w badaniach sondażowych.' In: *Analizy i próby technik badawczych w socjologii*. Tom XII. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Jabkowski, Piotr. 2007. 'The Impact of not Carrying out Some Part of Interviews on the Validity of Statistical Inference in Social Research. PAPI and CATI Surveys in View of Systematic Errors.' *ASK. Research & Methods* Vol. 16: 67–86.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best and Peyton Craighill. 2006. 'Gauging the Impact of Growing Nonresponse on Estimates from a National Telephone Survey.' *Public Opinion Quarterly* 70(5): 759–779.

Kuha, Jouni and David Firth. 2011. 'On the Index of Dissimilarity for Lack of Fit in Loglinear and Log-multiplicative Models.' *Computational Statistics and Data Analysis* 55(1): 375–388.

Lissowski, Grzegorz. 1971. 'Problem jednostek niedostępnych w reprezentacyjnych badaniach socjologicznych.' In: *Metody matematyczne w socjologii – zagadnienia wybrane*, edited by K. Szaniawski. Warszawa: IFiS PAN.

Lynn, Peter, Paul Clarke, Jean Martin and Patrick Sturgis. 2002. 'The Effects of Extended Interviewer Efforts on Nonresponse Bias.' In: *Survey Nonresponse*, edited by R. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. New York: Willey.

Lynn, Peter. 2003. 'PEDAKSI: Methodology for Collecting Data about Survey Non-Respondents.' *Quality & Quantity* 37(3): 239–261.

Mulekar, Madhuri S., John C. Knutson and Jyoti A. Champanerkar. 2008. 'How Useful are Approximations to Mean and Variance of the Index of Dissimilarity?' *Computational Statistics & Data Analysis* 52(4): 2098–2109.

Sawiński, Zbigniew. 2005. '*Metody doboru respondentów*.' In: *Fieldwork jest sztuką. Jak dobrać respondenta, skłonić do udziału w wywiadzie, rzetelnie i sprawnie zrealizować badanie,* edited by P.B. Sztabiński, Z. Sawiński and F. Sztabiński. Warszawa: Wydawnictwo IFiS PAN.

Stoop, Ineke.2004. 'Surveying Nonrespondents.' *Field Methods* 16(1): 23–54.
Stoop, Ineke. 2005. 'The hunt for the Last Respondent. Nonresponse in Sample Surveys.' The Hague: Social and Cultural Planning Office.
Sztabiński, Paweł B. 2001. 'European Social Survey.' *ASK. Research & Methods* Vol. 10: 117–121.
Sztabiński, Paweł B., Franciszek Sztabiński and Dariusz Przybysz. 2007. 'Are Non-Respondents Similar to Respondents. Findings from the ESS-2004 in Poland.' *ASK. Research & Methods* Vol. 16: 25–54.
Sztabiński, Paweł B., Franciszek Sztabiński and Dariusz Przybysz. 2009. 'How Does Length of Fieldwork Period Influence Non-Response? Findings from ESS 2 in Poland.' *ASK. Research & Methods* Vol. 18(1): 67–95.
Tancreto, Jennifer G. and Michaerl Bentley. 2005. 'Determining the Effectiveness of Multiple Nonresponse Followup Contact Attempts on Response and Data Quality.' *American Statistical Association Section on Survey Research Methods*: 3626–3632.
The American Association for Public Opinion Research. 2011. Standard Definitions: FinalDispositions of Case Codes and OutcomeRates for Surveys. 7th edition. AAPOR.
Vehovar, Vasja, Zenel Bategelj, Katja Lozar Manfreda and Metka Zaletel. 2002. 'Nonresponse in Web Surveys.' Pp. 229–242. In: *Survey Nonresponse*, edited by R. Groves, D.A. Dillman, J.L. Eltinge and R.J.A. Little. New York: Willey.
Wood, Angela M. and Ian R. White. 2004. 'Using Number of Failed Contact Attempts to Adjust for Non-ignorable Non-response.' *Journal of the Royal Statistical Society. Series A*. 169(3): 525–542.

Piotr Jabkowski is an assistant professor at the Institute of Sociology of Adam Mickiewicz University in Poznań. After obtaining his masters degree in applied mathematics at the Faculty of Mathematics and Computer Science of Adam Mickiewicz University in Poznań (2001), he began his Ph.D. studies at the Faculty of Social Sciences of Adam Mickiewicz University in Poznań and in 2006 he obtained his Ph.D. degree in sociology. He has been associated with the Center For Quality of Life Studies - a university unit founded to conduct social studies. His research interests encompass problems related to the quality of data obtained in survey studies, including unit non-response issues.

Institute of Sociology, Adam Mickiewicz University in Poznan,
Szamarzewskiego 89c, PL 60-568 Poznań, Poland
E-mail: pjabko@amu.edu.pl

APPENDIX

Table A1 List of variables – the percentages

Variable name	Variable description
CIA1.1.	Percentage of persons who reported watching local television stations
CIA6.1.	Percentage of persons who reported participating in Polish presidential election in 2005
CIA10.1.	Percentage of persons who reported membership in a social organization or association
CIA10.2.	Percentage of persons who reported membership in a non-governmental organization
CIA11.4.	Percentage of persons who reported giving 1% of their income tax to a chosen Public Benefit Organization
CIA11.5.	Percentage of persons who reported offering financial support to a chosen charity organization
HEL3.1.	Percentage of persons who reported health-related difficulties in fulfilling their school, work or household obligations
HEL3.3.	Percentage of persons who reported health-related difficulties in interpersonal contacts with their relatives
HEL5.1.	Percentage of persons who reported physical exhaustion and fatigue
HEL5.2.	Percentage of persons who reported painful ailments
SOC4.4.	Percentage of persons who reported income-related difficulties in fulfilling their school, work or household obligations
SOC4.5.	Percentage of persons who reported income-related difficulties in interpersonal contacts with their relatives.
CRI3.1.	Percentage of persons who reported victimization experiences
JOB.1.	Percentage of occupationally active persons
EDU1.1.	Percentage of households with children attending elementary school
EDU1.2.	Percentage of households with children attending middle school

Table A2 List of variable – Likert scales

Variable name	Variable description
WBI1	Assessment of current life situation
WBI2	Life situation improvement prognosis
WBI4	Assessment of satisfaction with one's apartment
WBI5	Assessment of one's neighborhood
CIA5	Assessment of the actions of the Mayor of Poznań
CIA9	Readiness to engage in civic activities
HEL4	Self-assessment of health status
SOC1	Self-assessment of financial capability
SOC2	Assessment of fund management
SOC5	Readiness to engage in activities aimed at helping the poor and the needy
CRI2	Assessment of one's sense of security in the evening and at night
ENV1	Assessment of air purity in the city
ENV2	Assessment of noise level in the city
ENV3	Assessment of order and cleanliness in the city
ENV4	Assessment of water reservoir and lake purity in the city
ENV6	Readiness to engage in activities aimed at protecting the environment

Table A3 Percentage values bias (Sbias) due to disproportions in sample and population structure

Variable name	1 visit (37,1%)	2 visits (65,5%)	3 visits (83,2%)	4 visits (92,5%)	5 visits (97,2%)	6 visits (98,6%)	7 visits (100%)	*Avg(AbsSbias)*
CIA1.1	1,733	0,780	0,995	0,813	0,639	0,457	0,562	**0,854**
CIA6.1	1,840	3,171	2,233	2,286	2,471	2,688	2,676	**2,481**
CIA10.1	1,289	-0,569	-0,137	-0,300	-0,402	-0,506	-0,437	**0,520**
CIA10.2	2,028	0,057	0,038	0,088	0,048	0,109	-0,199	**0,367**
CIA11.4	0,346	0,265	0,76	0,958	1,481	1,374	0,886	**0,867**
CIA11.5	0,855	0,053	0,565	0,665	0,978	0,955	0,871	**0,706**
HEL3.1	0,596	0,174	0,32	0,399	0,09	0,071	0,091	**0,249**
HEL3.3	-1,163	-0,633	-0,094	-0,202	-0,576	-0,558	-0,485	**0,530**
HEL5.1	0,413	0,44	0,761	0,881	0,382	0,136	0,507	**0,503**
HEL5.2	0,964	1,201	1,559	1,742	1,459	1,081	1,343	**1,336**
SOC4.4	-1,334	-0,230	0,072	0,112	0,423	0,431	0,367	**0,424**
SOC4.5	0,794	0,997	0,516	0,474	0,509	0,567	0,42	**0,611**
CRI3.1	-0,924	0,063	-0,443	-0,189	0,043	0,162	0,064	**0,270**
JOB1	-2,372	-3,139	-1,825	-1,747	-0,937	-0,879	-0,706	**1,658**
EDU1.1	-2,412	-1,316	-0,868	-1,277	-1,468	-1,185	-1,253	**1,397**
EDU1.2	-0,863	-0,761	-0,851	-0,758	-0,584	-0,667	-0,715	**0,743**
EDU1.3	-1,749	-2,254	-1,995	-1,707	-2,169	-2,427	-2,667	**2,138**
Avg(AbsSbias)	**1,275**	**0,947**	**0,825**	**0,859**	**0,862**	**0,838**	**0,838**	-

Table A4 Likert scales bias (Sbias) due to disproportions in sample and population structure

Variable name	**1 visit** (37,1%)	**2 visits** (65,5%)	**3 visits** (83,2%)	**4 visits** (92,5%)	**5 visits** (97,2%)	**6 visits** (98,6%)	**7 visits** (100%)	*Avg(AbsSbias)*
WBI1	0,740	0,262	0,021	0,058	0,217	0,412	0,108	0,260
WBI2	-1,805	-2,078	-1,901	-2,013	-1,818	-1,465	-1,323	1,772
WBI4	0,867	0,989	0,411	0,581	0,526	0,603	0,632	0,658
WBI5	0,409	1,240	1,379	1,405	1,338	1,401	1,241	1,202
CIA5	-0,342	-0,886	-0,032	-0,181	-0,148	-0,047	-0,240	0,268
CIA9	2,227	0,540	-0,028	-0,263	-0,205	-0,063	-0,027	0,479
HEL4	-0,804	-1,554	-1,502	-1,636	-1,372	-1,075	-1,240	1,312
SOC1	-0,663	-0,765	-0,403	-0,598	-0,295	-0,168	-0,263	0,451
SOC2	0,800	0,306	0,306	-0,460	0,312	0,212	0,048	0,349
SOC5	-0,118	-0,290	-0,191	-0,349	-0,454	-0,251	-0,240	0,270
CRI2	-1,781	-1,059	-0,794	-1,128	-0,661	-0,809	-0,841	1,010
ENV1	-1,108	-0,995	-0,263	-0,067	0,178	-0,036	-0,220	0,410
ENV2	-0,205	-0,193	-0,444	-0,524	-0,293	-0,171	-0,568	0,343
ENV3	-1,460	-0,685	-0,332	-0,318	-0,120	-0,470	-0,459	0,549
ENV4	0,953	0,289	0,085	0,039	0,304	0,084	0,091	0,264
ENV6	-0,842	-0,283	-0,655	-0,572	-0,547	-0,314	-0,287	0,500
Avg(AbsSbias)	**0,945**	**0,776**	**0,547**	**0,637**	**0,549**	**0,474**	**0,489**	-