International Statistical Review
Revue Internationale de Statistique

International Statistical Review (2012), 80, 3, 382–399 doi:10.1111/j.1751-5823.2012.00189.x

Evaluating, Comparing, Monitoring, and Improving Representativeness of Survey Response Through R-Indicators and Partial R-Indicators

Barry Schouten,[1] Jelke Bethlehem,[1] Koen Beullens,[2] Øyvin Kleven,[3] Geert Loosveldt,[2] Annemieke Luiten,[1] Katja Rutar,[4] Natalie Shlomo[5] and Chris Skinner[5]

[1]*Statistics Netherlands, Den Haag, The Netherlands*
[2]*Katholieke Universiteit Leuven, Leuven, Belgium*
[3]*Statistics Norway, Oslo, Norway*
[4]*Statistical Office of the Republic of Slovenia, Ljubljana, Slovenia*
[5]*University of Southampton, Southampton, United Kingdom*
E-mail: jg.schouten@cbs.nl

Summary

Non-response is a common source of error in many surveys. Because surveys often are costly instruments, quality-cost trade-offs play a continuing role in the design and analysis of surveys. The advances of telephone, computers, and Internet all had and still have considerable impact on the design of surveys. Recently, a strong focus on methods for survey data collection monitoring and tailoring has emerged as a new paradigm to efficiently reduce non-response error. Paradata and adaptive survey designs are key words in these new developments. Prerequisites to evaluating, comparing, monitoring, and improving quality of survey response are a conceptual framework for representative survey response, indicators to measure deviations thereof, and indicators to identify subpopulations that need increased effort. In this paper, we present an overview of representativeness indicators or *R*-indicators that are fit for these purposes. We give several examples and provide guidelines for their use in practice.

Key words: Adaptive survey design; non-response; non-response reduction; paradata; representativity.

1 Introduction

Non-response is one of the most widely recognized sources of error in surveys. It has been investigated extensively in the literature and features importantly in quality–cost trade-offs in the design of surveys (Groves, 1989; Groves *et al*., 2002; Biemer & Lyberg, 2003). Because surveys, especially those administered by interviewers, are expensive tools, survey institutes constantly seek a balance between non-response error and costs of survey data collection. In recent literature, a shift of focus has taken place from non-response analysis and non-response adjustment to non-response monitoring, and tailoring of data collection (Groves & Heeringa, 2006; Kreuter, *et al.* 2010).

The new paradigm of survey data collection monitoring and tailoring has led to a growing interest in quality indicators for evaluating and monitoring response and in sources of data for constructing such indicators. In particular, there has been increased interest in the use of data about the data collection process and observations on sample cases additional to those from the questionnaire. These data are usually referred to as paradata. We refer to Groves *et al.* (2008), Schouten *et al.* (2009), and Wagner (2010).

Indicators for evaluating non-response error serve four purposes:

(1) To compare response between different surveys that share the same target population, for example, households or businesses.
(2) To compare the response to a survey longitudinally, for example, monthly, quarterly, or annually.
(3) To monitor the response to a survey during data collection, for example, after various days, weeks, or months of fieldwork.
(4) To adapt the data collection design by tailoring based on historic data, and available frame data and paradata.

This paper provides an overview of *R*-indicators and partial *R*-indicators that were designed to be fit for these purposes. They were introduced by Schouten *et al.* (2009) and Schouten *et al.* (2011). The indicators are based on measures of variability in response propensities. In this paper, we convey the usefulness of the indicators through their application across a range of examples. We bring together results from various technical papers, provide simple guidelines on their use in practical survey settings and discuss their strengths and weaknesses.

As motivation, we consider survey settings where four properties of the indicators are required. The indicators should be easy to interpret, they must be based on available auxiliary data and survey data only, they should be relevant or in other words lead to effective survey designs, and they should allow for analysis at different levels of detail. The last property is especially important when many auxiliary variables are available and the number of possible indicators increases very rapidly.

Alternative indicators exist. The most well established non-response quality indicator is the response rate. This has the advantage of simplicity and ease of calculation, but suffers from often having only a limited relation to non-response bias (e.g. Groves, 2006; Groves & Peytcheva, 2008). The latter might be taken to be the ideal measure of non-response error. However, it is rarely measurable directly and, moreover, most surveys are designed to produce a large number of survey estimates and the corresponding number of non-response biases might be too great to serve many needs of quality indicators, for example, between-survey comparisons. Nonetheless, indicators that focus on the impact of non-response on specific survey statistics have been proposed recently, see Wagner (2010) and Andridge & Little (2011). These indicators are especially useful when surveys have only a few key substantive variables.

For monitoring and improving survey response, it is not sufficient to have overall measures of quality. Indicators need to be detailed in order to find subgroups that affect representativeness of response. The indicators that currently come closest to such indicators are subgroup response rates, for example, the response rates for different regions of the country or different types of businesses. There are three main drawbacks to using subgroup response rates in monitoring and targeting non-response. First, subgroup response rates do not depend on the size of the subgroup, that is, small subgroups may appear equally important as large subgroups. Second, subgroup response rates cannot be given at the variable level. As a consequence different variables cannot be evaluated and compared in their impact on response. Third, subgroup response rates

are univariate and do not allow for conditioning on other variables in an easy way. There is, therefore, a need for other quality indicators to supplement their use.

The indicators discussed in this paper possess the four properties as they are linked to an easy to interpret definition of representative response and allow for different levels of detail. In the various examples we have selected, we show how these properties translate to survey practice.

A few remarks are in place. The indicators are not designed to be tools in non-response adjustment. The indicators do evaluate the extent to which survey researchers need to rely on usual assumptions made under such methods, but they are neither related to specific adjustment methods nor to specific survey variables. Indicators that are specifically designed to support selection of auxiliary variables in non-response adjustment were proposed by Schouten (2007) and Särndal & Lundström (2010), and discussed by Särndal (2011). The indicators presented here evaluate representativeness relative to a set of auxiliary variables as we will explain. The auxiliary variables may consist of frame data, registry data, and paradata. Furthermore, the indicators, just like any population statistic, have a precision that is dependent on the size of the survey sample. Small surveys, therefore, do not allow for strong conclusions about representativeness of response and cannot be tailored without considering surveys similar in design.

In Section 2, we define representative response and explain how to measure deviations from it. In sections 3–6, we discuss and illustrate the various uses of the indicators. We end with an extended discussion in section 7 on the strengths and weaknesses of the indicators and the relation to non-response adjustment.

2 *R*-Indicators and Partial *R*-Indicators

The indicators that we propose build on two definitions: representative response and conditionally representative response. The first definition is introduced by Schouten *et al.* (2009); the second is discussed by Schouten *et al.* (2011).

Let X be a vector of auxiliary variables, for example, a vector consisting of age, gender, degree of urbanization of residence area, and the observed status of the dwelling for a household survey or the type of business and the number of employees for a business survey. Response is called representative with respect to X when the response propensities of all subpopulations formed by the auxiliary variables are constant (and, hence, equal to the overall response rate).

Let Z be one of the variables in the vector X, for example, age or number of employees, and let X^{-} represent vector X where Z is deleted. The response is called conditionally representative for Z given X^{-}, when response propensities for the subpopulations formed by Z are constant within the subpopulations formed by the other auxiliary variables in X. Hence, age is conditionally representative with respect to gender, urbanization, and dwelling status when the propensities for different age categories are constant when the gender, degree of urbanization and the status of the dwelling are fixed.

In other words, representative response can be viewed as arising when the respondents form a random subsample of the survey sample itself and conditionally representative response arises when they form a stratified random subsample.

Representativeness indicators or *R*-indicators (Schouten *et al.*, 2009) measure the extent to which survey response deviates from representative response. The response to a survey is less representative for X when the response propensities show more deviation to the overall response rate.

The *R*-indicator is a simple measure and is based on the standard deviation of response propensities.

$$R(\rho_X) = 1 - 2S(\rho_X), \tag{1}$$

where ρ_X indicates the individual response propensities given the auxiliary variables X and $S(\rho_X)$ is the standard deviation of these propensities. The transformation in (1) assures values between 0 and 1, where 1 is representative response.

As we will show in the following sections, *R*-indicators can be used to compare representativeness of surveys, but they cannot be used to identify subgroups in monitoring and improving representativeness of response. For this reason, *R*-indicators were supplemented by partial *R*-indicators; see Schouten *et al.* (2010). Partial *R*-indicators are available at the variable level and at the level of categories of variables in order to allow for detailed levels of analysis. Furthermore, partial *R*-indicators can be computed unconditionally and conditionally, thereby relating to representative and conditionally representative response.

Essentially, partial *R*-indicators decompose the variation in response propensities according to different values of the auxiliary variables. The standard deviation $S(\rho_X)$ in (1) can be rewritten as

$$S^2(\rho_X) = S^2_W(\rho_X|H) + S^2_B(\rho_X|H), \tag{2}$$

where H is some stratification of the population into subgroups and $S^2_W(\rho_X)$ and $S^2_B(\rho_X)$ are, respectively, the within and between variances of the response propensities. The between variance measures the variation in response propensities between the subgroups and the within variance measures the remaining variance within the subgroups.

The unconditional partial *R*-indicator for a variable Z is based on the between variance for a stratification of the population using Z, for example, the between variance for different age groups or different numbers of business employees. It measures the extent to which the response is representative for Z. It is defined as

$$P_u(Z, \rho_X) = S_B(\rho_X|Z). \tag{3}$$

The conditional partial *R*-indicator for Z given X^- is based on the within variance for a stratification of the population using X^-, for example, the remaining variance due to age within subgroups formed by gender, urbanization, and dwelling status or due to number of employees within groups formed by business type. The conditional partial *R*-indicator measures the extent to which response is conditionally representative. It is defined as

$$P_c(Z, \rho_X) = S_W(\rho_X|X^-). \tag{4}$$

Variable level partial *R*-indicators follow directly from the decomposition in (2). Partial *R*-indicators for categories of variables are composed by isolating the contribution to between and within variances attributable to those categories. In all cases, partial *R*-indicators ideally have values equal to zero. The unconditional category-level partial *R*-indicators have a positive or negative sign in order to indicate the direction of the representation of the category, for example, a negative sign for males implies that males responded less to the survey. The conditional category-level partial *R*-indicators do not have a sign and are always positive. Conditionally, it does not have a meaning to attach a direction to the indicator as a category may perform better in one subpopulation but do less in another. For example, young males may do better than average while older males may do worse than average. We refer to Schouten *et al.* (2011) for detailed descriptions of the various indicators.

The definitions of representative and conditionally representative response do not focus on survey variables that are missing, but they can be rephrased in terms of well-known

missing-data mechanisms. Little & Rubin (2002) introduced Missing-Completely-at-Random (MCAR), Missing-at-Random (MAR) and Not-Missing-at-Random (NMAR). When response is representative for X then the missing data is MCAR(X). When response is conditionally representative for Z given X, then the missing data is MAR(Z, X). When the latter is not true, then the missing data is NMAR(Z, X). The partial R-indicators, in fact, measure Euclidean distances to MCAR and MAR. The unconditional partial R-indicator P_u measures the distance to MCAR(X), while the conditional partial R-indicator P_c equals the distance to MAR(Z, X). Hence, the conditional indicator reflects the extent to which missing data is NMAR for Z given X.

One important remark needs to be made concerning the estimation of the R-indicators and partial R-indicators. Because individual response propensities are unknown, they need to be estimated from the survey data. The estimation comes at a price. The indicators themselves are subject to the variation in samples. This variation needs to be accounted for. It goes beyond the scope of this paper, to discuss estimation of indicators. We refer to Shlomo *et al.* (2011) for a detailed description and discussion. The website www.risq-project.eu contains open source code in SAS and R that computes the various indicators and accounts for the sampling variation.

3 Comparing Representativeness of Different Surveys

In this section, we discuss and illustrate the first of the purposes of the proposed indicators, that of comparing different surveys. We start by elaborating on the choice of auxiliary variables and next provide a cross-country example.

3.1 The Choice of Auxiliary Variables

In section 2, we defined R-indicators as measures of representative response. An R-indicator cannot be viewed separate from the auxiliary vector X that was used to evaluate the representativeness of response. As such indicator values should always be reported with reference to the auxiliary variables. Consequently, when comparing multiple surveys within one survey institute, over survey institutes or even over countries, one needs to fix the set of auxiliary variables used for each of the surveys. Because variables represent characteristics of population units, it does not make sense to compare indicator values for different population entities, like households versus businesses. Also, one needs to realize that different subpopulations of a larger population, like single households and multiperson households or countries in the EU, consist of intrinsically different sets of population units.

There are two conflicting aims that need to be balanced when selecting auxiliary variables in the comparison of different surveys. On the one hand, it is desirable to choose auxiliary variables that are maximally correlated with the variables of analytic interest in each survey. On the other hand, the choice is constrained to the set of auxiliary variables that is available for each of the surveys.

The wider the scope of the comparison, the more restrictive the availability of variables will be. Within one survey institute one is likely to use one sampling frame, have access to the same register data and collect similar paradata for surveys. Multiple countries, however, may have completely different traditions and legislation, which will limit the set of auxiliary variables that is shared.

3.2 Example

To illustrate the comparison of different surveys, we present R-indicators for four surveys from different countries: the Dutch Health Survey 2005, the Belgian and Norwegian European

Table 1
Sample sizes, response rates and R-indicators for six surveys. R-indicators are evaluated with respect to gender, age, and degree of urbanization.

	Sample Size	Response Rate	R-Indicator
Health survey 2005	15,411	67.3%	0.832
ESS 2006 (Belgium)	2,927	61.4%	0.807
ESS 2006 (Norway)	2,673	65.6%	0.762
Survey on level of living 2004	4,837	69.1%	0.872
LFS Quarter 3, 2007	2,219	70.1%	0.854
LFS Quarter 4, 2007	2,215	69.3%	0.807

Social Surveys 2006, the Norwegian Survey of Level of Living 2004, and the Slovenian Labor Force Survey 2007. Gender, age, and degree of urbanization are available in all surveys and are used as the auxiliary variables. We present a short description of each survey.

The Dutch Health Survey 2005: The Dutch Health Survey is a continuous survey among Dutch residents of 4 years and older with questions about health, life style and use of medical care. In 2005 it was conducted using face-to-face interviewing.

Belgian and Norwegian European Social Surveys (ESS) 2006: The European Social Survey is a biennial face-to-face, multicountry survey of individuals covering over 30 nations. It is an academically driven social survey designed to chart and explain the interaction between Europe's changing institutions and the attitudes, beliefs and behaviour patterns of its diverse populations.

Norwegian Survey of Level of Living 2004: The Survey of Level of Living has two main purposes. One is to throw light on the main aspects of the living conditions in general and for various groups of people. Two is to monitor development in living conditions, both in level and in distribution. The survey topics change during a 3-year cycle. Housing conditions, participation in organizations, leisure activities, offences, and fear of crime were topics in 2004. The survey is a mix of face-to-face and telephone interviews.

Slovenian Labour Force Survey (LFS) 2007: The Slovenian Labour Force Survey is an EU harmonized rotating panel survey conducted continuously through the year. The data contains employment related characteristics and demographic characteristics of all individuals 15 years or older living in selected households. The panel consists of a face-to-face first wave and subsequent telephone waves.

Table 1 shows the sample sizes, response rates, and *R*-indicators for the selected surveys. The ESS response is evaluated for the Belgium and Norway response sets. The Slovenian Labor Force survey is analized for two of the quarters of 2007.

What does the comparison tell us about the differences between the surveys? The values of the *R*-indicator tell us that for gender, age, and urbanization the surveys show different variation in response propensities. The results also indicate that there is at best a weak relation between representativeness on the selected variables and the response rate, a finding that confirms those in Groves & Peytcheva (2008). The differences may be partially due to demographic differences between countries; the same gender, age, and urbanization groups may show exactly the same response behaviour over countries but their population distributions may be very different. Part of the differences may be due to the data collection design, most notably the survey mode. Finally, within countries and designs, the surveys may show different representativeness because of the survey topics. The *R*-indicators, however, do not tell us that the key variables in the surveys have different non-response biases. They tell us only that there is a different risk of bias for an arbitrary variable when it would be measured in each of the surveys.

The four surveys share only a small number of auxiliary variables. Statements about the representativeness of the surveys would be more relevant, if the set of variables would be

Table 2
The response rate and R-indicator for the VAT register of January, June and December after 25, 30, and 60 days of data collection. The R-indicator is evaluated with respect to reported VAT of previous year and total wages.

	January			June			December		
	25 days	30 days	60 days	25 days	30 days	60 days	25 days	30 days	60 days
Rate	19.7%	26.1%	28.1%	64.1%	81.5%	83.2%	48.1%	84.1%	88.0%
R	0.683	0.604	0.614	0.739	0.716	0.731	0.846	0.769	0.815

extended with variables like income, educational level, ethnicity, and type of household. For instance, for the ESS, it would be informative to define a "gold standard" of auxiliary variables for which representativeness is evaluated.

4 Comparing Representativeness Over Time

In this section, we restrict ourselves to a single survey and compare the representativeness longitudinally, that is, per month, quarter, or year. This setting also applies to the evaluation of the representativeness of a survey panel with respect to the original recruited set of households or persons when the panel is subject to attrition.

4.1 The Choice of Auxiliary Variables

The choice of auxiliary variables for a single survey is subject to the same trade-off as for multiple surveys. The ideal set of auxiliary variables and the available set of auxiliary variables may only partially overlap. There is, however, an important difference; the survey variables of interest are to a large extent constant; although they may be subject to topic rotation over time. As a consequence, although the two sets may overlap, the auxiliary variables of interest are not just those that relate to the main subpopulations of interest in publications and analysis, but also those that relate to the key survey variables. Paradata, related to the topic of the survey, may be added as auxiliary variables. Examples of such variables are the observed status of dwellings in a housing survey, the observed presence of children in surveys on family growth, the number of visits to contact in labour force surveys and the number of employees in short term business statistics.

4.2 Example

We computed the *R*-indicator for the 2007 Dutch business register of VAT reports for the months January, June, and December. Businesses have to report their VAT to the Tax Board on a monthly, quarterly, or annual basis depending on their size. Small companies report only once a year while big companies have to submit VAT reports every month. Statistics Netherlands uses the VAT records as input to statistics about business revenue. For monthly statistics the VAT reports need to be available between 25 and 30 days after the end of the reference month. After 25 days, processing of data begins and after 30 days the statistics are made public. Because the reporting frequency depends on the size of the company, the months January, June, and December are very different. For January only, monthly reports are available, while for June and December also, respectively, the quarterly and annual reports can be used. We view the completion of the register at a given point in time as response and *R*-indicators as measures of the representativeness of the available reports.

The response rates and *R*-indicators are given in Table 2. For the estimation of the response probabilities we used VAT reported 1 year earlier in the same month and the total wages of the

same reporting month. The total wages are also reported to the Tax Board and are available quickly after the end of the reporting month. The total wages need to be reported every month; there are no differences in reporting frequency between businesses. The response rates are given after 25, 30, and 60 days. The response rate for January is extremely low, only 20% of the businesses have submitted a tax report after 25 days. For June and December, these rates are much higher. After 30 days almost 85% of the businesses has reported for December.

From Table 2, we can conclude that the representativeness is lowest for January and highest for December. For each of the 3 months it does not pay off to wait longer than 25 days when it comes to representativeness.

Different from the example in section 3.2, when evaluating a single survey or register, variables can be selected that are relevant to the statistics based on that survey. For the VAT register we focus on variables that relate to business revenue. Obviously, the VAT registered in the previous year and the total wages in the data collection month are almost ideal predictors; making the evaluation of representativeness very relevant.

5 Monitoring a Survey During Data Collection

In this section we explain and show how indicators can be used for monitoring the representativeness of response during data collection. First, we discuss how to implement the indicators in data collection monitoring and propose a strategy to detail the level of analysis in order to improve the utility of the various indicators. Next, we illustrate the strategy by an example from survey practice.

5.1 From Evaluating to Monitoring

Keywords in monitoring response are data collection subprocesses, data collection design features and paradata.

In order to go from evaluating to monitoring, the data collection design needs to be divided into its basic subprocesses: assessing eligibility, making contact, judging the level of ability in terms of physical, and mental condition and in terms of language, and obtaining response. These subprocesses are sequential, although some subprocesses coincide in some cases. When data is collected through panels, then additional subprocesses are panel recruitment and panel attrition.

Furthermore, different design features like the advance letter, survey mode, contact protocol, incentives, refusal conversion strategies, and interviewer may be crossed with each of the subprocesses, if applicable, and monitored separately. For example, in a mixed-mode survey where all households are invited to fill in a web questionnaire and non-responding households receive a face-to-face follow-up, representativeness may be evaluated for web response, for the combined web and face-to-face response, and for face-to-face response relative to web non-response. Within the face-to-face follow-up, one may evaluate representativeness after a certain number of visits or data collection weeks.

Paradata are data about the data collection process. Such data are needed to distinguish the data collection subprocesses and to derive the impact of various design features. However, additionally paradata are viewed as observations on the sample units and may be treated as auxiliary variables in the assessment of representativeness. For example in housing surveys one may employ interviewer observations on the dwelling of households.

5.2 Representativeness at Different Levels of Detail

The *R*-indicators, unconditional, and conditional variable-level partial indicators and the unconditional and conditional category-level partial indicators form a set of measures that allow

Table 3
Response to survey of consumer confidence 2005.

	Number	Rate
Not eligible	267	1.5%
Not contacted	954	5.4%
Language problems	143	0.8%
Not able during fieldwork period	955	5.4%
Refusal	3 735	21.2%
Cooperation	11 870	67.2%

for analysis at different levels of detail. Monitoring may be done using a number of steps that should be repeated during data collection:

(1) Compute the *R*-indicator and compare to previous waves of the same survey.
(2) Assess the unconditional variable-level partial *R*-indicators for all selected auxiliary variables; the variables that have the highest values are the strongest candidates to be involved in design changes and increased follow-up efforts.
(3) Assess the conditional variable-level partial *R*-indicators for all selected auxiliary variables; the conditional values are needed in order to check whether some of the variables are strongly collinear. If indicator values remain high then the strongest variables are selected. If indicator values vanish by conditioning, then it is sufficient to focus only on a subset of the variables.
(4) Repeat steps 1 and 2 but now for the category-level partial *R*-indicators and for the auxiliary variables selected in step 3 only; the subgroups that need to be targeted in design changes are those categories that have large negative unconditional values and large conditional values.

5.3 Example

The use of indicators for monitoring survey data collection is illustrated through the 2005 Dutch Survey of Consumer Confidence. The Survey of Consumer Confidence (SCC) is a monthly cross-sectional survey among 1 500 households of which a listed telephone number can be found. Fieldwork is conducted in the first 10 workdays of each month.

Table 3 reports the response to the SCC in 2005. The response rate was 67.2%. The majority of non-response came from refusals. In the following, we focus on the two subprocesses that are known to have the strongest impact on response rates and response representativeness: contact and cooperation. We combine the less influential causes for non-response, language problems, and not able during fieldwork with refusal into the subprocess of obtaining cooperation. Hence, when monitoring cooperation conditional on contact and eligibility, we contrast response to refusal or language problem or not able.

Following the strategy of section 5.1, we first analyse variable level partial *R*-indicators. We selected the variables type of household, age, ethnicity, urban density of residence area, gender, and average house value at zip code. From literature, these variables are known to relate to contact and cooperation. In this particular study, we did not have paradata observations, but they can be added for the two subprocesses in a straightforward way.

In Table 4, the indicators are given for contact and cooperation and for the overall response. Cooperation is evaluated conditionally on contact and eligibility and contact is evaluated conditionally on eligibility.

Somewhat surprisingly in obtaining contact all variables play almost an equal role except for average house value. However, after conditioning it is age that contributes most to the

Table 4
Variable level partial R-indicators for six auxiliary variables in the SCC 2005. The six variables are type of household, age, ethnicity, urban density, gender, and average house value at zip code. The response propensities were also estimated using these six variables.

	Unconditional			Conditional		
	Contact	Cooperation	Response	Contact	Cooperation	Response
Household	0.0018	0.0054	0.0063	0.0008	0.0012	0.0015
Age	0.0017	0.0069	0.0058	0.0021	0.0049	0.0036
Ethnicity	0.0014	0.0031	0.0037	0.0010	0.0025	0.0029
Urban	0.0014	0.0013	0.0019	0.0009	0.0010	0.0010
Gender	0.0021	0.0049	0.0058	0.0008	0.0006	0.0010
House value	0.0007	0.0026	0.0027	0.0008	0.0007	0.0008

Table 5
Category level partial R-indicators for age in the SCC 2005. The response propensities were estimated using type of household, age, ethnicity, urban density, gender, and average house value at zip code.

	Unconditional			Conditional		
	Contact	Cooperation	Response	Contact	Cooperation	Response
< 25	–0.0027	0.0025	0.0003	0.0020	0.0057	0.0044
25–29	–0.0084	0.0088	0.0020	0.0082	0.0101	0.0048
30–34	–0.0088	0.0139	0.0065	0.0092	0.0126	0.0064
35–39	–0.0065	0.0168	0.0109	0.0088	0.0109	0.0057
40–44	–0.0014	0.0184	0.0163	0.0048	0.0118	0.0085
45–49	0.0016	0.0108	0.0114	0.0031	0.0080	0.0055
50–54	0.0041	0.0049	0.0077	0.0039	0.0090	0.0057
55–59	0.0029	0.0108	0.0124	0.0038	0.0129	0.0118
60–64	0.0008	0.0063	0.0066	0.0026	0.0107	0.0087
65–69	0.0004	0.0024	0.0025	0.0024	0.0087	0.0070
69	0.0083	–0.0595	–0.0508	0.0116	0.0366	0.0278

variance in response propensities. The partial *R*-indicators for cooperation are larger than for contact. Unconditionally, the strongest variables are age, type of household, and gender, but after conditioning it is age that contributes most. Remarkable is also the large value for ethnicity. It turns out that ethnicity has an impact on cooperation that is almost orthogonal to the impact of the other variables.

Summarizing, in improving representativeness for contact we should look at age and for cooperation we should consider age and ethnicity. The other variables, however, also play a role in each of the subprocesses but not as distinct. Looking at the three subprocesses, the biggest potential increase in representativeness can be obtained from reducing refusal.

If we join the impact of contact and cooperation, then all unconditional partial *R*-indicators are larger than for each of the subprocesses alone except for age. For age, the value for cooperation is larger than for overall response, indicating that the different subprocesses counteract each other to some extent. After conditioning the strongest overall variables are age and ethnicity. Hence, if one would not decompose non-response into non-eligible, non-contact, and refusal, then these two variables should be targeted.

Tables 5 and 6 show the category level partial *R*-indicators for age and ethnicity for contact, cooperation, and overall response.

A closer look at the contact category-level indicators for age reveals that persons younger than 45 affect representativeness in a negative way. Especially persons between 25 and 40 show low contact propensities and, as a consequence, increase the variation in contact propensities. Conditioning on other variables has a minor impact. Hence, when improving contact representativeness, the focus should be on persons between 25 and 40 years of age.

Table 6
Category level partial R-indicators for ethnic origin in the SCC 2005. The response propensities were estimated using type of household, age, ethnicity, urban density, gender, and average house value at zip code.

	Unconditional			Conditional		
	Contact	Cooperation	Response	Contact	Cooperation	Response
Native	0.0043	0.0038	0.0067	0.0050	0.0069	0.0108
Non-western						
1st generation	–0.0069	–0.0127	–0.0166	0.0019	0.0054	0.0096
2nd generation	–0.0052	0.0005	−0.0033	0.0048	0.0091	0.0038
Western	–0.0055	–0.0173	–0.0199	0.0054	0.0141	0.0171
Other	–0.0010	0.0084	0.0072	0.0025	0.0032	0.0036
Unknown	–0.0086	–0.0202	–0.0246	0.0032	0.0162	0.0159

For obtaining cooperation, we consider both age and ethnic origin. For age one group immediately jumps out, persons of 70 years and older. This relatively small group has a large negative value. Conditional on the other variables the value remains large. With respect to ethnic origin the subgroups with large negative unconditional indicator values are first generation non-western non-natives, western non-natives, and persons for which the ethnic origin is unknown. The conditional values for western non-natives and unknown ethnic origin remain relatively large. Therefore, we conclude that measures to improve cooperation representativeness should aim at elderly persons, western non-natives, and persons with unknown ethnic origin.

Finally, we analyse the overall representativeness. For age, cooperation and contact attenuate indicator values. However, the overall impact of elderly persons remains strong. For ethnic origin, contact and cooperation tend to enforce each other; almost all category-level values point in the same direction. Again the negative values come from first generation non-western non-natives, western non-natives, and unknown ethnic origin. Conditionally, the values remain largest for first generation non-western non-natives and unknown ethnic origin. We can conclude that for overall response one should target the same groups as for cooperation.

Tables 4–6 form a subset of all partial *R*-indicator values computed for the SCC 2005. The monitoring strategy of section 5.1 led to an efficient search for subpopulations that affect the representativeness in the various subprocesses.

6 Improving Representativeness through Adaptive and Responsive Survey Designs

The last purpose for the indicators to be discussed is improving representativeness. Partial *R*-indicators may be tools to identify subgroups that need to be targeted or prioritized in non-response follow-up.

6.1 From Monitoring to Designing and Intervening

The strategy of section 5.1 naturally leads to the identification of population subgroups that may require additional efforts. Obviously, there is more to improving quality of survey response than the identification of such subgroups.

First, there is the quality–cost trade-off. If the budget of the survey is fixed, one may reallocate resources in such a way that the selected subgroups get more attention while the other subgroups get less attention. For example, some subgroups may get a larger number of contact attempts, visits, or reminders while other subgroups get a smaller number. One may decide to assign the best performing interviewers to the lowest responding subgroups. Alternatively, one may decide to increase budget in order to increase efforts for the selected subgroups and to maintain efforts for the other subgroups at traditional levels.

Second, in the quality–cost trade-off one needs to choose a quality objective function. We propagate to use the representativeness of response as the objective function. However, care is needed when aiming at a more representative response. Without any specific population parameter in mind, one may seek to obtain as high an *R*-indicator as possible given the budget. Because response propensities are unknown, they are estimated with a set of auxiliary variables, say gender, age, urbanization, and status of the dwelling. Now, a higher *R*-indicator can easily be obtained by subsampling or erasing subgroups. Suppose young males in big cities have the lowest response propensity of 20%. All other subgroups may be subsampled with a rate proportional to their response propensity. If young females in rural areas have a response propensity of 50%, then their subsample rate would be 40%. After this artificial response enhancement procedure the *R*-indicator equals one and the response rate dropped to 20%. Hence, representativeness may always be improved by reducing response. Clearly, increasing budget by extending effort for those same subgroups is always profitable in terms of non-response bias. Therefore, we propose that it should be the objective to improve the *R*-indicator while restricting the response rate not to drop.

Third, here we completely focus on the impact of non-response. Doing so, we ignore the data quality, for example, the measurement error that may be the result of different data collection designs. In changing data collection, one must be sure that measurement errors do not counteract non-response errors. For instance, a follow-up of specific non-respondents may result in a more representative response, but the converted refusers may provide socially desirable answers. As a consequence, the impact of non-response has decreased but the impact of measurement errors has increased, and the net effect may still be negative in terms of total survey error.

Fourth, differentiating effort over subgroups naturally leads to responsive survey designs (Groves & Heeringa, 2006; Mohl & Laflamme, 2007; Tabuchi *et al.*, 2009; Laflamme & Karaganis, 2010; Peytchev *et al.*, 2010) and adaptive survey designs (Wagner, 2008). Responsive survey designs use paradata and costs assessments in early data collection phases to detect subprocesses that need more attention and to act based on the observed effectiveness of design features. Adaptive survey designs tailor design features based on historical data before data collection starts. The important observation here is that changing data collection with the goal to get a more representative response implies a differential data collection design. Currently, there is no tradition in differentiating effort over sample units.

Given these considerations, the next step is to search for design features that have a positive impact on the response propensities of subgroups that were identified using the partial *R*-indicators. One way to do this is by means of response models that include design features as treatment variables. Loosveldt & Beullens (2009) and Loosveldt *et al.* (2010) investigate the impact of interviewer assignment and re-assignment, contact mode, and timing of contacts on the representativeness of response.

6.2 Example

We consider again the Survey of Consumer Confidence for illustration of the use of indicators in improving representativeness of survey response. In an experiment an adaptive mixed mode design was compared to the standard CATI SCC. Historical data were used to predict response propensities in CATI, web and mail. In doing so, the subprocesses eligibility, contact, ability to respond, and cooperation were distinguished. The goal of the pilot was to increase response representativeness in terms of the *R*-indicator while preserving costs and response rate. In other words, the experimental design costs should not exceed the costs of the regular SCC and the design should lead to a response rate that is larger than or equal to that of the regular SCC.

Table 7
R-indicators for control and experimental group. The indicators were evaluated with respect to type of household, age, ethnicity, urban density, gender, and income. The response propensities were also estimated using these six variables.

	Control Group	Experimental Group
Contact	0.83	0.87
Able to cooperate	0.86	0.85
Cooperation	0.87	0.89
Response	0.77	0.85

In order to achieve the aim of better representativeness with lower costs, a mixed mode design was chosen for the pilot, in which a mail and/or web first round was followed by a telephone follow-up of non-respondents. Mail and web questionnaires not only cost less to administer than telephone questionnaires, they can also reach respondents that are otherwise hard to contact and/or to convince to cooperate.

On the basis of the predicted web and mail response probabilities, the following design was decided upon for the first web/mail wave:

(1) Households with a high chance of cooperation would receive an invitation for the web survey.
(2) Households with a medium chance of cooperation would receive an invitation for the web survey and a mail questionnaire. Either could be filled in.
(3) Households with a low chance of cooperation received only a mail questionnaire. This simplified the advance letter to a great extent, and it was expected that that would be beneficial to response.

In the second wave, the non-response was followed up by telephone. In this wave it was attempted to

(1) increase probability of contact for sample units with a low contact propensity,
(2) dampen the number of contact attempts for units with a high contact propensity,
(3) stimulate cooperation for sample units with a low cooperation propensity, and
(4) dampen cooperation for sample units with a high cooperation propensity.

For different subgroups, different contact strategies were chosen. One strategy was defined for elderly native households. These were called primarily during daytime. Another strategy was defined for single households, households of non-native origin, households in highly urban areas and households consisting of young people. They were to be called in every shift (morning, afternoon and evening), every day of the fieldwork period. Four different strategies were such defined.

In order to influence cooperation probability, the assignment of households to specific interviewers was manipulated. Based on their cooperation rates, interviewers were classified in three categories. The interviewers with the lowest response rates called households with the highest probability of cooperation and vice versa.

Table 7 shows the *R*-indicator values for the various subprocesses and the overall response for both the control group and the experimental group. The control group and the experimental group were equally sized and consisted of 3 000 sample cases each. Only subprocess Ability to respond shows a slightly lower *R*-indicator in the experimental group. However, this difference is not significant at the 5% level. The only difference that is significant at the 5% level is the *R*-indicator for overall response. The *R*-indicator increases from 0.77 to 0.85. It can be concluded that the experiment was successful in improving representativeness.

Table 8
Unconditional variable level partial R-indicators for six auxiliary variables in control and experimental group. The variables are type of household, age, ethnicity, urban density, gender, and income. The response propensities were also estimated using these six variables.

	Control Group				Experimental Group			
	Contact	Able	Coop	Response	Contact	Able	Coop	Response
Household	0.0049	0.0035	0.0018	0.0088	0.0038	0.0037	0.0029	0.0052
Age	0.0052	0.0038	0.0035	0.0058	0.0033	0.0045	0.0021	0.0036
Ethnicity	0.0033	0.0040	0.0013	0.0071	0.0015	0.0025	0.0017	0.0043
Urban	0.0030	0.0018	0.0016	0.0032	0.0015	0.0011	0.0026	0.0024
Gender	0.0108	0.0101	0.0041	0.0209	0.0071	0.0091	0.0058	0.0134
Income	0.0029	0.0035	0.0014	0.0067	0.0014	0.0038	0.0014	0.0054

Table 9
Conditional variable level partial R-indicators for six auxiliary variables in control and experimental group. The variables are type of household, age, ethnicity, urban density, gender, and income. The response propensities were also estimated using these six variables.

	Control group				Experimental group			
	Contact	Able	Coop	Response	Contact	Able	Coop	Response
Household	0.0016	0.0014	0.0024	0.0022	0.0021	0.0015	0.0022	0.0018
Age	0.0040	0.0026	0.0029	0.0024	0.0031	0.0029	0.0016	0.0013
Ethnicity	0.0005	0.0019	0.0006	0.0017	0.0008	0.0018	0.0009	0.0015
Urban	0.0013	0.0011	0.0009	0.0014	0.0007	0.0010	0.0018	0.0014
Gender	0.0013	0.0017	0.0010	0.0031	0.0014	0.0011	0.0021	0.0012
Income	0.0005	0.0017	0.0013	0.0021	0.0010	0.0026	0.0020	0.0028

Tables 8 and 9 show the unconditional and conditional variable-level partial *R*-indicators, respectively. For making contact the partial *R*-indicators are lower in the experimental group. For ability to respond and cooperation the results are mixed. For overall response the experimental group always scores better with the exception of the conditional value for income. We conclude that the impact of the auxiliary variables has decreased, except possibly for income. Hence, the adaptive survey design was successful as it led to a more representative overall response.

7 Discussion

In this paper, we demonstrated that *R*-indicators and partial *R*-indicators can be used for different purposes. We discussed and illustrated the use of indicators in evaluating different surveys or a single survey in time, in monitoring and in adapting survey data collection. Four properties are indispensable for any indicator when having these purposes in mind: They should be easy to interpret, they must be based on available auxiliary data and survey data only, they should be relevant or in other words lead to effective survey designs, and they should allow for analysis at different levels of detail.

Because it is not straightforward how to go from evaluating to monitoring and from monitoring to improving the design of a survey, we explained how the *R*-indicator and partial *R*-indicators can be used to detail the level of analysis. Because there may be many population subgroups to look at, given the available variables, it benefits monitoring greatly if indicators can be measured at different levels; from full population to specific categories. This property is helpful in a search for subgroups that may be targeted in adaptive and responsive survey designs.

In the next sections, we discuss issues that have come up at various presentations of the indicators.

7.1 Strengths and Weaknesses

R-indicators and partial *R*-indicators do not solve the non-response problem. The indicators rely on information that is auxiliary to a survey or register in order to evaluate the representativeness. As a consequence, the indicators depend strongly on the auxiliary variables that are selected. As we have argued throughout the paper, the choice of variables depends on the intended use of the indicators and obviously on the availability of variables. When comparing different surveys, auxiliary variables need to be available in all surveys and be of general interest to all surveys. When the quality of a single survey is assessed, the scope can and must be extended to variables that relate also to the key survey topics.

The dependence on auxiliary variables is a weakness of the indicators, but no indicator can be constructed that is independent of external information, and, hence, of auxiliary variables. However, empirical evidence from follow-up surveys is needed to assess whether smaller *R*-indicators indeed correspond to larger average biases of survey target parameters.

The strength of the indicators lies in its mathematical basis and its structured use of auxiliary information in evaluating representativeness of response. Like the *R*-indicator can be viewed as supplemental to response rates, so can partial *R*-indicators be viewed as supplemental to subgroup response rates. We believe that partial *R*-indicators are more relevant and allow for different levels of analysis, whereas the most obvious alternative, the set of subgroup response rates, lacks those two properties. The main differences between subgroup response rates and partial *R*-indicators are:

(1) Partial *R*-indicators are linked to *R*-indicators, that is, they represent the contribution of variables to the lack of representativeness, while subgroup response rates are linked to response rates. In other words they are conceptually different, although they have response propensities as basic ingredients.
(2) Partial *R*-indicators are available at the variable level.
(3) Partial *R*-indicators are computed both unconditionally and conditionally.
(4) Partial *R*-indicators are weighted differences between subgroup response rates and overall response rates. The weight is proportional to the size in the population. Subgroup response rates are not weighted, and do not account for the size and, hence, impact of a subgroup.

The second difference makes *R*-indicators and partial *R*-indicators fit for detailing the level of analysis. The third and fourth differences make partial *R*-indicators relevant in searching for groups that help improving representativeness of survey response.

7.2 Indicators and Non-response Adjustment

A criticism that is often put forward, is that any traces of non-representative response found by the indicators can be corrected for using standard non-response adjustment methods. It is true that candidate variables for evaluation of representativeness are also candidates for non-response adjustment procedures. There are, however, two important differences between evaluation of representativeness as presented in this paper and non-response adjustment.

The evaluation of representativeness is not directed at a particular survey variable, a particular population parameter or a particular estimator. Representativeness is assessed for the full range of variables, parameters and estimators. For this reason, *R*-indicators and partial *R*-indicators provide an incomplete picture when the purpose is adjustment of non-response bias of a certain population parameter using a specified estimator. Non-response adjustment requires a detailed look at the relation between a specific survey variable and non-response. Clearly, when a survey contains a wide range of key variables, then non-response adjustment tends to focus on the

explanation of non-response behaviour rather than on the explanation of a survey variable. In these settings the indicators may provide some input to non-response adjustment. However, literature, for example, see Schouten (2007) & Särndal & Lundström (2010), has proposed measures that are better suited to select variables in specific non-response adjustment methods like general regression estimators. Särndal & Lundström (2010) propose a selection criterion in the estimation of a population mean, using calibration estimators, that is not survey-variable specific: the coefficient of variation of the non-response adjustment weights. It can be shown that this measure is asymptotically equivalent to the variance of response propensities divided by the response rate squared, when a saturated model of auxiliary variables is used. In other words the variation in propensities, which forms the basis to *R*-indicators, is one of the components to this measure.

The second difference lies in the assumptions underlying to non-response adjustment. Most non-response adjustment methods assume a missing-at-random mechanism; the probability to respond does not depend on the survey variable within strata formed by auxiliary variables. In general, this assumption fails to hold, but any weaker assumption cannot be tested without additional information or intrinsic knowledge on the nature of the non-response behaviour. The indicators in this paper measure the extent to which survey researchers have to rely on such assumptions. Surveys that have a more representative response have a smaller risk of bias remaining after non-response adjustment.

7.3 Future Research

More empirical evidence and experience are needed to get a full insight into the potential role of *R*-indicators and partial *R*-indicators in data collection.

Although the values of the indicators are especially interesting in a relative sense, that is, from one design to another, it is paramount to get a better feeling about acceptable levels of the indicators. What levels would mean that immediate action is needed.

Currently additional work is done on confidence intervals for partial *R*-indicators. Partial *R*-indicators have a precision that depends on the sample size. In order to increase the relevance of these indicators, they need to be supported by error margins.

The indicators may be compared to in-depth analyses of non-response, especially to those analyses that led to design changes. Such comparisons would illustrate to what extent suggested actions coincided with indications coming from the partial *R*-indicators.

The indicators may also be applied to other survey errors like item non-response and undercoverage, subject to the same side remarks as for unit non-response. For undercoverage it would be needed that the non-covered units can be identified and can be linked to auxiliary variables. For example, at Statistics Netherlands *R*-indicators are assessed in telephone surveys in order to evaluate the undercoverage of households without a registered phone number.

Finally, the non-response error is one component of the total survey error. A blind focus on a single error source is too simplistic. Care is needed when altering data collection designs in order to improve representativeness. Future research should focus on combining these efforts with assessment of other relevant errors, most notably measurement errors.

Acknowledgements

This research was supported under the research project Representativity Indicators for Survey Quality (RISQ, www.risq-project.eu), funded by the European Commission in the 7th EU

Research Framework Programme. We thank colleagues on the RISQ project for valuable input of ideas to this research.

References

Andridge, R. & Little, R. (2011). Proxy pattern-mixture analysis for survey nonresponse. *J. Official Statist.,* **27**(2), 153–180.

Biemer, P.P. & Lyberg, L.E. (2003). *Introduction to Survey Quality*. Hoboken, NJ: Wiley.

Groves, R.M. (1989). *Survey errors and survey costs*. New York, USA: Wiley.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Pub. Opin. Quart.*, **70**, 646–675.

Groves, R.M., Brick, J., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Kreuter, F., Pennell, B., Raghunathan, T., Schouten, B., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheva, E., Ziniel, S. & Wagner, J. (2008). Issues facing the field: alternative practical measures of representativeness of survey respondent pools. Survey Practice, October 2008. Available online at: http://surveypractice.org/2008/10/30/issues-facing-the-field/.

Groves, R.M., Dillman, D., Eltinge, J. & Little, R. (2002). *Survey Nonresponse*. New York: Wiley Series in Probability and Statistics.

Groves, R.M. & Heeringa, S.G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *J. Roy. Statist. Soc. Ser. A*, **169**, 439–457.

Groves, R.M. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias. *Pub. Opin. Quart.*, **72**, 1–23.

Kreuter, F., Olsen, K., Wagner, J., Yan, T., Ezzati-Rice, T.M., Casa-Cordero, C., Lemay, M., Peytchev, A., Groves, R.M. & Raghunathan, T.E. (2010). Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys. *J. Roy. Statist. Soc. Ser. A*, **173**, 389–407.

Laflamme, F. & Karaganis, M. (2010). Implementation of responsive collection design for CATI surveys at Statistics Canada, Paper presented at Q2010, 3–6 May, Helsinki, Finland.

Little, R.J.A. & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. New York: Wiley Series in Probability and Statistics.

Loosveldt, G. & Beullens, K. (2009). Fieldwork Monitoring, Deliverable 5, RISQ project. Available online at: http://www.risq-project.eu.

Loosveldt, G., Beullens, K., Luiten, A. & Schouten, B. (2010). Improving the fieldwork using R-indicators: applications, RISQ project. Available online at: http://www.risq-project.eu.

Mohl, C. & Laflamme, F. (2007). Research and responsive design options for survey data collection at Statistics Canada. In *Proceedings of ASA Joint Statistical Meeting, Section 293, July 29–August 2*, Salt Lake City, USA.

Peytchev, A., Riley, S., Rosen, J., Murphy, J. & Lindblad, M. (2010). Reduction of nonresponse bias in surveys through case prioritization. *Surv. Res. Meth.*, **4**(1), 21–29.

Särndal, C.E. (2011). The 2010 Morris Hansen Lecture. Dealing with survey nonresponse in data collection, in estimation. *J. Official Statist.*, **27**(1), 1–21.

Särndal, C.E. & Lundström, S. (2010). Design for estimation: identifying auxiliary vectors to reduce nonresponse bias. *Survey Methodol.*, **36**(2), 131–144.

Schouten, B. (2007). A selection strategy for weighting variables under a not-missing-at-random assumption. *J. Official Statist.*, **23**(1), 1–19.

Schouten, B., Cobben, F. & Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodol.*, **35**, 101–113.

Schouten, B., Shlomo, N. & Skinner, C. (2011). Indicators for monitoring and improving survey response. *J. Official Statist.*, **27**(2), 231–253.

Shlomo, N., Skinner, C.J. & Schouten, B. (2012). Estimation of an indicator of the representativeness of survey response. *J. Statist. Plann. Inference*, **142**, 201–211.

Tabuchi, T., Laflamme, F., Philips, O., Karaganis, M. & Villeneuve, A. (2009). Responsive design for the survey of labour and income dynamics. In *Proceedings of the Statistics Canada Symposium 2009*, Longitudinal Surveys: from design to analysis, Statistics Canada.

Wagner, J. (2008). *Adaptive survey design to reduce nonresponse bias*. PhD thesis, University of Michigan, USA.

Wagner, J. (2010). The fraction of missing information as a tool for monitoring the quality of survey data. *Pub.Opin. Quart.*, **74**(2), 223–243.

Résumé

La non-réponse est une source d'erreur importante dans de nombreuses enquêtes. Étant donné que les enquêtes sont souvent des opérations coûteuses, le compromis entre qualité et coût est omniprésent dans leur conception aussi bien que dans leur analyse. Les progrès du téléphone, des ordinateurs et d'internet tous ont eu, et ont encore, un impact considérable sur la conception des enquêtes. Récemment, l'accent a été mis sur les méthodes de collecte de données d'enquêtes de surveillance et l'adaptation est apparue comme un nouveau paradigme réduisant de façon efficace les erreurs liées à la non-réponse. «Paradonnées» (paradata) et plans de sondage adaptatifs sont les mots-clés de ces nouveaux développements. Les conditions préalables à l'évaluation, à la comparaison, à la surveillance et à l'amélioration de la qualité de la réponse du sondage sont un cadre conceptuel pour l'étude de la représentativité des résultats d'enquêtes et de leurs mesures de déviation, ainsi que pour l'identification des sous-populations requérant un effort accru. Dans cet article, nous présentons un aperçu des indicateurs de représentativité ou R-indicateurs qui sont propres à ces fins. Nous donnons plusieurs exemples, et des lignes directrices pour leur mise en pratique.

[*Received August 2011, accepted May 2012*]