



RISQ

Representativity Indicators
for Survey Quality

RISQ manual

Tools in SAS and R for the computation of R-indicators and partial R-indicators

Work package 8

Deliverable 12.1

Vincent de Heij, Barry Schouten

Centraal Bureau voor de Statistiek, The Netherlands

Natalie Shlomo

University of Southampton, United Kingdom

May 10, 2010

Acknowledgements: The authors thank Katja Rutar and Johan Fosen for their comments and input.

Table of contents

Acknowledgements: The authors thank Katja Rutar and Johan Fosen for their comments and input.
..... 0

Table of contents 1

1. Introduction 1

2. Downloading and installing the RISQ suite 2

3. Getting started 2

 3.1 Getting started in R 2

 3.2 Getting started in SAS 3

4. The R-indicator 5

 4.1 Output in R 6

 4.2 Output in SAS 7

5. Bias adjustment and confidence intervals of R-indicators 7

6. Unconditional partial indicators on the variable level 8

 6.1 Output in R 9

 6.2 Output in SAS 9

7. Unconditional partial indicators within categories 10

 7.1 Output in R 10

 7.2 Output in SAS 11

8. Conditional partial indicators on the variable level 12

 8.1 Output in R 13

 8.2 Output in SAS 13

9. Conditional partial indicators within categories 14

 9.1 Output in R 14

 9.2 Output in SAS 15

10. Bias adjustment and confidence intervals of partial R-indicators 16

11. Maximal absolute bias and maximal absolute contrast 17

12. General guidelines to R-indicators and partial R-indicators 18

12. Visualising R-indicators in R-cockpit 19

13. Future releases of RISQ_R-indicators in SAS and R 19

1. Introduction

This document is one of the two manuals of software developed within project RISQ (Representativity Indicators for Survey Quality). It describes the R and SAS software libraries that can be used for the computation of R-indicators and partial R-indicators. The other manual describes the graphical tool called R-cockpit. The RISQ project is financed by the 7th EU Research Framework Programme.

The RISQ suite is developed in SAS and in R and is available at www.risq-project.eu. In this manual we give basic background to the various indicators developed under the project, we explain how the suite can be used and adapted to any survey data set, and we illustrate its use for the anonymised data set that can be downloaded from the website.

Detailed background to the concepts and ideas behind representativity indicators can be found in the following documents:

- Shlomo, N., Skinner, C., Schouten, B., Bethlehem, J., Zhang, L.C. (2009), Statistical properties of representativity indicators, RISQ deliverable 2.1
- Shlomo, N., Skinner, C., Schouten, B., Carolina, N., Morren, M. (2009), Partial indicators for representative response, RISQ deliverable 4

Guidelines and a general overview are contained in the following documents:

- Schouten, B., Morren, M., Bethlehem, J., Shlomo, N., Skinner, C. (2009), How to use R-indicators?, RISQ deliverable 3
- Schouten, B., Bethlehem, J. (2009), Representativeness indicators for measuring and enhancing the composition of survey response, RISQ deliverable 9

Examples of the use of representativity indicators in survey data collection monitoring are given in the following documents:

- Loosveldt, G., Beullens, K. (2009), Fieldwork monitoring, RISQ deliverable 5
- Loosveldt, G., Beullens, K., Luiten, A., Schouten, B. (2010), Improving the fieldwork using R-indicators: applications, RISQ deliverable 6

All documents are available at www.risq-project.eu.

2. Downloading and installing the RISQ suite

The SAS and R programs can be downloaded from the RISQ website. From the RISQ website also an anonymised SPSS survey data set can be downloaded. It is called RISQ-test.sav and contains approximately 35,000 persons. In the following we will refer to it as RISQ-test. The data set can be used to test the RISQ suite. It will be used in the examples below.

For the moment a single file contains all the R-code which is needed to determine the R-indicators. In the near future the single file will be replaced by a package. Sourcing the single file will make the functions available in R;

```
> source("RISQ_R-indicators_v1.0.r")
> ls()

[1] "%sub%"                "getBiasRPopulationBased"
[3] "getBiasRSampleBased"  "getPartialRConditional"
[5] "getPartialRs"         "getPartialRUnconditional"
[7] "getRIndicator"        "getRPopulationBased"
[9] "getRSampleBased"      "getSampleDesign"
[11] "getTrace"             "getVariables"
[13] "getVarianceRPopulationBased" "getVarianceRSampleBased"
[15] "isSampleBased"        "newResponsModel"
[17] "weightedVar"
```

Only two functions are relevant for a user of the R-code, `getRIndicator` and `newResponsModel`. The user never has to call directly any of the other functions.

In SAS all computations are done within program `RISQ_R-indicators_v1.0.sas`

3. Getting started

3.1 Getting started in R

To load the RISQ-test data set, the function `read.spss` from the package `foreign` is needed. To load the RISQ-test data set `read.spss("RISQ-test.sav")` can be used in the folder where file is stored. To

transform the list of vectors which `read.spss` returns, into a data frame¹, the function `as.data.frame` can be used.

```
> library(foreign)
> sampleData <- read.spss("RISQ-test.sav")
> sampleData <- as.data.frame(sampleData)
> summary(sampleData[c("respons", "gender", "age", "urb")])
```

respons	gender	age	urb
N/a: 0	Male :17667	35-39 years: 3572	Very strong:5637
No :16076	Female:17788	40-44 years: 3424	Strong :9419
Yes:19379		30-34 years: 3352	Average :7443
		50-54 years: 3174	Little :7864
		45-49 years: 3106	Not :5092
		55-59 years: 2942	
		(Other) :15885	

Before the R-indicators can be calculated the response model has to be defined. The left hand side of the formula (the part left of the symbol`~`) states the response variable, the right hand side (the part right of the symbol`~`) states the model which will be used to fit the response. A model may consist of main effect terms and interaction effect terms. For example, the next three formulas are allowed;

```
> respons ~ gender * age # Full model
> respons ~ gender + age # Only main effects
> respons ~ gender:age # Only interaction effects
```

All variables which are used in the formula have to be members of the data frame with sample data. The variables on the right hand side of the formula should be factors, the response variable on the left hand side of the formula should either be a factor (logistic regression) or a numeric variable with values zero or one (linear regression). A variable, e.g. age, is transformed into a factor by

```
> sampleData$age <- factor(sampleData$age)
```

A response model is created with the function `newModel`;

```
> responsModel <- newResponsModel(respons ~ gender + age + urb,
+ family = 'binomial')
```

The first argument of `newResponsModel` is the formula which defines the response model, the second argument either has the value 'binomial' for logistic regression or 'gaussian' for linear regression.

3.2 Getting started in SAS

The following steps are needed to prepare RISQ-test.sav for computing R-indicators and partial R-indicators:

Step 0: Transfer the data set to SAS in SPSS by saving it as a SAS data file.

Step 1: The first part of the preparation to run the SAS program is for the user to input information about the dataset, the relevant variables to be used in the logistic regression model and other data set parameters. We refer to the screen shot in figure 3.2.1 as an example.

¹ In R, a data set will usually be an object of the type "data frame". A data frame is usually more convenient than a matrix.

1. Define the name of the SAS library which contains the dataset and will include the outputs. In figure 3.2.1, the first line of the program defines the **libname** as **RISQtest**.
2. Define the following:
 - Size of population – **popsize**
 - Size of sample – **samsize**
 - Number of independent variables in the logistic regression model (including interactions) – **variablenum**. The names of the variables in the model should be in quotes under **var1**, **var2**, etc. In the example in figure 3.2.1, **variablenum=3** and the names of the variables: **var1='gender'**; **var2='agea'**; **var3='hhtype*urban'**; Note, the variables defining interactions are separated by an asterisk '*': **'hhtype*urban'**.
 - Number of variables in the logistic regression model that are main effects only – **variablenoint**. In the example in figure 3.2.1, there are two main effects in the logistic regression model so **variablenoint=2**;
 - Number of variables that are used for stratification of the unconditional partial indicator, Pu (see section 6), that are **NOT** used in the logistic regression model – **variablestrat**. The names of the variables should be in quotes under **strat1**, **strat2**, etc. In the example in figure 3.2.1, **variablestrat=1** and the name of the variable is **strat1='marstat'**
 - Number of variables that are included in the interactions – **variableinter**. The names of the variables in the interactions should be in quotes under **vvar1**, **vvar2**, etc. In the example in figure 3.2.1, **variableinter=2** and the names of the variables: **vvar1='hhtype'**; **vvar2='urban'**; You should not count the same variable twice, for example, if there were two interactions in the model, eg. **var3='hhtype*urban'**; and **var4='hhtype*region'**; then **variableinter=3** and **vvar1='hhtype'**; **vvar2='urban'**; **vvar3='region'**;

Step 2: The second part of the preparation to run the program is to define the labels for the categories of the variables defined in step 1 according to the SAS **Proc Format** statement. See figure 3.2.2 for an example of step 2. In **Proc Format** every variable in the response model has to be mentioned, and for each of them, all its categories have to be stated followed by a label, e.g. 1="a".

Step 3: The last part of the preparation to run the program is to define the dataset, and any necessary transformations or relabeling that may need to be carried out. For instance, the variable **age** in the RISQ-test file was changed to **agea** by collapsing the first three categories (not shown). The user needs to define a response indicator denoted as **responsesamp1** where 1 is a response and 0 is a non-response. In the RISQ-test data file **respons** is the 0-1 indicator for response. The user also needs to define the sample design weights, i.e. the inverse of the sample inclusion probabilities, for all sample units (respondents and non-respondents). For simple random sampling, **piinv** is equal to $1/\pi$ which is the **popsize/samsize** defined in step 1. For any other design, the design weight **d** should be included on the dataset and **piinv** is equal to **d**. See figure 3.2.3 how step 3 is implemented for the RISQ-test file.

Figure 3.2.1: First part of program *RISQ_R-indicators_v1.0.sas*

```

/***** change the name of the library *****/
libname RISQtest 'h:\Documents\risq\risq-test'; run;
/***** fill out this section *****/
%let popsize=35455;
%let samsize=35455;
%let variablenum=3; /**total number of variables in model (including interactions) **/
%let variablenoint=2; /**number of main effects variables in model **/
%let variablestrat=1; /** number of stratifying variables not in the model for partial indicators**/
%let variableinter=2; /** number of variables that are in interactions (do not repeat variables)**/
/** names of main effects variables **/
%let var1='gender';
%let var2='agea';
/** interactions **/
%let var3= 'hhtype*urb';
/** variables in interaction (do not repeat names) **/
%let vvar1='hhtype';
%let vvar2='urb';
/** names of stratifying variable for partial indicator not in the model**/
%let strat1='marstat';

```

Figure 3.2.2: Labelling the categories of the variables

```

PROC FORMAT;
  VALUE hhtype
  1="Single" 2="Couple no children" 3="Couple with children" 4="Single parent" 5="Other";
  VALUE agea 1="12-14" 2="15-17" 3="18-19" 4="20-24" 5="25-29"
  6="30-34" 7="35-39" 8="40-44" 9="45-49" 10="50-54" 11="55-59"
  12="60-64" 13="65-69" 14="70-74" 15="75+" ;
  value urb
  1="very strong" 2="Strong" 3="Average" 4="Little" 5="Not";
  value gender
  1="male" 2="female";
  value marstat
  1="Not married" 2="Married" 3="Widowed" 4="Divorced";
run;

```

Figure 3.2.3: Defining the dataset, transformation, response variable and design weights

```

data RISQtest.att;
set RISQtest.RISQtest;
responsesamp1= respons;
  /**** responsesamp1 is the indicator for response, 1=response, 0=non-response**/

agea=age-2; /**transformations on the data*/
pi=&samsize/&popsize;
piinv=1/pi; /** or define piinv= dweight if there are differential design weights in the file */

run;

```

4. The R-indicator

The R-indicator is a transformation of the variance of estimated response propensities to the [0,1] interval. A value equal to 1 implies representative response. A value equal to 0 implies a maximal deviation from representative response.

Suppose the estimated response probabilities for the n elements in the sample are denoted by $\rho_1, \rho_2, \dots, \rho_n$ and the sample design inclusion weights are denoted by d_1, d_2, \dots, d_n . The design weights are the inverse of the probabilities that a population unit is contained in the survey sample. Then the R-indicator is computed as

$$R = 1 - 2S(\rho) = 1 - 2\sqrt{\frac{1}{N-1} \sum_{i=1}^n d_i (\rho_i - \bar{\rho})^2}, \quad (1)$$

with $\bar{\rho} = \frac{1}{N} \sum_{i=1}^n d_i \rho_i$ the weighted sample mean of the estimated response probabilities and N the size of the population.

Response probabilities can be estimated in the R component of the RISQ suite by either a linear or a logistic regression. The default in R is a logistic regression. In SAS response propensities are always estimated by a logistic regression. Let $X = (X_1, X_2, \dots, X_m)'$ be the vector of independent variables. X needs to be provided by the user. Main effect terms as well as interaction effect terms may be included.

4.1 Output in R

Once the response model is defined, the R-indicator can be determined;

```
> indicator <- getRIndicator(responsModel,
+   sampleData, sampleWeights, sampleStrata)
```

Properties of the sampling design, the inclusion weights and strata, can be specified by the optional arguments `sampleWeights` and `sampleStrata`. These vectors should have a length equal to the number of rows in the data frame `sampleData`. The type of sampling, simple random sampling (SI), stratified simple random sampling (STSI) or something else, is inferred from the values of `sampleWeights` and `sampleStrata`. If there is only one stratum and all inclusion weights are the same, then SI sampling is assumed. If there is more than one stratum and within each stratum the inclusion weights are the same then STSI sampling is assumed.

The return value of the function `getRIndicator` is a list called `indicator`. The most important components are

<code>sampleDesign</code>	the inferred sample design;
<code>R</code>	a bias adjusted estimate for the R-indicator; a bias-adjusted estimate will be determined if the inferred sampling design equals SI or STSI;
<code>RUnadj</code>	an estimate for the R-indicator, without any bias adjustment;
<code>SE</code>	standard error of the estimation of the R-indicator; the standard error will be determined if the inferred sampling design equals SI;
<code>prop</code>	an estimate for the response propensities;
<code>propMean</code>	the mean of the estimated response propensities which equals the response rate

The components of `indicator` can be assessed by concatenating the name of the component with a "\$" to `indicator`. The output is

```
> indicator$sampleDesign
[1] "SI"
```

```
> c(indicator$R, indicator$RUnadj, indicator$SE, indicator$propMean)
[1] 0.839748039 0.838129332 0.004107512 0.54658
```

4.2 Output in SAS

In SAS the R-indicator output is stored in the file **RISQtest.Finalfiler_ind**. The output for the RISQ-test survey has the following form

```
Obs  r_indicator  r_unadjusted  propmean    SE_r  LB_r  UB_r
1    0.85445    0.85155    0.54658  .004351777  0.84592  0.86298
```

where `r_indicator` is the adjusted R-indicator value, `r_unadjusted` is the unadjusted R-indicator, `propmean` is the response rate, `SE_r` the estimated standard error of the R-indicator and `LB_r` and `UB_r` the 95% confidence interval based on a normal approximation. We will return to the construction of the confidence intervals in section 5.

Figure 4.2.1 shows the file **RISQtest.Finalfiler_ind**. Figure 4.2.2 shows a detail of the SAS dataset with estimated response propensities.

Figure 4.2.1: *RISQtest.Finalfiler_ind* - R-indicator and confidence interval.

	bia	svvar	r_indicator	svvarb	variance_r	ci_r	LB_r	UB_r	propmean	r_unadjusted	SE_r
1	7.5698379351	0.0052962053	0.8544499356	0.0055097168	0.000018938	0.0085294823	0.8459204533	0.8629794178	0.5465808552	0.8515450661	0.0043517767

Figure 4.2.2: *RISQtest.Rindicator* - original dataset with estimated response probabilities *rphatsamp*

	gender	urb	respons	marstat	hhstype	age	responsesamp1	agea	pi	piinv	Response Value	Estimated Probability	os	rphatsamp
1	1	1	0	1	1	3	0	1	1	1	0	0.989987397	1	0.010012603
2	1	1	0	1	1	4	0	2	1	1	0	0.4491296225	1	0.5508703775
3	1	1	0	1	1	4	0	2	1	1	0	0.4491296225	1	0.5508703775
4	2	1	1	1	1	4	1	2	1	1	0	0.4315797395	1	0.5684202605
5	1	1	0	1	1	5	0	3	1	1	0	0.5371054211	1	0.4628945789
6	1	1	0	1	1	5	0	3	1	1	0	0.5371054211	1	0.4628945789
7	1	1	0	1	1	5	0	3	1	1	0	0.5371054211	1	0.4628945789
8	1	1	1	1	1	5	1	3	1	1	0	0.5371054211	1	0.4628945789
9	1	1	1	1	1	5	1	3	1	1	0	0.5371054211	1	0.4628945789
10	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
11	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
12	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
13	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
14	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
15	2	1	0	1	1	5	0	3	1	1	0	0.5193588921	1	0.4806411079
16	2	1	1	1	1	5	1	3	1	1	0	0.5193588921	1	0.4806411079
17	1	1	0	1	1	6	0	4	1	1	0	0.5711851456	1	0.4288148544
18	1	1	0	1	1	6	0	4	1	1	0	0.5711851456	1	0.4288148544
19	1	1	0	1	1	6	0	4	1	1	0	0.5711851456	1	0.4288148544
20	1	1	0	1	1	6	0	4	1	1	0	0.5711851456	1	0.4288148544

5. Bias adjustment and confidence intervals of R-indicators

R-indicators have a bias that is due to the estimation of response probabilities. In the RISQ suite the bias is approximated analytically. The standard output contains adjusted R-indicator values.

Suppose the link function h is used in the general linear model for the estimation of the response propensities ρ_i

$$\text{linear regression: } h(x^T \beta) = x^T \beta$$

$$\text{logistic regression: } h(x^T \beta) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}.$$

Hence, $h(x_i^T \beta)$ is used as a predictor for ρ_i with β a vector that is estimated. Let $\hat{\beta}$ be the estimator and ∇h be the gradient, i.e. the vector with first order derivatives with respect to β .

For simple random samples without replacement, i.e. $d_i = N/n$, the adjusted R-indicator equals

$$R_B = 1 - 2\sqrt{\left(1 + \frac{1}{n} - \frac{1}{N}\right)S^2(\rho) - \frac{1}{n} \sum_{i \in S} z_i^T \left[\sum_{j \in S} z_j x_j^T \right]^{-1} z_i}, \quad (2)$$

with $z_i = \nabla h(x_i^T \hat{\beta}) x_i$.

Since R-indicators are based on weighted sample variances of estimated probabilities, they also have a standard error and precision. The RISQ suite provides standard errors for the R-indicator. The standard errors (c.f. the previous sections on output) can be used to construct confidence intervals.

If σ_R is the estimated standard error of the R-indicator, then $[R - \xi_{1-\alpha/2} \sigma_R, R + \xi_{1-\alpha/2} \sigma_R]$ is an $100(1-\alpha)\%$ confidence interval based on a normal approximation. $\xi_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution. The estimated standard error σ_R is `indicator$SE` in R and `SE_r` in SAS..

6. Unconditional partial indicators on the variable level

The unconditional partial R-indicator measures the amount of variation of the response probabilities between the categories of a variable. The larger the between-category variation is, the stronger the relationship is and the stronger the impact of the variable on response.

As realier, let X_k be one of the components of the vector X . Suppose X_k is categorical and has H categories. Let n_h denote the weighted sample size in category h , for $h = 1, 2, \dots, H$. That means

$$n_h = \sum_{i=1}^n d_i \Delta_{h,i}, \quad (3)$$

where $\Delta_{h,i}$ is the 0-1 indicator for sample unit i being a member of stratum h . Then $n_1 + n_2 + \dots + n_H = N$.

Let $\bar{\rho}$ again be the weighted mean response probability in the sample. Furthermore, let $\bar{\rho}_h$ the weighted mean of the response probabilities in category h of X_k .

The unconditional partial indicator for variable X_k is measuring the variation between the response categories of the H categories, and is defined as

$$P_U(X_k) = \sqrt{\frac{1}{N} \sum_{h=1}^H n_h (\bar{\rho}_h - \bar{\rho})^2}. \quad (4)$$

It holds that $P_U(X_k) \leq S(\rho) \leq 0.5^2$. i.e. the total variation between categories is always smaller than the total variation. The larger the value of (4), the stronger the impact of the variable on nonresponse. By computing and comparing the unconditional partial indicators for a set of variables it can be established for which variables the relationships are strongest.

6.1 Output in R

To determine unconditional partial indicators, the optional argument `withPartials` of the function `getRIndicator` should be set to `TRUE`;

```
> indicator <- getRIndicator(responsModel, sampleData,
+   sampleWeights,
+   sampleStrata,
+   withPartials = TRUE)
```

Just as earlier, the return value `indicator` of the function `getRIndicators` contains a component `partials` containing the estimates for the partial R-indicators. The component `partials$byVariables` of the list `indicator` is a data frame with the unconditional and conditional partial indicators for each variable in the model. The data frame contains the following columns:

variable	the name of the variable;
Pu	a bias adjusted estimate for the unconditional, partial indicator; a bias-adjusted estimate will be determined if the inferred sampling design equals SI or STSI;
PuUnadj	an estimate for the partial unconditional, indicator, without any bias adjustment;

The data frame `partials$byVariables` are found by

```
> indicator$partials$byVariables

variable      Pu      PuUnadj      Pc      PcUnadj
1  gender 0.009573745 0.00967045 0.008876557 0.00896622
2    age 0.059934052 0.06053945 0.059026296 0.05962252
3    urb 0.053289584 0.05382786 0.052455829 0.05298569
```

which contains both unconditional and conditional partial R-indicators. We return to conditional partial R-indicators in section 8.

6.2 Output in SAS

In SAS the unconditional partial R-indicator is stored in the file **RISQtest.partialbetween**. The output for the RISQ-test survey

Variable Level Unconditional Partial Indicators

Obs	t	Pu	Pu_unadjusted
1	gender	0.009481	0.009670
2	age	0.059349	0.060533
3	hhstype	0.008658	0.008831
4	urb	0.035983	0.036702
5	marstat	0.016504	0.016834

² $S(\rho)$ attains its maximum value when half of the ρ_i 's are 0 and the rest are 1.

where t is the variable label, P_u is the adjusted unconditional partial R-indicator and $P_{u_unadjusted}$ is the unadjusted unconditional partial R-indicator. Figure 6.2.1 shows the contents of the output file `RISQtest.partialbetween`.

Figure 6.2.1: *RISQtest.Partialbetween* - unconditional partial indicators at the variable level.

	withinvar	betweenvar	m	withinbiasadj	betweenbiasadj	sqrtwithinbiasadj	bias	t	row	P_u	$P_{u_unadjusted}$
1	0.0054161974	0.0000935168	0.0055097142	0.0052063157	0.000089893	0.072154804	0.0002135055	gender	1	0.00948119	0.009670409
2	0.0018453152	0.0036642983	0.0055096135	0.0017738065	0.0035223015	0.0421165826	0.0002135055	agea	2	0.0593489803	0.060533448
3	0.0054317239	0.0000779908	0.0055097146	0.0052212406	0.0000749686	0.0722581522	0.0002135055	hhype	3	0.0086584401	0.0088312392
4	0.0041626786	0.0013470003	0.0055096789	0.0040013706	0.0012948027	0.0632563882	0.0002135055	urb	4	0.0359833672	0.0367015028
5	0.0052263417	0.0002833672	0.0055097089	0.0050238169	0.0002723865	0.0708788889	0.0002135055	marstat	5	0.0165041349	0.0168335132

7. Unconditional partial indicators within categories

The unconditional partial R-indicator can give more information about the relationship of a variable X_k and response behaviour if this indicator is computed for each category of X_k separately. It is clear from (4) that each category h contributes an amount

$$\frac{n_h}{n} (\bar{\rho}_h - \bar{\rho})^2 \quad (5)$$

to $P_U(X_k)$. The unconditional partial indicators within categories are obtained by taking the square root of the quantities in (5), giving

$$P_U(X_k, h) = \sqrt{\frac{n_h}{N}} (\bar{\rho}_h - \bar{\rho}). \quad (6)$$

$P_U(X_k, h)$ can assume positive and negative values. A positive value means that the particular category is over-represented. A negative value means that the particular category is under-represented.

7.1 Output in R

The component `partials$byCategories` is a list, containing the partial indicators within categories for each variable in the model. Each component in the list `partials$byCategories` is a data frame with the unconditional and conditional partial indicators within categories of a variable.

Each component of `partials$byCategories` is a data frame whose name equals the name of the variable. One example is `indicator$partials$byCategories$gender`. Most of the columns in the data frame equal the columns in the data frame `indicator$partials$byVariables`. The column `variable` is replaced by the column `category` containing the names of the categories.

```
> indicator$partials$byCategories
$gender
  category      Pu      PuUnadj      Pc      PcUnadj
1  Female  0.006758099  0.006826362  0.006268555  0.006331874
2   Male -0.006781202 -0.006849699  0.006284782  0.006348265
```

```

$age
      category      Pu      PuUnadj      Pc      PcUnadj
1  0-15 years -5.116213e-02 -5.167892e-02 0.0513946246 0.0519137627
2  15-17 years 1.748735e-02 1.766399e-02 0.0169592639 0.0171305697
3  18,19 years 2.768542e-03 2.796507e-03 0.0026033585 0.0026296551
4  20-24 years -6.409295e-03 -6.474036e-03 0.0044646787 0.0045097765
5  25-29 years -1.341988e-02 -1.355544e-02 0.0110030829 0.0111142252
6  30-34 years -3.463283e-03 -3.498266e-03 0.0024941232 0.0025193164
7  35-39 years 2.693295e-03 2.720500e-03 0.0030374423 0.0030681235
8  40-44 years 4.577895e-05 4.624136e-05 0.0003472266 0.0003507339
9  45-49 years 4.936054e-03 4.985914e-03 0.0042700934 0.0043132256
10 50-54 years -2.785296e-03 -2.813430e-03 0.0039937361 0.0040340770
11 55-59 years -2.233244e-03 -2.255802e-03 0.0035006909 0.0035360515
12 60-64 years 6.934018e-03 7.004059e-03 0.0060249980 0.0060858566
13 65-69 years 8.200488e-03 8.283321e-03 0.0074697953 0.0075452478
14 70-74 years 1.637653e-02 1.654195e-02 0.0159097009 0.0160704051
15 75 year ... 5.761386e-05 5.819582e-05 0.0002794007 0.0002822230

$urb
      category      Pu      PuUnadj      Pc      PcUnadj
1  Average 0.009982662 0.010083497 0.009954187 0.010054734
2  Little 0.016760638 0.016929938 0.016507174 0.016673913
3  Not 0.017890627 0.018071340 0.017485893 0.017662518
4  Strong -0.001583986 -0.001599985 0.002478790 0.002503829
5  Very strong -0.046223627 -0.046690533 0.045476813 0.045936175

```

7.2 Output in SAS

In SAS the unconditional partial R-indicator for a category is stored in the file **RISQtest.partial2all**. The output for the RISQ-test survey for variables gender and age

Category Level Unconditional Partial Indicators

```

Obs  gender      Pu      Pu_unadjusted
1  male  -0.006715644  -0.006849670
2  female 0.006692764  0.006826334

```

Category Level Unconditional Partial Indicators

```

Obs  age      Pu      Pu_unadjusted
1  15-17 -0.050661  -0.051672
2  18-19 0.017318  0.017664
3  20-21 0.002742  0.002796
4  22-24 -0.006348  -0.006474
5  25-29 -0.013290  -0.013556
6  30-34 -0.003430  -0.003498
7  35-39 0.002667  0.002720
8  40-44 0.000045  0.000046
9  45-49 0.004888  0.004986
10 50-54 -0.002759  -0.002814
11 55-59 -0.002212  -0.002256
12 60-64 0.006867  0.007004
13 65-69 0.008121  0.008283
14 70-74 0.016218  0.016542
15 75+ 0.000057  0.000058

```

where P_u is the adjusted unconditional partial R-indicator per category and $P_{u_unadjusted}$ is the unadjusted partial R-indicator per category. In section 10 we will explain the bias adjustment of partial R-indicators. Figure 7.2.1 contains the output stored in file **RISQtest.partial2all** for the RISQ-test data set.

Figure 7.2.1 **RISQtest.Partial2all** – all unconditional partial indicators at the category level condensed into one file.

	gender	p1zkbiasadj	p1zk	Pu	Pu_unadjusted	agea	hhstype	urb	marstat
1	1	0.0000450999	0.000046918	-0.006715644	-0.00684967
2	2	0.0000447931	0.0000465988	0.006692764	0.0068263335
3	.	0.0025665248	0.0026699908	-0.05066088	-0.051671954	1	.	.	.
4	.	0.0002999215	0.0003120124	0.0173182409	0.0176638728	2	.	.	.
5	.	7.5167766E-6	7.8198054E-6	0.002741674	0.0027963915	3	.	.	.
6	.	0.0000402912	0.0000419155	-0.006347538	-0.00647422	4	.	.	.
7	.	0.0001766341	0.0001837549	-0.013290377	-0.013555622	5	.	.	.
8	.	0.000011765	0.0000122393	-0.00343002	-0.003498476	6	.	.	.
9	.	7.1131798E-6	7.3999382E-6	0.0026670545	0.0027202827	7	.	.	.
10	.	2.0365736E-9	2.1186754E-9	0.0000451284	0.0000460291	8	.	.	.
11	.	0.0000238941	0.0000248573	0.0048881551	0.0049857114	9	.	.	.
12	.	7.6097628E-6	7.9165403E-6	-0.00275858	-0.002813635	10	.	.	.
13	.	4.8923049E-6	5.0895317E-6	-0.002211856	-0.002255999	11	.	.	.
14	.	0.0000471535	0.0000490544	0.0068668409	0.0070038871	12	.	.	.
15	.	0.0000659521	0.0000686108	0.0081210869	0.0082831649	13	.	.	.
16	.	0.0002630278	0.0002736315	0.016218133	0.0165418093	14	.	.	.
17	.	3.2349472E-9	3.3653598E-9	0.0000568766	0.0000580117	15	.	.	.
18	.	0.0000406109	0.0000422481	-0.00637267	-0.006499851	.	1	.	.
19	.	0.0000197902	0.000020588	0.0044486127	0.0045373949	.	2	.	.
20	.	4.5676906E-7	4.7518274E-7	0.0006758469	0.000689335	.	3	.	.
21	.	6.7768626E-6	7.0500575E-6	-0.002603241	-0.002655194	.	4	.	.
22	.	7.3338734E-6	7.629523E-6	0.0027081125	0.0027621591	.	5	.	.
23	.	0.000961605	0.0010003703	-0.031009756	-0.031628631	.	.	1	.
24	.	3.1757528E-6	3.3037775E-6	-0.001782064	-0.00181763	.	.	2	.
25	.	0.0000461371	0.000047997	0.0067924264	0.0069279858	.	.	3	.
26	.	0.0001575818	0.0001639344	0.0125531578	0.0128036866	.	.	4	.
27	.	0.0001263032	0.0001313949	0.0112384684	0.0114627594	.	.	5	.
28	.	0.0001761618	0.0001832634	-0.013272597	-0.013537483	.	.	.	1
29	.	0.0000719983	0.0000749007	0.0084851783	0.0086545197	.	.	.	2
30	.	0.0000239639	0.00002493	0.0048952976	0.0049929947	.	.	.	3
31	.	2.6243492E-7	2.7301444E-7	-0.000512284	-0.000522508	.	.	.	4

8. Conditional partial indicators on the variable level

Conditional partial indicators can only be computed for variables that are included in the response model. These indicators measure the relative importance of a variable, i.e. the impact of a variable conditional on all other variables in the response model. As such conditional partial R-indicators attempt to isolate the part of the deviation of representative response that is attributable to a variable alone.

The conditional partial indicator for a variable X_k is obtained by cross-classification of all model variables, but with the exception of X_k itself. Suppose, this cross-classification results in L cells U_1, U_2, \dots, U_L . Let n_l denote the weighted sample size in cell l , for $l = 1, 2, \dots, L$. Then again $n_1 + n_2 + \dots + n_L = N$. Furthermore, let $\bar{\rho}_l$ the mean of the response probabilities in cell l .

The conditional partial indicator for variable X_k is now defined as

$$P_C(X_k) = \sqrt{\frac{1}{N} \sum_{l=1}^L \sum_{i \in U_l} d_i (\rho_i - \bar{\rho}_l)^2} . \quad (7)$$

To say it in words: $P_C(X_k)$ is the remaining within cell variation of the response probabilities if the variable X_k is removed from the cross-classification. If, on the one hand, the remaining variation is large, this can apparently not be accounted for by the other variables. So, there is an important role for X_k . If, on

the other hand, the remaining variation is small, the other variables are capable of explaining the variation. It can be concluded that there need not be a role for X_k in reducing the lack of representativity.

Also here it can be remarked that $P_C(X_k) \leq S(\rho) \leq 0.5$. i.e. the total variation within categories is smaller than the total variation. And again a larger value for $P_C(X_k)$ implies a stronger conditional impact.

8.1 Output in R

To determine unconditional partial indicators, the optional argument `withPartials` of the function `getRIndicator` should again be set to `TRUE`;

```
> indicator <- getRIndicator(responsModel, sampleData,
+   sampleWeights,
+   sampleStrata,
+   withPartials = TRUE)
```

As we did for the unconditional partial R-indicators

The return value of the function `getRIndicators` contains a component `partials` containing the estimates for the partial R-indicators. The component `partials$byVariables` of the list `indicator` is a data frame with the unconditional and conditional partial indicators for each variable in the model. The data frame contains the following columns:

variable	the name of the variable;
Pc	a bias adjusted estimate for the conditional, partial indicator; a bias-adjusted estimate will be determined if the inferred sampling design equals SI or STSI;
PcUnadj	an estimate for the partial conditional, indicator, without any bias adjustment.

The output is

```
> indicator$partials$byVariables

variable      Pu      PuUnadj      Pc      PcUnadj
1  gender 0.009573745 0.00967045 0.008876557 0.00896622
2    age 0.059934052 0.06053945 0.059026296 0.05962252
3   urb 0.053289584 0.05382786 0.052455829 0.05298569
```

8.2 Output in SAS

In SAS the conditional partial R-indicator is stored in in the file **RISQtest.partialwithin**

The output for the RISQ-test survey for variables gender and age

Variable Level Conditional Partial Indicators

Obs	t	Pc	Pc_unadjusted
1	gender	0.008277	0.008442
2	agea	0.058485	0.059652
3	hhtype	0.021494	0.021923
4	urb	0.039547	0.040336

where P_c is the adjusted conditional partial R-indicator and $P_{c_unadjusted}$ the bias-adjusted conditional partial R-indicator. Figure 8.2.1 contains the output file **RISQtest.partialwithin** for the RISQ-test data set.

Figure 8.2.1: *RISQtest.Partialwithin* - conditional partial indicators at the variable level.

	withinvara	betweenvara	m	withinbiasadj	betweenbiasadj	mm	sqbetweenbiasadj	bias	t	row	Pc	Pc_unadjusted
1	0.000071275	0.005438288	0.005509563	0.0000685132	0.0052275447	0.005296058	0.0723017615	0.0002135	gender	1	0.00827727	0.0084425
2	0.003558415	0.001951247	0.005509662	0.0034205221	0.0018756342	0.0052961563	0.0433089928	0.0002135	agea	2	0.05848523	0.0596524
3	0.000480636	0.005028939	0.005509575	0.0004620101	0.0048340594	0.0052960695	0.0695274007	0.0002135	hhlype	3	0.02149442	0.0219234
4	0.001627028	0.00388258	0.005509607	0.001563978	0.0037321238	0.0052961018	0.0610911106	0.0002135	urb	4	0.03954716	0.0403364

9. Conditional partial indicators within categories

The conditional partial indicators can give even more insight when they are computed for each category of a variable separately. The remaining within cell variation of the response probabilities after removing a variable X_k from the cross-classification, is computed for each category of X_k separately. Let again X_k have H categories, labelled $h=1, 2, \dots, H$, and $\Delta_{h,i}$ be the 0-1 indicator for category h . From (7) it can be seen that each category h contributes an amount

$$\frac{1}{N} \sum_{l=1}^L \sum_{i \in U_l} d_i \Delta_{h,i} (\rho_i - \bar{\rho}_l)^2 \quad (8)$$

to $P_C(X_k)$. The conditional partial indicators within categories are then obtained by taking the square root of (8)

$$P_C(X_k, h) = \sqrt{\frac{1}{N} \sum_{l=1}^L \sum_{i \in U_l} d_i \Delta_{h,i} (\rho_i - \bar{\rho}_l)^2} . \quad (9)$$

The category-level conditional partial R-indicators are always larger than or equal to zero. A large value of (9) does not correspond to either under- or over-representation. Such an interpretation cannot be given as within some cells l the category may be over-represented while in other cells it may be under-represented. Hence, the subpopulation corresponding to a category may be overrepresented in some cells and underrepresented in others. The conditional partial indicator within a category $P_C(X_k, h)$ must be interpreted as the impact of that category on the deviation from representative response after conditioning on the other variables. The larger the indicator the larger the impact of that category and the more interesting the corresponding subpopulation becomes in nonresponse reduction methods.

9.1 Output in R

As we did for the unconditional partial indicator at the category level, we will consider the data frame `partials$byCategories`, but this time we focus on the last two columns of the data frame: `Pc` and `PcUnadj`. The component `partials$byCategories` is a list, containing the partial indicators within categories for each variable in the model. Each component of `partials$byCategories` is a data frame whose name equals the name of the variable. One example is `indicator$partials$byCategories$gender`. Most of the columns in the data frame equal the columns in the data frame `indicator$partials$byVariables`. The column `variable` is replaced by the column `category` containing the names of the categories.

```
> indicator$partials$byCategories
```

```
$gender
```

	category	Pu	PuUnadj	Pc	PcUnadj
1	Female	0.006758099	0.006826362	0.006268555	0.006331874
2	Male	-0.006781202	-0.006849699	0.006284782	0.006348265

```
$age
```

	category	Pu	PuUnadj	Pc	PcUnadj
1	0-15 years	-5.116213e-02	-5.167892e-02	0.0513946246	0.0519137627
2	15-17 years	1.748735e-02	1.766399e-02	0.0169592639	0.0171305697
3	18,19 years	2.768542e-03	2.796507e-03	0.0026033585	0.0026296551
4	20-24 years	-6.409295e-03	-6.474036e-03	0.0044646787	0.0045097765
5	25-29 years	-1.341988e-02	-1.355544e-02	0.0110030829	0.0111142252
6	30-34 years	-3.463283e-03	-3.498266e-03	0.0024941232	0.0025193164
7	35-39 years	2.693295e-03	2.720500e-03	0.0030374423	0.0030681235
8	40-44 years	4.577895e-05	4.624136e-05	0.0003472266	0.0003507339
9	45-49 years	4.936054e-03	4.985914e-03	0.0042700934	0.0043132256
10	50-54 years	-2.785296e-03	-2.813430e-03	0.0039937361	0.0040340770
11	55-59 years	-2.233244e-03	-2.255802e-03	0.0035006909	0.0035360515
12	60-64 years	6.934018e-03	7.004059e-03	0.0060249980	0.0060858566
13	65-69 years	8.200488e-03	8.283321e-03	0.0074697953	0.0075452478
14	70-74 years	1.637653e-02	1.654195e-02	0.0159097009	0.0160704051
15	75 year ...	5.761386e-05	5.819582e-05	0.0002794007	0.0002822230

```
$urb
```

	category	Pu	PuUnadj	Pc	PcUnadj
1	Average	0.009982662	0.010083497	0.009954187	0.010054734
2	Little	0.016760638	0.016929938	0.016507174	0.016673913
3	Not	0.017890627	0.018071340	0.017485893	0.017662518
4	Strong	-0.001583986	-0.001599985	0.002478790	0.002503829
5	Very strong	-0.046223627	-0.046690533	0.045476813	0.045936175

9.2 Output in SAS

In SAS the conditional partial R-indicator for a category is stored in in the file **RISQtest.partial3all**

The output for the RISQ-test survey

Category Level Conditional Partial Indicators

Obs	gender	Pc	Pc_unadjusted
1	male	.005823778	.005940008
2	female	.005881909	.005999299

Category Level Conditional Partial Indicators

Obs	age	Pc	Pc_unadjusted
1	15-17	0.050271	0.051274
2	18-19	0.017632	0.017984
3	20-21	0.003354	0.003421
4	22-24	0.005042	0.005143
5	25-29	0.011812	0.012048
6	30-34	0.003373	0.003441

7	35-39	0.004008	0.004088
8	40-44	0.001832	0.001868
9	45-49	0.005366	0.005473
10	50-54	0.004132	0.004215
11	55-59	0.004231	0.004316
12	60-64	0.005585	0.005696
13	65-69	0.006863	0.007000
14	70-74	0.015181	0.015484
15	75+	0.001708	0.001742

where P_c is the bias-adjusted conditional partial R-indicator at the category levels and $P_{c_unadjusted}$ is the unadjusted conditional partial R-indicator at the category levels. See section 10 for detail on the bias adjustment. Figure 9.2.1 shows the output file **RISQtest.partial3all** for the RISQ-test data set.

Figure 9.2.1: *RISQtest.Partial3all* - all conditional partial indicators at the category level condensed into one file.

	gender	p3zk	p3zkbiasadj	Pc	Pc_unadjusted	agea	hhstype	urb
1	1	0.0000352837	0.0000339164	0.0058237781	0.0059400082	.	.	.
2	2	0.0000359916	0.0000345969	0.0058819087	0.005999299	.	.	.
3	.	0.0026290667	0.0025271875	0.0502711399	0.0512744259	1	.	.
4	.	0.0003234337	0.0003109002	0.0176323635	0.0179842613	2	.	.
5	.	0.000011701	0.0000112476	0.0033537422	0.0034206745	3	.	.
6	.	0.0000264485	0.0000254236	0.0050421856	0.005142815	4	.	.
7	.	0.0001451593	0.0001395343	0.0118124618	0.0120482089	5	.	.
8	.	0.0000118382	0.0000113795	0.0033733509	0.0034406745	6	.	.
9	.	0.0000167081	0.0000160607	0.004007579	0.0040875603	7	.	.
10	.	3.4896633E-6	3.354435E-6	0.0018315117	0.001868064	8	.	.
11	.	0.0000299538	0.0000287931	0.0053659158	0.005473006	9	.	.
12	.	0.0000177645	0.0000170761	0.0041323243	0.0042147951	10	.	.
13	.	0.0000186237	0.000017902	0.0042310796	0.0043155214	11	.	.
14	.	0.000032448	0.0000311906	0.0055848516	0.0056963112	12	.	.
15	.	0.0000489981	0.0000470994	0.0068628991	0.0069998653	13	.	.
16	.	0.0002397457	0.0002304553	0.015180754	0.0154837238	14	.	.
17	.	3.0354591E-6	2.9178317E-6	0.0017081662	0.0017422569	15	.	.
18	.	0.0000316881	0.0000304601	0.0055190653	0.0056292141	.	1	.
19	.	7.9472358E-6	7.639266E-6	0.0027639222	0.0028190842	.	2	.
20	.	0.0000392883	0.0000377658	0.0061453898	0.0062680387	.	3	.
21	.	0.0000627583	0.0000603263	0.0077670035	0.0079220163	.	4	.
22	.	0.0003389536	0.0003258185	0.0180504433	0.0184106918	.	5	.
23	.	0.0011663873	0.001121188	0.0334841453	0.0341524123	.	.	1
24	.	0.0000161063	0.0000154822	0.0039347367	0.004013265	.	.	2
25	.	0.0000730511	0.0000702203	0.0083797529	0.0085469936	.	.	3
26	.	0.000214751	0.0002064291	0.0143676407	0.0146543859	.	.	4
27	.	0.0001567321	0.0001506585	0.0122743024	0.0125192694	.	.	5

10. Bias adjustment and confidence intervals of partial R-indicators

As for the R-indicators, partial R-indicators have a bias and standard error. In the current version of the RISQ suite the approximation of standard errors is not included. As a consequence, the partial R-indicators need to be handled with some care. Especially, category-level partial R-indicators may have a low accuracy as the number of sample units in the corresponding cell may be small.

The SAS program does allow for an approximation of the standard error using resampling methods. In order to obtain confidence intervals for the partial indicators, bootstrapping can be carried out. The latter part of the programme (currently closed under the comments `/* */`) can produce these confidence intervals. The bootstrapping carries out 500 replicates and calculates the variance and confidence intervals (the 2,5 and 97,5 percentiles) for each of the partial indicators that are calculated in each replicate. The number of replicates can be increased to gain more precision. Computation times may be considerable.

In the RISQ suite the bias of partial R-indicators is adjusted by prorating the overall R-indicator bias over the partial R-indicators. That means that the estimated bias of the variance of response probabilities $B(S^2(\rho))$ is multiplied by the ratio between the square of the partial R-indicator and $S^2(\rho)$. This approximation is motivated by the fact that the partial R-indicators are between and within variances which are components of the total variance of response probabilities $S^2(\rho)$. The resulting, prorated bias is then subtracted from the between variance (unconditional partial R-indicators) or the within variance (conditional partial R-indicators). And the partial R-indicators are computed by taking the square root of the adjusted between or within variance.

Let $S_{W,unadj}^2(\rho)$ and $S_{B,unadj}^2(\rho)$ denote, respectively, the unadjusted within variance and the unadjusted between variance of the estimated response propensities. Both variance terms are adjusted for bias in the following way

$$S_W^2(\rho) = S_{W,unadj}^2(\rho) - B(S^2(\rho)) \frac{S_{W,unadj}^2(\rho)}{S^2(\rho)} \quad (10)$$

$$S_B^2(\rho) = S_{B,unadj}^2(\rho) - B(S^2(\rho)) \frac{S_{B,unadj}^2(\rho)}{S^2(\rho)} \quad (11)$$

and the adjusted partial R-indicators at the variable level are computed by taking square roots. The bias adjustment for the category level indicators is analogous, i.e. the bias is prorated according to the relative size of the variance with respect to the overall variance.

11. Maximal absolute bias and maximal absolute contrast

In RISQ deliverables R-indicators are interpreted in terms of the impact of nonresponse on survey estimation by considering the standardized bias of the design-weighted response mean \hat{y}_r of survey items y

$$\frac{|B(\hat{y}_r)|}{S(y)} = \frac{|Cov(y, \rho_Y)|}{\bar{\rho}S(y)} = \frac{|Cov(y, \rho_{\aleph})|}{\bar{\rho}S(y)} \leq \frac{S(\rho_{\aleph})}{\bar{\rho}} = \frac{1 - R(\aleph)}{2\bar{\rho}}, \quad (12)$$

with $\bar{\rho}$ the average response propensity and \aleph the vector of auxiliary variables explaining response behaviour. The vector \aleph is unknown and, as a consequence, we do not know ρ_{\aleph} . Since we are interested in the general representativeness of a survey, i.e. not the representativeness with respect to single survey items, we use as an approximation for (12)

$$B_m(X) = \frac{1 - R(\rho_X)}{2\bar{\rho}}. \quad (13)$$

B_m represents the maximal absolute standardized bias under the scenario that non-response correlates maximally to the selected auxiliary variables. ρ_X are the estimated response propensities with a response model based on X .

Additionally, the RISQ deliverables consider the maximal contrast between respondents and non-respondents. The contrast for a variable Y is the expected difference between the response mean and nonresponse mean of that variable. The bias of the response mean can be rewritten as the product of the non-response rate $1 - \bar{\rho}$ and the contrast.

$$B(\hat{y}_r) = (1 - \bar{\rho})(E(\hat{y}_r) - E(\hat{y}_{nr})).$$

Hence, the maximal absolute standardized contrast is defined as the maximal absolute standardized bias divided by the non-response rate. We denote it by $C_m(X)$

$$C_m(X) = \frac{1 - R(\rho_X)}{2\bar{\rho}(1 - \bar{\rho})}. \quad (14)$$

B_m and C_m are referred to as the maximal bias and the maximal contrast. Both measures are not in the standard output of `RISQ_R-indicators` but can easily be computed by

$$B_m = \frac{1 - R}{2\text{propmean}}$$

$$C_m = \frac{1 - R}{2\text{propmean}(1 - \text{propmean})}$$

The R-indicator, the maximal bias and the maximal contrast provide means to evaluate the quality of response

12. General guidelines to R-indicators and partial R-indicators

The following recommendations must be kept in mind when using the R-indicators and partial R-indicators:

- R-indicators and partial R-indicators cannot be evaluated or presented separately from the variables X that were used in the response model and should always be presented together with X .
- When comparing different surveys, one should use the same response model, where the variables X , have the same categories.
- R-indicators should be adjoined by a confidence interval in order to indicate the uncertainty due to the estimation based on a sample.
- The inclusion of response-unrelated variables into the response model leads to an increase of the standard errors of R-indicators. It is recommendable to restrict analysis to variables X for which it is known from the literature that they relate to response behaviour.
- R-indicators measure the distance to a fully representative response; they do not reflect the impact of non-response on the bias of (weighted) means or the contrast of survey variables, and nor does the response rate. The maximal absolute bias combines the response rate and the R-indicator and is designed to make comparisons of non-response bias under worst case scenarios. The maximal absolute contrast does the same for the contrast under worst case scenarios.

The various indicators may be used to compare different surveys or a single survey in time. When comparing different surveys, we recommend to fix a number of sets of auxiliary variables beforehand (including interactions) and to add all variables to the models. One should restrict to demographic and socio-economic characteristics that are generally available in many surveys.

When comparing a survey in time, we recommend to fix a number of sets of auxiliary variables. However, now the sets may also include variables that correlate to the main survey items, and variables that relate to the data collection (paradata). When many variables are available, parsimonious models may be favoured.

Partial R-indicators provide insight that is helpful in the reduction of nonresponse. We provide the following simple guidelines:

- In the comparison of different surveys, partial R-indicators are supplementary to R-indicators. Response models are simple and employ general auxiliary variables only.

- In the comparison of a survey in time, partial R-indicators are again supplementary to R-indicators. Response models may be more complex, e.g. define multiple model equations or levels, and may employ paradata additionally to auxiliary variables.
- Conditional partial R-indicators should be used in conjunction with unconditional partial R-indicators. They are always smaller than the unconditional partial R-indicators and comparing the two shows to what extent the apparent impact of a single variable is taken away by the others.
- When many variables are added to models for response, then conditional partial R-indicators naturally are smaller. When two or more variables are included that correlate strongly, then the conditional partial R-indicators will be small for both variables. It is recommendable not to include many related variables.

As a general guideline we conclude with the remark that in improving representativity of response it must always be the objective to increase the response rate and to decrease the R-indicators simultaneously.

12. Visualising R-indicators in R-cockpit

Partial R-indicators are easier to interpret when they are visualised. The R-cockpit program developed in the project RISQ is a graphical tool that enables a quick and easy display of both unconditional and conditional R-indicators. R-cockpit is available at the RISQ website www.risq-project.eu. It is written in R and assumes that the survey data set is converted to R. With the program an R function called *export.R* is provided that executes export of SPSS and SAS data files to R. We refer to the R-cockpit manual for further details.

13. Future releases of RISQ_R-indicators in SAS and R

Future releases of RISQ_R-indicators are planned. In the autumn of this year a second release will be provided on www.risq-project.eu that contains analytic standard error approximations for both unconditional and conditional partial R-indicators as set out in

- Shlomo, N., Skinner, N., Schouten, B., Carolina, N., Morren, M. (2009), Partial indicators for representative response, RISQ deliverable 4

Spring 2011 a third release is planned that includes population-based R-indicators. Population-based R-indicators measure representativeness based on population counts and population tables only. They widen the scope of the indicators to settings where samples cannot be linked to administrative data. Population-based R-indicators are discussed in

- Shlomo, N., Skinner, C., Schouten, B., Heij, V. de, Bethlehem, J., Ouwehand, P. (2009), Indicators for representative response based on population totals, RISQ deliverable 2.2