

Institute of Public Health of the Republic of Slovenia

**Eurostat Grant 2007: Support for the implementation of the
European Health Interview Survey**

FINAL REPORT

Report prepared by:

- Darja Lavtar, IPH
- Tina Zupanič, IPH
- Daša Moravec Berger, IPH
- Jana Trdič, IPH

Ljubljana, December 2008

CONTENTS

1 INTRODUCTION 3

2 ACTIVITIES CARRIED OUT 4

3 RESPONSE / NON-RESPONSE 8

4 CALCULATION OF WEIGHTS10

**5 DESCRIPTION OF DIFFICULTIES IN THE IMPLEMENTATION OF THE MODULES
AND RECOMMENDATION FOR THE SECOND ROUND OF EHS13**

6 ANNEXES14

1 INTRODUCTION

Institute of Public Health of the Republic of Slovenia (IPH) implemented the European Core Health Interview Survey (EHIS), where the national linguistic version of the EHIS questionnaire for the 1st Round¹ was adopted. As the continuation of the EHIS implementation, the Eurostat grant 2007 on Support for the implementation of the European Health Interview Survey was carried out.

The process of support activities on implementation of the European Health Interview Survey in the year 2008 was done by the schedule and carried out in three stages.

In the first stage from January 2008 to March 2008 the supervision of the completed questionnaires was done; numeric codes to variables, such as occupation, were assigned; the work of the interviewers was precisely supervised and the data entry begun.

In the second stage from April 2008 to June 2008 the data entry was done. The data entry was followed by data editing, non-response analysis, weights calculation and the data file in SPSS was prepared.

In the third stage from July 2008 to October 2008 the tabulation of data and data analyses were done and the final report prepared.

¹ Available at CIRCA (<http://forum.europa.eu.int/Public/irc/dsis/Home/main>) > Public Health Statistics > Library > Methodologies and data collections > Health Interview Survey > European Health Interview Survey (EHIS) 2007-2008 methodology > EHIS questionnaire. 1st round 2007-2008.pdf

2 ACTIVITIES CARRIED OUT

As already described in the introduction, the process of support activities on implementation of the European Health Interview Survey in the year 2008 was done by the schedule and carried out in three stages. Since we used the PAPI questionnaires the exact supervision of the completed questionnaires was first done.

For the purposes of supervision questionnaires two kinds of material were prepared already during the Eurostat grant 2006 “Implementation of the modules on health determinants, health care and background variables for the European Health Interview Survey (EHIS)”, *Conceptual manual for controlling and coding the questionnaires* and *Technical manual for controlling the questionnaires*. Both manuals were used by the supervisors in the process of supervision of the questionnaires. Also the database which was developed during the Eurostat grant 2006 in order to have accurate review over the state of the data collection phase was used to accurately keep a record over the returned questionnaires at IPH and their state in the process of supervision. Whenever detail revision over the supervision phase could be done, by using of PivotTable and PivotChart report.

Each questionnaire was equipped with label composed of several dates and statuses which were fulfilled in the process of the supervision (see the label below).

Return date: _____
Takeover date: _____
Questionnaire status: _____
Takeover date (quest. needs completion): _____
Takeover date in the archives: _____
Date of entering data in the base: _____

Supervisors used the Conceptual manual for controlling and coding the questionnaires which contains also some general directions for supervisors’ for the supervising process and directions about variables which need to be coded and where can the code books be found. They also used Technical manual for controlling the questionnaires which contain detailed directions for controllers, on particular questions (variables) where possible mistakes could be done in during the fieldwork phase. The supervisors also used so-called “Control paper” in which all the mistakes from the interviewers’ and if happen marks of questions that need to be completed were registered, upon which the control coordinator made the completion of the questionnaire. Each interviewer got an estimation of particular questionnaire (that was one of the factors that influenced upon interviewers’ payment). All the needed documents were available to the supervisors on the network (including both manuals, control paper, code books,...).

In the process of supervision the supervisors need to assign numeric codes to some variables, such as occupation, economic sector, country of birth and citizenship. For the purposes of assigning numeric codes to above mentioned variables several codification books were available at the network, such as ISCO-88 for occupation,


NACE Rev.2 (2 digits) for economic sector, internal code book for country of birth and citizenship (see the attached Annex 1).

In case the questionnaire needed to be completed, or when during the supervision some mistakes or missing values were discovered, supervision coordinator contacted the respondent in order to get the insufficient data. Almost all the mistakes or missing values were completed during the supervision.

After the interview each respondent received the followed-up letter (control letter) upon which we estimated the quality of the interviewers work and respondents' perceptions of the survey. Answers were entered into the database which was analysed afterward. Upon analysis we have done the estimation of the quality of the interviewers work.

Figure 1: Example of Control letter

INŠTITUT ZA VAROVANJE ZDRAVJA
REPUBLIKE SLOVENIJE



IVZ RS

INSTITUTE OF PUBLIC HEALTH OF
THE REPUBLIC OF SLOVENIA

Spoštovani!

Pred kratkim je Inštitut za varovanje zdravja Republike Slovenije izvajal anketo, v kateri smo prebivalce in prebivalke Slovenije spraševali o zdravju in zdravstvenem varstvu.

V tem času vas je obiskal naš sodelavec (anketar) in anketo skupaj z vami tudi izpolnil. Za vaše sodelovanje se vam iskreno zahvaljujemo. Kot slučajno izbrana oseba ste namreč omogočili vpogled v nekatera področja zdravja in zdravstvenega varstva.

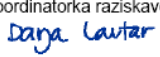
Ker se želimo prepričati o kakovosti izvedbe naše ankete, bi radi preverili korektnost dela naših anketarjev, zato vas vljudno prosimo, da odgovorite na naslednja kratka vprašanja, in sicer tako, da pri vsakem vprašanju obkrožite ustrezen odgovor.

Ali vas je nedavno obiskal naš pooblaščen anketar?	Da.				Ne.
Ali je anketiral vas osebno?	Da.				Ne.
Ali vas je anketiral (tudi) po telefonu?	Da, v celoti.	Delno po telefonu.			Ne.
Kako bi (z oceno od 1 do 5) ocenili anketarjevo korektnost in obnašanje?	1 zelo neprimerno	2	3	4	5 zelo primerno
Koliko časa vam je vzelo sodelovanje z anketarjem?	Do pol ure.	Od pol ure do ene ure.		Več kot eno uro.	
Bi nam želeli še kaj sporočiti?					


Vljudno vas prosimo, da pismo vložite v priloženo ovojnico in to čim prej odpošljete.

Za sodelovanje se vam najlepše zahvaljujemo!

Darja Lavtar
koordinatorka raziskave



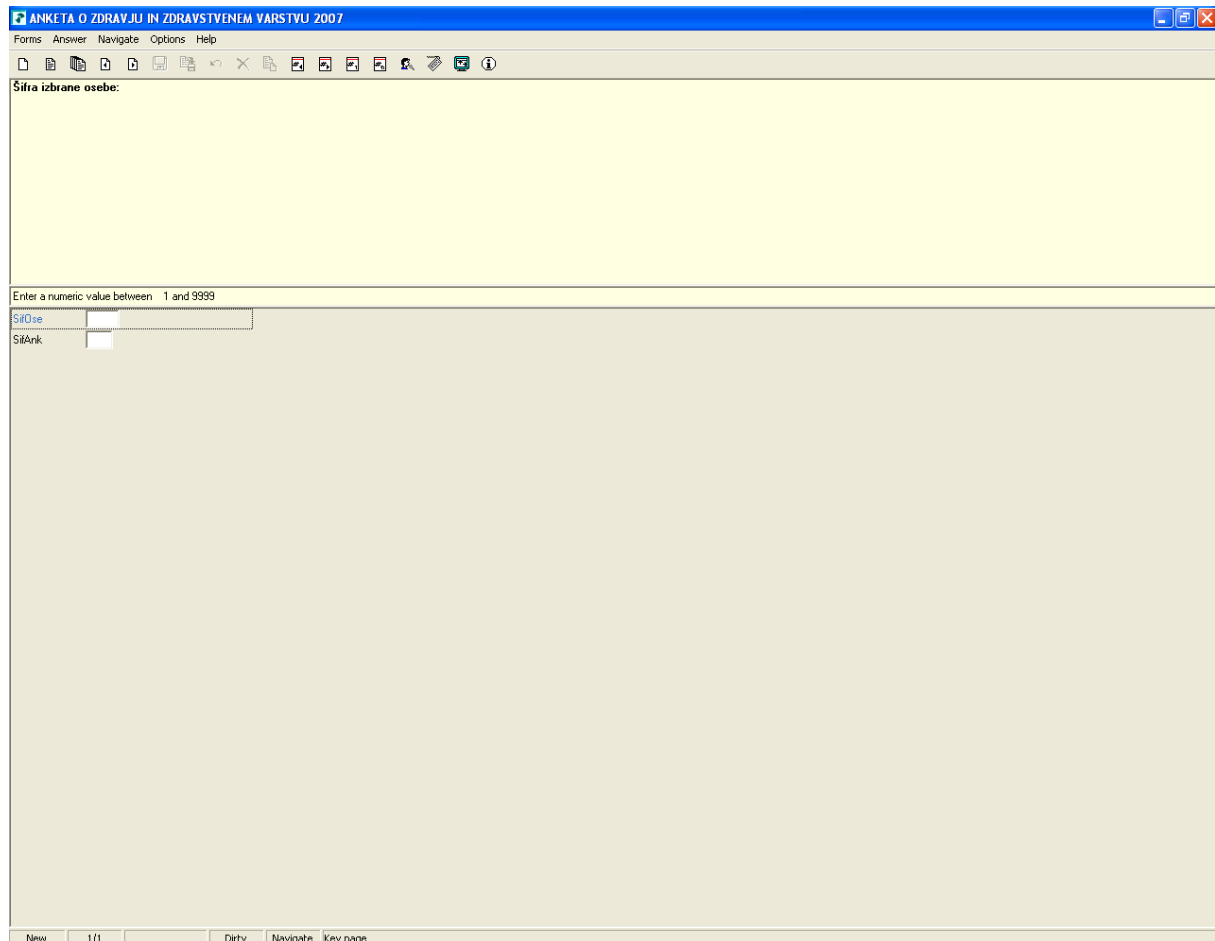
Ada Hočevnar Grom
v. d. direktorice



After the supervision process was done, each questionnaire was brought to archive where it waited until the data entering phase begun. The data entering phase take

place between March and April 2008. The data entry programme was developed upon final version of main and both self-completion questionnaires in Blaise for Windows software, which is a powerful and flexible system for computer assisted survey processing. All the logical controls from the printed questionnaires were integrated into the data entry programme. All the interviews were entered into the database (two different databases were made, in order to get two separated databases, one with respondents and one with non-respondents, to get an accurate overview over the state of the survey).

Figure 2: Screenshot of the first page of the data entry programme



When the data entering phase was done, the data were exported from Blaise to SPSS where some basic data editing was done (finding the duplicates, falsely entered questionnaires e.g. respondents into the database of non-respondents and the other way round, etc.). After that in SPSS 16.0 some other data editing procedures, using SPSS Data Preparation were done. Several rules were applied to the data. First we need to define »single-variable« validation rules and then »cross-variable« validation rules. After that the defined validation rules were applied to particular variables to validate the data. During the data validation process first the allowed maximum percentage of missing values, maximum percentage of cases in a single category and maximum percentage of categories with count of 1 were applied to all the variables. The data editing procedure flags the IDs' of incomplete cases and also the duplicate IDs. The indicator variables which present a complete record of individual validation rule violations were saved as individual variables, where each

variable corresponds to an application of a validation rule and has a value of 1 if the case violates the rule and a value 0 if it does not. No cases in our database were marked as duplicate, so all the cases were primary cases. After the data validation was done, the case report with all rule violations was precisely revised and then the data entry corrected. Also the identification of unusual cases was done; upon which anomalous cases were identified (these cases needed further examination).

Anytime during the data collection phase and also later in other subsequent processes of EHIS implementation, the response rate could easily be computed using *response categories* respectively *status of the returned questionnaire* for each individual questionnaire included in the database developed in order to have accurate review over the state of the data collection phase and subsequent processes. Each questionnaire which came to IPH was entered into this database, where the *return date*, *status of the returned questionnaire* (“the questionnaire is entirely performed”, “the questionnaire is partially performed”, “refusal”, “no contact”, “unsuitability”), *date of delivering questionnaire to the supervisor*, *date of delivering the supervised questionnaire to the archive*, *date of sending the control letter*, *date of the return of the control letter*, *evaluation of the interviewer from the control letter* and *date of the data entering* were entered into the database for each questionnaire, what enable simple and quick overview over different processes of the EHIS implementation (anytime also the state of supervision and data entering could be observed using this database, as all the data for each questionnaire were daily entered into the database).

After the data entering and data editing the calculation of weights was done. First the *personal weight* for the individual person was calculated using Bascula, the software package inside the Blaise for Windows, and by using the application CALMAR which is an application inside the SAS and uses generalized calibration for non-response adjustment the *household weight* was calculated.

Afterward the data file in SPSS with labels in accordance with European Health Interview Survey Guidelines for the input file List of variables², was prepared. From July to October 2008 data tabulation and different analyses were done and the final report on the implementation of the 1st round of the EHIS in Slovenia prepared.

² Available at CIRCA (http://circa.europa.eu/Members/irc/dsis/healthtf/library?l=/technicalsgroupshis/2008_15-16_2008/tech-his-2008-47-1/_EN_1.0_&a=d > Health Interview Survey > Library > Technical Group HIS > 2008: 15-16 May 2008 > Tech-HIS-2008-4.7-1 - Guidelines for data delivery)

3 RESPONSE / NON-RESPONSE

Based upon the *status of the returned questionnaire* (“non-contact”, refusal, not eligible, completed interview, partial interview), the response rate was calculated. The table below presents the final status of the 1st round of the EHIS in Slovenia.

Table 1: Final status of the survey

Status of the returned questionnaire	Total sum of cases	Total percentage among all cases	Total percentage among eligible cases
Non-contact	202	6%	6%
Refusal	798	24%	26%
Not eligible	268	8%	
Complete interview	2117	62%	68%
Partial interview	2	0%	0%
Total	3387	100%	100%
Eligible cases	3119		
RESPONSE RATE	68%		

Table 2: Calculation of response, cooperation rates

Category	EHIS 2007
I=Complete Interviews	2117
P=Partial Interviews	2
R=Refusal and break off	798
NC=Non Contact	202
O=Other	0
RESPONSE RATES	
Response Rate 1	
$I/(I+P) + (R+NC+O)$	0,679
Response Rate 2	
$(I+P)/(I+P) + (R+NC+O)$	0,679
COOPERATION RATES	
Cooperation Rate 1	
$I/(I+P)+R+O)$	0,726
Cooperation Rate 2	
$(I+P)/((I+P)+R+O))$	0,726
Cooperation Rate 3	
$I/((I+P)+R))$	0,726
Cooperation Rate 4	
$(I+P)/((I+P)+R))$	0,726

Table 3: Calculation of refusal and contact rates

Category	EHIS 2007
REFUSAL RATES	
Refusal Rate 1	
$R/((I+P)+(R+NC+O))$	0,256
Refusal Rate 2	
$R/((I+P)+(R+NC+O))$	0,256
Refusal Rate 3	
$R/((I+P)+(R+NC+O))$	0,256
CONTACT RATES	
Contact Rate 1	
$(I+P)+R+O / (I+P)+R+O+NC$	0,935
Contact Rate 2	
$(I+P)+R+O / (I+P)+R+O+NC$	0,935
Contact Rate 3	
$(I+P)+R+O / (I+P)+R+O+NC$	0,935

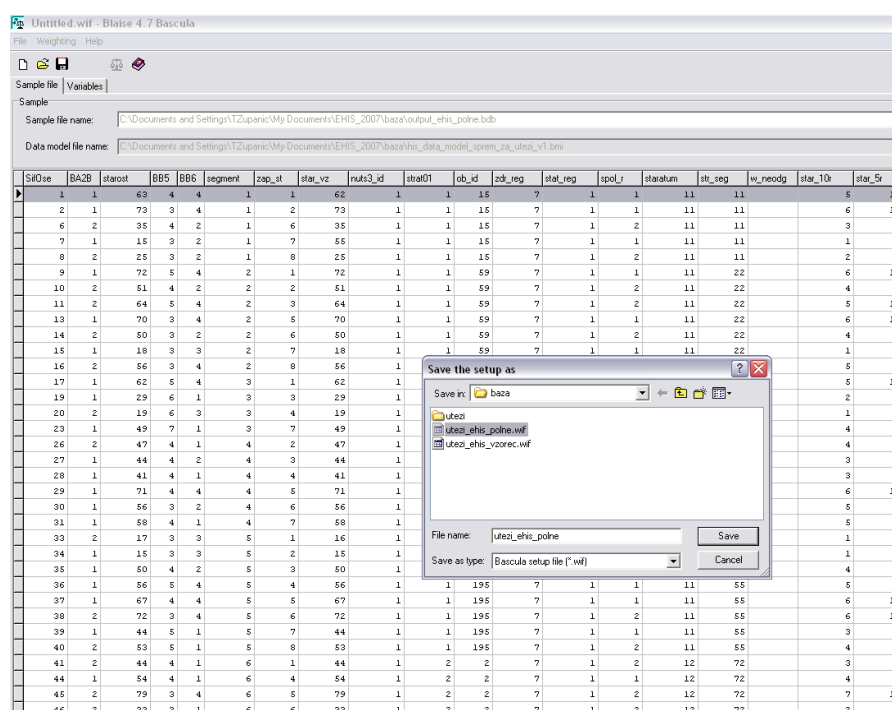
Based upon survey categories, we calculated several response, cooperation, refusal and contact rates based upon AAPOR methodology: *Standard Definitions* (Final Dispositions of Case Codes and Outcome Rates for Surveys; Revised 2008). The **minimum response rate** of European Health Interview Survey 2007 in Slovenia was **68 %**; the **minimum cooperation rate** was **73 %** (a cooperation rate is the proportion of all cases interviewed of all eligible units ever contacted); the **refusal rate** was **26 %** (a refusal rate is the proportion of all cases in which a housing unit or respondent refuses to do an interview, or breaks-off an interview of all potentially eligible cases); and **contact rate** which assumes that all cases of indeterminate eligibility are actually eligible was **94 %** (a contact rate measures the proportion of all cases in which some responsible member of the housing unit was reached by the survey).

4 CALCULATION OF WEIGHTS

After the data entering and data editing the calculation of weights was done. The weighting was done in cooperation with the Statistical Office of the Republic of Slovenia.

First the **personal weight** for the individual person was calculated using Bascula, the software package inside the Blaise for Windows. Bascula is a software package for calculating weights for all units in a sample using auxiliary information. The auxiliary information is gathered in a weighting model, which forms the basis for the weighting procedure. Several weighting methods are supported, most of them based on the general regression estimator.

Figure 3: Screenshot of Bascula



We computed the weight based upon several auxiliary variables: *gender*, *five years age groups*, *type of the settlement* (for all so far mentioned auxiliary variables the population values based upon Central Register of the Population were used) and *education* (due to the fact the information about individuals' education is not included into Central Register of the Population the estimated population values for people age 15 years and older was obtained form Labour Force Survey in Slovenia, 2007). The weight on selection probability into the sample (prepared by Statistical Office of the Republic of Slovenia) was also considered in the weighting process. Due to the fact that post-stratification could not be used as weighting method, as we did not had information in all cells in the cross-classification of all auxiliary variables, we used an alternative approach to develop weights that make several marginal totals in the sample conform to the corresponding marginal population totals, called *linear weighting*. It is based on the general regression estimator and is justified by a regression relationship between the target variable y and the auxiliary vector x . For a

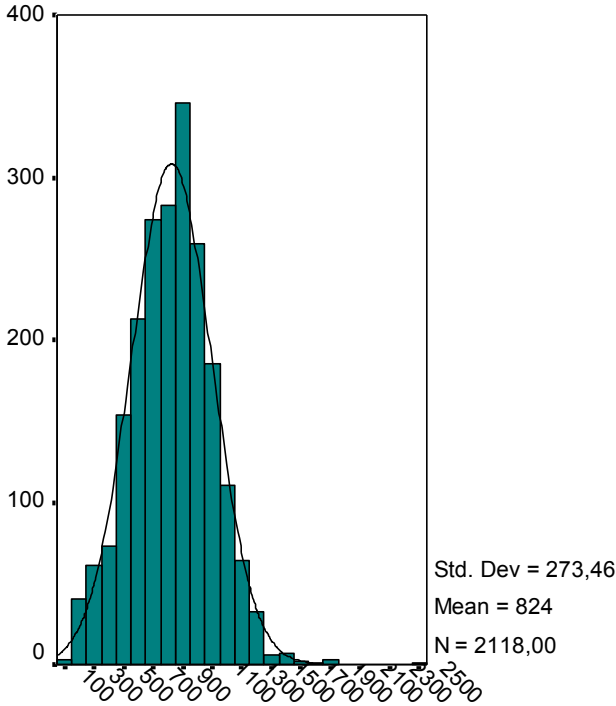
sufficiently large sample the regression estimator is (approximately) unbiased. It is possible to express the regression estimator as the (regression) weighted sample sum of the target variable, namely

$$\hat{t}_{yR} = \sum_{k \in S} w_k y_k .$$

The weights w_k do not depend on the target variables. This means that once a set of weights has been computed the total of any target variable can be estimated. Linear weighting provides a general framework with both post-stratification and ratio estimation.

The figure below presents the distribution of personal weight after the linear weighting was done.

Figure 4: Final (personal) weights distribution after linear weighting



The **household weight** was computed by several steps:

First the *sample weight* was calculated for each household that responded the number of people age 15 years and older was counted (os15), where the formula for calculation weight for individual household is $w_{gosp} = \frac{1}{os15}$ (after standardization upon number of households we got the weight w_{gosp_n}).

Secondly the *weight due to non-response* was calculated upon stratum level. First the whole sample was distributed as regarding response, refusal and non-eligible as

regarding the value of the variable *status of the returned questionnaire*. This weight was calculated as $w_neodg = \frac{odgovor + neodgovor}{odgovor}$.

Then the final weight was calculated, as product of sample weight and weight due to non-response, as $hhwgt = w_gosp_n \cdot w_neodg$.

After that the data were prepared for calibration. Due to the fact that the data about all members of particular household were available, the data on gender age structure were considered in calibration. Also the data about distribution as regarding type of the settlement and statistical region based upon Statistics on Income and Living Conditions 2007 (this survey is also the reference for the total number of households for the year 2007, which was 694741) were considered in calibration.

Finally, by using the application CALMAR in SAS the calibration was executed upon population values presented in earlier step, of which result was the ***household weight***.

5 DESCRIPTION OF DIFFICULTIES IN THE IMPLEMENTATION OF THE MODULES AND RECOMMENDATION FOR THE SECOND ROUND OF EHIS

The following difficulties encountered when implementing EHIS modules:

- The interviewers' recruitment presented the biggest problem and no permanent base of the interviewers' could be established, as even very serious and experienced students are not permanent staff for interviewing (some other ways of interviewers' recruitment should be considered in the future).
- In the case of question HS.7 (our mark CC1) answer "Not applicable" should be possible at least in case of Accident at work and Accident at school, for people that do not work and do not attend school, as these two questions are unnecessary for them.
- In case of questions related to different diseases or conditions (questions HS.4, HS.5 and HS.6 – our marks CB1, CB2 and CB3) the different sorting of questions or different logical controls should be considered, as there exist possibilities in case a person has a particular disease, which was not diagnosed by a medical doctor for the disease or condition to have it in the past 12 month (as the questions are formulated at present no such conditions which appeared in the last 12 month could be recognized).
- The respondents hardly differentiate physical activities (vigorous, moderate, walk) – in some cases enormous numbers of hours doing separate activity occurred.

Some useful recommendations could be derived from the EHIS modules implementation:

- Based upon pilot study which was conducted in the beginning of 2007, where proxy respondents were allowed only in case of severe health problems and permanent absence of the selected person and where proxy was used only in two cases, we decided not to allow proxy, as it is not suitable to be used especially in case of subjective questions.
- Based upon feedback information from the interviewers, the implementation of both self-completion questionnaires in the main questionnaire should be considered, as frequently respondents asked interviewers to interview them also on both self-completion questionnaires (it looks like, that these questions are not so sensitive for Slovenian residents).
- The two colours that were used in the questionnaires were very much approved by the interviewers, as they contribute to better look over the questionnaires (the detailed logical controls used in the questionnaires and guidelines to individual questions, which were marked with the green colour were also very much approved by the interviewers).

6 ANNEXES

Annexed to this final report are:

1. Annex 1: The codification scheme used (in Slovene);
2. Annex 2: Frequency tables describing the results of the survey at national level;
3. Annex 3: Description of the survey according to EHIS quality categories and criteria developed by Partnership on Public Health Statistics.