COMMISSION REGULATION (EC) No 28/2004

of 5 January 2004

implementing Regulation (EC) No 1177/2003 of the European Parliament and of the Council concerning Community statistics on income and living conditions (EU-SILC) as regards the detailed content of intermediate and final quality reports

THE COMMISSION OF THE EUROPEAN COMMUNITIES,

Having regard to the Treaty establishing the European Community,

Having regard to Regulation (EC) No 1177/2003 of the European Parliament and the Council of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC) [1], and in particular Article 15(2)(b) thereof,

Whereas:

(1) Regulation (EC) No 1177/2003 established a common framework for the systematic production of Community statistics on income and living conditions, encompassing comparable and timely cross-sectional and longitudinal data on income and on the level and composition of poverty and social exclusion at national and European Union levels.

(2) Pursuant to Article 15(2)(b) of Regulation (EC) No 1177/2003, implementing measures are necessary to determine the detailed content of the intermediate report, relating to the common cross-sectional EU indicators based on the EU-SILC cross-sectional component, and the detailed content of the final quality report focusing on internal accuracy, which covers both cross-sectional and longitudinal components.

(3) The measures provided for in this Regulation are in accordance with the opinion of the Statistical Programme Committee,

HAS ADOPTED THIS REGULATION:

Article 1

The definitions to be applied to the intermediate and final quality reports of the Community statistics on income and living conditions (EU-SILC) shall be as laid down in Annex I.

Article 2

Quality evaluation criteria and the detailed content of the intermediate quality report to be produced by Member States, relating to the common cross-sectional EU indicators based on the cross-sectional component of the EU-SILC shall be as laid down in Annex II.

Article 3

Quality evaluation criteria and the detailed content of the final quality report to be produced by Member States, relating to the EU-SILC cross-sectional and longitudinal components focussing on internal accuracy shall be as laid down in Annex III.

Article 4

The content of the comparative intermediate and final quality reports to be produced by the Commission (Eurostat) shall be as laid down in Annex IV.

Article 5

This Regulation shall enter into force on the 20th day following that of its publication in the *Official Journal of the European Union*.

This Regulation shall be binding in its entirety and directly applicable in all Member States.

Done at Brussels, 5 January 2004.

For the Commission
Pedro SOLBES MIRA
Member of the Commission

[1] OJ L 165, 3.7.2003, p. 1.

ANNEX I

DEFINITIONS

(a) 'Substitution' refers to replacement by other units of the original units selected in the sample which do not supply the required information, either because the address cannot be located or is inaccessible, or because the household refuses to cooperate, the entire household is temporarily away or the household is unable to respond

(b) 'Imputation' means estimating plausible (but artificial) substitute values for missing values

(c) 'Equivalised disposable income' is defined as the household's total disposable income divided by its 'equivalent size'

(d) 'Equivalent size' refers to the OECD modified scale (which gives a weight of 1,0 to the first adult, 0,5 to other persons aged 14 or over who are living in the household and 0,3 to each child aged under 14)

(e) 'Sampling frame' is the population of units from which a sample can be selected

(f) 'Accuracy' denotes the closeness of computations or estimates to the exact or true values

(g) 'Sampling errors' refers to the variability that occurs at random because of the use of a sample rather than a census

(h) 'Non-sampling errors' are errors that occur in all phases of the data collection and production process

Non-sampling errors are basically of four types:

— 'Coverage errors' are errors due to differences between the target population and the sampling frame. Coverage errors include over-coverage, under-coverage and misclassification:
  — 'over-coverage' relates either to wrongly classified units that are in fact out of scope, or to units that do not exist in practice
  — 'under-coverage' refers to units not included in the sampling frame
  — 'misclassification' refers to incorrect classification of units that belong to the target population

— 'Measurement errors' are errors that occur at the time of data collection. There are a number of sources for these errors, such as the survey instrument, the information system, the interviewer and the mode of collection

— 'Processing errors' are errors in post-data-collection processes such as data entry, keying, editing and weighting

— 'Non-response errors' are errors due to an unsuccessful attempt to obtain the desired information from an eligible unit. Two main types of non-response errors are considered:
  — 'unit non-response' refers to the absence of information of the whole units (households and/or persons) selected into the sample
  — 'item non-response' refers to the situation where a sample unit has been successfully enumerated, but not all the required information has been obtained

(i) 'Relevance' is the degree to which statistics meet current and potential users' needs. It refers to whether all statistics that are needed are produced and the extent to which concepts used (definitions, classifications, etc.) reflect users' needs

(j) 'Timeliness and punctuality':

— 'Timeliness of information' reflects the length of time between its availability and the event or phenomenon it describes

— 'Punctuality' refers to the time lag existing between the actual Eurostat delivery date of data and the target date when it should have been delivered, for instance, with reference to dates announced in some official release calendar, laid down by regulations or previously agreed among partners

(k) 'Accessibility and clarity':

— 'Accessibility' refers to the physical conditions in which users can obtain data: where to go, how to order, delivery time, clear pricing policy, convenient marketing conditions (copyright, etc.), availability of micro or macro data, various formats (paper, files, CD-ROM, internet), etc.

— 'Clarity' refers to the data's information environment whether data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality also available (including limitation in use) and the extend to which additional assistance is provided by the National Statistical Institutes

ANNEX II

Quality evaluation criteria and content of the intermediate quality report to be produced by Member States

1. COMMON CROSS-SECTIONAL EUROPEAN UNION INDICATORS

1.1. **Common cross-sectional EU indicators based on the cross-sectional component of EU-SILC**

Members States will provide the common cross-sectional EU indicators, based on the cross-sectional sample of year *N*, which will be included in the annual spring report of year (*N* + 2) to the European Council.

The common cross-sectional EU indicators refer to those indicators adopted by the Council in the context of the open method of coordination, which can be derived from the EU-SILC instrument.

1.2. **Other indicators**

1.2.1. *Equivalised disposable income*

1.2.2. *The unadjusted gender pay gap*

Members States that compute the unadjusted gender pay gap indicator on the basis of EU-SILC will provide that indicator.

2. ACCURACY

2.1. **Sample design**

The following information will be provided:

2.1.1. *Type of sampling design (stratified, multi-stage, clustered)*

2.1.2. *Sampling units (one stage, two stages)*

2.1.3. *Stratification and substratification criteria*

2.1.4. *Sample size and allocation criteria*

2.1.5. *Sample selection schemes*

2.1.6. *Sample distribution over time*

2.1.7. *Renewal of sample: rotational groups*

2.1.8. *Weightings*

2.1.8.1. Design factor

2.1.8.2. Non-response adjustments

2.1.8.3. Adjustments to external data (level, variables used and sources)

2.1.8.4. Final cross-sectional weight

2.1.9. *Substitutions*

For Members States where substitution is applied in cases of unit non-response, the following information will be provided:

2.1.9.1. Method of selection of substitutes

2.1.9.2. Main characteristics of substituted units compared to original units, by region (NUTS 2) if available

2.1.9.3. Distribution of substituted units by record of contact at address (DB120), household questionnaire result (DB130) and household interview acceptance (DB135) of the original units

2.2. **Sampling errors**

2.2.1. *Standard error and effective sample size*

The following information will be provided:

— effective sample size for the common cross-sectional EU indicators based on the cross-sectional component of EU-SILC, for equivalised disposable income and for the unadjusted gender pay gap (if applicable),

— standard errors for the common cross-sectional EU indicators based on the cross-sectional component of EU-SILC, for equivalised disposable income and for the unadjusted gender pay gap (if applicable).

2.3. **Non-sampling errors**

2.3.1. *Sampling frame and coverage errors*

A description of the sampling frame (including information on the procedure used to update the frame, frequency and duplicates), and a description of the main coverage problems (misclassification, under-coverage and over-coverage), if available, will be provided.

For Member States using a rotational design, information on the sampling frame will be provided for the new replications only.

2.3.2. *Measurement and processing errors*

2.3.2.1. Measurement errors

The following information will be provided:

— a description of different sources of measurement errors likely to be found in the survey,

— a description of the way the questionnaire was built up, the use of a cognitive laboratory (if applicable), field testing of the questionnaire, the effect of its design, content and wording,

— information on the intensity and efficiency of interview training: number of training days, skills testing before starting the fieldwork (rate of success and so on),

— information on studies, such as re-interviews, record check studies, or split-sample experiments, if available,

— results from models, for instance to assess the impact of using a financial year instead of a calendar year, if available.

2.3.2.2. Processing errors

The following information will be provided:

— a description of data entry controls, coding controls and the editing system applied to the data. Main errors detected in the post-data-collection process,

— the rates of failed edits for income variables.

2.3.3. *Non-response errors*

The following information will be provided:

2.3.3.1. Achieved sample size

— number of households for which an interview is accepted for the database. Rotational group breakdown (if applicable) and total,

— number of persons of 16 years or older who are members of the households for which the interview is accepted for the database, and who completed a personal interview. Rotational group breakdown (if applicable) and total,

— number of selected respondents (if applicable) who are members of the households for which the interview is accepted for the database, and who completed a personal interview. Rotational group breakdown (if applicable) and total.

2.3.3.2. Unit non-response

For Member States using a rotational design, information on unit non-response will be provided for the new replications in accordance with the formulas described below.

For the total sample, the unit non-response will be calculated by removing, from the numerator and the denominator of the formulas described below, those units that according to the tracing rules are out of scope.

— Household non-response rates (*NRh*) will be computed as follows:

NRh = (1-(Ra * Rh)) * 100

Where:

$$\text{Ra} = \frac{\text{Number of addresses successfully contacted}}{\text{Number of valid addresses selected}} = \frac{\sum[\text{DB120} = 11]}{\sum[\text{DB120} = \text{all}] - \sum[\text{DB120} = 23]}$$

Ra is the address contact rate

$$\text{Rh} = \frac{\text{Number of household interviews completed and accepted for database}}{\text{Number of eligible households at contacted addresses}} = \frac{\sum[\text{DB135} = 1]}{\sum[\text{DB130} = \text{all}]}$$

Rh is the proportion of complete household interviews accepted for the database

DB120 is the record of contact at the address

DB130 is the household questionnaire result, and

DB135 is the household interview acceptance result.

For those Members States where substitutions are made in case of unit non-response, non-response rates will be calculated before and after substitutions.

— Individual non-response rates (*NRp*) will be computed as follows:

NRp=(1-(Rp)) * 100

Where:

$$\text{Rp} = \frac{\text{Number of personal interviews completed}}{\text{Number of eligible individuals in the households whose interviews were completed and accepted for the database}} = \frac{\sum[\text{RB250} = 11 + 12 + 13]}{\sum[\text{RB245} = 1 + 2 + 3]}$$

Rp is the proportion of complete personal interviews within the households accepted for the database

RB245 is the respondent status, and

RB250 is the data status.

— Overall individual non-response rates (**NRp*) will be computed as follows:

* NRp = (1-(Ra * Rh * Rp)) * 100

For those Members States where substitutions are made in cases of unit non-response, non-response rates will be calculated before and after substitutions.

For those Members States where a sample of persons rather than a sample of households (addresses) was selected, the individual non-response rates will be calculated for 'the selected respondent' (RB245 = 2), for all individuals aged 16 years or older (RB245 = 2 + 3) and for the non-selected respondent (RB245 = 3).