

Issues in data quality and comparability in EU-SILC

Vijay Verma¹
University of Siena

ABSTRACT

This paper develops and discusses a framework for the assessment of statistical quality in EU-SILC, with focus on comparability as a central dimension of quality. We view data quality as a multidimensional concept, covering not only statistical accuracy but also the relevance, timeliness, comprehensiveness, etc., of the data. There is a broad agreement on what dimensions make up the overall quality of statistical data, and we briefly review these dimensions, noting some relationships between them. All countries participating in EU-SILC produce national Quality Reports. We review them briefly since they provide valuable information on data quality, including comparability.

Comparability, as one of the dimensions of data quality, is a particularly important dimension in the context of a EU-wide undertaking such as EU-SILC. In ECHP, for instance, comparability was achieved through a standardised design and common technical and implementation procedures, with centralised support of the national surveys by Eurostat. The survey structure and implementation arrangements are more diverse in EU-SILC; we identify how this diversity makes the problem of comparability more complex and acute in EU-SILC.

We begin by clarifying what "comparability" means and how it may be achieved in practice, and identify the basic characteristics and requirements of EU-SILC in their effect on comparability. Comparability means the extent to which the results for different countries can be put together, compared, and interpreted in relation to each other and against common standards. An assessment of how far such comparability has been achieved in practice requires us to examine the data and procedures both from the "input" and the "output" sides. The former involves an analysis of the methodology and implementation of the process of production: how the data were collected, statistically treated, processed, and analysed. The latter involves a comparison of the substantive results actually obtained with appropriate standards such as alternative data sources, prior knowledge, and logical expectations. Both these aspects in the assessment of comparability are important.

On the basis of this framework, the paper proceeds to identify a number of specific aspects where problems of comparability are likely to arise in EU-SILC, and elaborates methodologies for the study of some of the more important ones among them.

¹ Dipartimento di Metodi Quantitativi, Università degli Studi di Siena, P.za S.Francesco 8,
53100 Siena, Italy.verma@unisi.it.

Contents

1	Data quality: a multidimensional concept.....	3
1.1	Quality dimensions	3
1.2	Relationship between different aspects of quality.....	6
1.3	Quality of EU-SILC Country Quality Reports	6
2	Comparability: a central requirement	9
2.1	What is comparability.....	10
2.2	Why comparability?	11
2.3	How comparability is achieved.....	11
3	Inter-country comparability of EU-SILC data.....	12
3.1	Lessons from ECHP.....	12
3.2	Characteristics and requirements of EU-SILC affecting comparability	13
4	Assessment of achieved comparability.....	15
4.1	Comparison of the production processes	16
4.2	Comparison of the substantive results	16
5	Potential sources of non-comparability in EU-SILC.....	17
5.1	Some potential sources of non-comparability in EU-SILC	17
5.2	Example of studies in comparability.....	20
(1)	Assessment of the comparability of income distribution by component.....	20
(2)	The incidence and treatment of negative and zero incomes	22
(3)	Impact on comparability of differences in the definition of household and household membership.....	24

1 Data quality: a multidimensional concept

A comprehensive assessment of data quality requires its diverse dimensions to be taken into account. Various organisations have developed their own specific lists of ‘quality dimensions’, but they all have a great deal in common. Essentially, they all share the view that “improving statistical quality” means “increasing the utility of statistical products and services for the community of their users”. Quality may be defined in terms of user needs as the “totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs”. Or more simply, especially in relation to statistics, as “fitness to use” for the purpose for which the data were created.²

Below is an illustration of overlapping concepts and categories used by different organisations to identify dimensions of quality.³

1.1 Quality dimensions

Canada	Netherlands	R. of Korea	IMF	Eurostat
			Prerequisites of quality	
Relevance	Relevant	Relevance		Relevance
Accuracy	Accurate	Accuracy	Accuracy and reliability	Accuracy
Timeliness	Timely	Timeliness	Serviceability	Timeliness and Punctuality
Accessibility		Accessibility	Accessibility	Accessibility and clarity
Coherence				Coherence
		Comparability	Methodological soundness	Comparability
Interpretability				
			Integrity	Completeness
	Cost-effectively	Efficency		
	Without too much a burden			

1 Relevance and use of the data

Relevance refers to the capacity of the data to meet users’ needs. It implies the identification of users and their needs, and assessment of the extent to which their needs are actually met. The concept also covers the *potential* of the data in meeting the relevant needs.

EU-SILC builds on the experience of ECHP, with a similar in scope and content. Judging from the enormous amount of academic and policy-relevant research which has been conducted using ECHP data, one can expect EU-SILC also to prove highly relevant for the purpose for which the instrument has been created. Perhaps so even to a greater extent than ECHP, because the experience of ECHP has contributed towards the creation of an improved instrument in the form of EU-SILC.

² Juran J.M., Gryna F.M. (1970). *Quality Planning and Analysis*. McGraw Hill.

³ This table is taken from

Lee D., Shon A. (2001). Korea’s experiences in statistical quality assessment. *Proceedings, Statistics Canada Symposium “Achieving data quality in a statistical agency: a methodological perspective”*.

This expectation by no means diminishes the need for continued assessment of the actual performance of EU-SILC, preferably on the basis of well-designed *user satisfaction surveys*, enumerating periodically at least the most important users.

2 Timeliness and punctuality

Punctuality refers to adherence to a pre-established time schedule for the release of statistics. Timeliness is a more objective criterion, assessing how fresh are the data and whether they became available when most needed. Punctuality acquires increased importance in the EU-wide context, in so far as the data are needed and used in a comparative context, and hence for many countries simultaneously.

On the other hand, it is important to note that the requirements of timeliness can conflict with those of accessibility and clarity, and above all with those of accuracy. At a minimum, the data must be checked and corrected to a high standard before their public release. We may call this “elementary quality control”. Obviously, releasing data or results without adequate editing and correction can be misleading and wasteful. *It can also damage the credibility of the producer organisation.*⁴

The requirements of punctuality have been expressed very strongly in EU-SILC Regulations. This is because of the desire (and the need) to improve upon the rather poor performance of ECHP in this respect. We will learn in the near future whether an appropriate balance has been reached in EU-SILC between the somewhat conflicting requirements of punctuality on the one hand, and pre-release quality control on the other.

3 Data accuracy

It has been customary in survey practice to focus on accuracy at the expense of – even to the exclusion of – other dimensions of quality. Even though some people may consider accuracy to be the most important and central aspect of data quality, it is essential to pay close attention to other dimensions of quality as well.

Just as quality is multi-dimensional, we may also view accuracy – one of the dimensions of quality – to be itself multidimensional. Firstly, it is useful to distinguish between two broad categories of error: errors of measurement and errors of estimation. *Errors of measurement* refer to the difference between the measurements and the actual (true) values for the given set of observation units. *Errors of estimation* refer to the biases and uncertainties involved in going from the observed sample to the whole target population. Next, within each of these broad types, there are errors arising from different sources: e.g., conceptual, interviewing (measurement, observation), respondent and processing errors under ‘measurement errors’; and coverage, sample selection, non-response and sampling error under ‘errors of estimation’.

Evaluation of the accuracy of survey data requires an assessment of the magnitude of errors arising from all these sources. In practice, in most surveys the assessment is limited to the more important and/or the more easily measured components. EU-SILC country data producing organisations no doubt collect a great deal of relevant information on different types of errors in their data, but we may expect serious limitations in the extent to which such detailed information can be made available in the public domain.

4 COMPARABILITY

Comparability is increasingly considered a most central requirement of data quality. This dimension of quality is particularly important in the context of a multidimensional

⁴ Ivan Fellegi identifies credibility as a ‘survival’ issue for a statistical organisation.

undertaking such as EU-SILC, where *issues relating to comparability underscore all aspects of data quality*. For instance, for most of the important uses of EU-SILC, the data can be relevant only if they are comparable across Member States, and also over time. Data for all countries need to be released at the same time: in international programmes it has often been a problem that the slowest country can delay the whole. Indications of accuracy (such as response rates) need to be defined and computed following identical procedures. And so on.

The survey structure and implementation arrangements are more diverse in EU-SILC, compared with ECHP for instance; this has implications for comparability. I will discuss diverse aspects related to comparability in the remaining sections of this paper.

5. Coherence with other statistics and over time

This refers to consistency with other sources providing similar and related information. *Consistency does not necessary mean identity*: often there are genuine and inherent differences in the information coming from sources of different types. What it means is whether different sources together lead to a consistent picture, with each making a contribution towards the development of the picture. In the case of the EU-SILC, the most relevant sources for external comparison include national household budget and labour force surveys, national accounts, and various administrative and other sources depending on the country.

In a panel (and in fact in any continuing survey) coherence over time is also a fundamental requirement. Only under this condition can we study trends, aggregate data over time, or construct micro-level longitudinal measures.

6. Accessibility and clarity

These aspects refer to the extent to which the statistical data are available in the form and under conditions which meet users' requirements, and to how well the data are described and documented for the purpose. 'Conditions of availability' include a whole range of factors such as restrictions on who can or cannot get access to the data, what items of information are suppressed, what restrictions apply on the conditions and purposes of data use, and also the difficulties, delays and the costs involved in gaining access to the micro data.

Well documented EU-SILC micro data became available for research very soon after they were ready for the purpose. *This a remarkable achievement* and, in my view, represents a major advance in the ESS. There are of course shortcomings, such as the suppression of a little too much information in the name of confidentiality, and the price charged for data use which can be rather high especially for academic researchers. Documentation and provision of meta data also needs improving and constant updating. Some of the most important requirements include the establishment of sufficiently high 'minimum standards' for data checking, a system of 'data alerts' warning users about data problems and limitations, and an actively interacting community of users.

7. Other quality aspects

A variety of other aspects are also covered in the various data quality frameworks.

Some of these overlap – different terms indicating more or less the same thing, perhaps from a somewhat different point of view or with a somewhat different emphasis. We have, for instance, 'completeness' in Eurostat terminology, 'integrity' in that of IMF and, with a little difference, 'interpretability' in that of Statistics Canada. Cost efficiency and minimisation of respondent burden are other aspects included as quality dimensions,

especially in national frameworks (The Netherlands, South Korea in the table). Surprisingly, not all frameworks explicitly refer to ‘comparability’ as a dimension.

1.2 Relationship between different aspects of quality

No quality dimension is an all-or-nothing property: each is a matter of degree.

To a certain extent, the *different dimensions of data quality compete against each other*, an obvious example being the common conflict between timeliness and data accuracy – ‘quickly released but rough data, versus refined data but much delayed’. And of course, statistically the most accurate data are not necessarily the most comparable. Often there is a clear trade-off between different dimensions of data quality – one dimension being enhanced at the expense of the other. The optimum choice of statistical design and procedures often takes the form of a balance between different quality requirements. *The appropriate balance is always specific to the particular circumstances and objectives, also the type and the particular uses of the data.* In some situations, punctuality for instance is absolutely critical, while in others it may hardly matter against the objective of producing robust and reliable information of long term utility.

It is not only conflict and competition which exists between quality dimensions, however. Even more importantly, different aspects of data quality can also *mutually support and reinforce each other*, one often forming a precondition for the other. For instance, it is hardly possible for two data sets to be comparable, when either or both lack statistical accuracy.

I would like to propose the following model as a tool for thinking about the issue more concretely.

Reduction in data quality in any dimension involves a loss in the utility of the information, a loss which may be more or less steep depending on the particular context. Often the resources saved by reducing quality in one dimension can be used to improve quality in other dimensions; however, some dimensions can also be linked in such a way that a quality loss in one dimension necessarily implies a loss in the other as well.

Beyond a certain point, there is likely to be a critical zone when further reduction in quality along a particular dimension would result in increasing drastically the loss in the overall utility of the data. A certain minimum degree of quality has to be present in *every* dimension for the statistical information to remain useful overall.

Hence the relationship between different aspects of data quality is a complex one, and is highly context specific. Practical understanding of the relationship between different dimensions of data quality on the one hand, and on the other the relationship between them and the different components of data costs, is what I would term as “statistical wisdom”.

Appropriate balances have to be sought for EU-SILC as well.

1.3 Quality of EU-SILC Country Quality Reports

National Intermediate Quality Reports for the EU-SILC surveys of 2004 have been produced by countries on the basis of the requirements specified in Framework Regulation (EC) No 1177/2003, Article 16, Commission Regulation (EC) No 28/2004,

and the technical elaboration provided in document EU-SILC 132/04 on intermediate and final quality reports.⁵

Below we provide, very briefly, a summary assessment of the completeness and quality of the Quality Reports.⁶ Of course, this assessment of the *reporting* of quality is not an assessment of the *quality* itself of the survey data. The latter type of assessment is a broader undertaking, and requires examining the survey procedures and processes, internal consistency of the data, their plausibility, and comparability of the results with other national sources, and also with similar data from other countries. Nevertheless, *an assessment of the reporting of quality forms an important part of the broader task of assessing data quality*; in fact the former may be considered the first step of the latter.

This is because the process of data quality assessment has two distinct aspects: evaluating the *survey process* through which the data are generated; and evaluating the *estimates and inferences* produced from the data. The survey process comprises the methodology of the survey, as well as the implementation of that methodology in all its aspects. We can assess data quality not only by evaluating the estimates and inferences from the data; a great deal is also learnt from examining the survey process (design, procedures, implementation, ...). From logic and experience, we often know that certain types of procedures and practices are more likely to yield good statistics than other types. At least we know that certain types of procedures are likely to produce data of poor quality, and it is critical in quality assessment to know whether such procedures, or some better alternatives, have been used. The national Quality Reports are designed primarily *to document the survey process* in the above sense. By reporting some important indicators computed from the survey data and comparing them with other sources, the national Quality Reports also provide some more direct information on data quality.

It is also useful to put together that information for a number of countries, so as to permit an examination of the national design and procedures in a *comparative context*. This greatly facilitates their assessment, especially for the purpose of assessing comparability in EU-SILC data.

It is our assessment that, overall, the national Quality Reports are of high quality. Most countries have provided most of the information requested by Eurostat, following systematically the sections and subsections specified in that document. However, in a few cases the country quality reports lack sufficient detail.

Common EU indicators

Nearly all the required Common EU indicators (at-risk-of poverty rate before and after social transfers, share ratio S80/S20, relative median at-risk-of poverty gap, Gini coefficient, unadjusted gender pay gap) have been reported in the Quality Reports. A priori, the values of the various indicators appear generally plausible. EU-SILC data are used for estimating the gender pay gap by about half of the countries (the other using alternative sources). It is possible that the values of this index are rather sensitive to outliers in EU-SILC data, and are less reliable than other indicators constructed from the same source.

⁵ The countries participating in the 2004 survey included: Austria, Belgium, Denmark, Estonia, Spain, Finland, France, Greece, Ireland, Iceland, Italy, Luxembourg, Norway, Portugal, Sweden.

⁶ These remarks are based on: Verma V. and co-researchers (2006). EU-SILC Quality of Quality Reports. Document prepared for Eurostat.

Survey structure

Most countries have adopted the 4-year rotational integrated design recommended by Eurostat for situations in which EU-SILC is based on a new survey. Modified designs have been used in some countries, primary for the purpose of integrating EU-SILC with an existing survey (e.g., Sweden, Finland), and/or incorporating into EU-SILC an existing sample (e.g., Norway, Denmark, Luxembourg). In France, the structure is the same as the standard integrated design, except that a much longer panel duration is used. In fact, apart from France, all other countries – including the ones mentioned above not using the standard design – have used 4-year panels. This of course is the minimum panel duration required for EU-SILC.

Sample selection and design

All samples are probability sample by design. In almost all countries, up-to-date sampling list are available for sample selection. The number of households or selected respondents interviewed varies from around 3,000 (Iceland) to 24,000 (Italy). In practically all cases, direct sampling of households/persons, or two-stage area samples have been used.

A serious shortcoming of the Quality Reports concerns the lack of information provided on *design effects* (the efficiency of the sample design used, compared to that of a simple random sample of households or persons). Standard errors for the common EU cross-sectional indicators have been reported only by half of the countries. More information is highly desirable.

Unit non-response and substitution

As noted, all samples are probability sample by design. *However, in some cases, high non-response rates and/or substitution may have damaged this desirable property of the sample.* Overall unit non-response rate for the personal interview varies from the low of 12% (Greece) to the high of 52% (Belgium).

Only two countries have chosen to allow substitution for selected sample units which cannot be enumerated: Spain and Ireland. In both cases, insufficient information has been provided on the substitution procedures used, and – much more importantly – on any measures applied to control the extent or manner of substitution. This is a major and potentially very damaging shortcoming.

Item non-response

Mostly, the requested information on the rates of item non-response has been provided in the Quality Reports. Where information on income components is obtained from registers, it can be assumed that there is no item non-response for these variables. Among the 'survey countries', the rates of unit response for self-employment income (perhaps the most problematic variable in this respect) varies from essentially zero in Greece and France (with very low values also in Estonia, Portugal, Italy and Ireland), to as high as 65% in Belgium, with high values also recorded in Spain (24%), Austria (38%), and Luxembourg (45%). The wide variation in item non-response rates is potentially a major factor adversely affecting comparability of the data. Note also that the 'total' response rate for an item is even lower – being the product of unit and item non-response rates considered separately.

Mode of data collection

There are surprising variations in mean interview duration, from 18 minutes in Norway to 60 minutes or more in France, Greece, Italy, Portugal and Sweden. The reason for this variation is not apparent. Interview by telephone has been the choice in Nordic countries, with the exception of Sweden where the report expresses a strong preference

for normal face-to-face interviewing. In the remaining countries, there is almost equal division between PAPI and CAPI.

In some countries (such as Estonia, Greece and Norway) individual interviewing by proxy has been largely avoided, but it is disturbingly high in a number of other cases: reaching 25% in Finland, France and Italy; and reaching or exceeding 30% Ireland and Spain. Proxy interviewing is likely to damage data quality, especially given the nature of the data collected in EU-SILC personal interview.

The quality Reports also provide wealth of information on data collection and processing procedures.

Comparability and coherence

The Quality Reports are particularly rich in this area: detailed information has been provided in most cases on the definition of income components, noting any departure from the standards, even when such departures are only minor.

Coherence, in the form of comparisons with external sources, is admirably detailed in a few country reports (such as Spain and Greece), but limited or absent in a number of other reports.

2 Comparability: a central requirement

The survey structure and implementation arrangements are more diverse in EU-SILC, compared with ECHP for instance. In the rest of the paper, I try to identify how this diversity makes the problem of comparability more complex and acute in EU-SILC.⁷

We begin by clarifying what "comparability" means and how it may be achieved in practice, and identify the basic characteristics and requirements of EU-SILC in their effect on comparability. Based on this framework, we identify a number of specific aspects where problems of comparability are likely to arise in EU-SILC, and elaborates methodologies for the study of some of the more important ones among them.

⁷ See for instance:

Verma V. (2004). Comparability of statistics at the international level: concepts, approaches, methods. *Invited Lecture* at Comparabilité, harmonisation et intégration de données dans la construction de systèmes statistiques. Neuchatel: Swiss Statistical Society, Section of Official Statistics.

Verma V. (2002). Comparability in Multi-country Survey Programmes. *Journal of Statistical Planning and Inference*, vol. 102(1), pp. 189-210.

Verma V. (2002). Comparability in International Survey Statistics. Keynote Address, *International Conference on Improving Surveys*, Copenhagen, 25-28 August.

Verma V. (1998). Robustness and comparability in income distribution statistics. Invited paper, *European Union High Level Think-Tank on Poverty Statistics*, Stockholm.

Verma V. (1998). Data sources and access for comparative analyses. In *Information Dissemination and Access in Russia and Eastern Europe* (ed. Walker, R. and Taylor, M.). Amsterdam: IOS Press.

Verma V. (1997). Comparability in multi-country survey programmes. *American Statistical Association Joint Statistical Meetings*, Special Session in Memory of Professor P.V. Sukhatme, Anaheim, California, USA.

Verma V. (1993). Comparative surveys in Europe: problems and possibilities. *Bulletin of the International Statistical Institute*, vol. 55.

Comparability is increasingly becoming one of the central dimensions of data quality. It is particularly important in the context of a EU-wide undertaking such as EU-SILC. The need for genuinely comparable data arises not only because it is important in itself, but also because such data give Member States and the EU the possibility of bench-marking and defining best practices in terms of social and economic policy. Apart from its obvious relevance at the EU level, comparable information is also invaluable for policy at the national level, as it helps each country to judge its place relative to others in the EU.

Diverse issues concerning comparability have been discussed in-depth by the current author, and these will provide a framework for the present discussion.

To begin with, it is useful to explain the concept of 'comparability': (1) what comparability is; (2) how comparability may be achieved in practice, and the role of standardisation; and (3) how we may assess the extent to which comparability has or has not been achieved. This defines the framework for understanding and evaluating the performance of EU-SILC in this respect.

2.1 What is comparability

Comparability of statistical data, i.e. their usefulness in drawing comparisons and contrast among different populations, is a complex concept, difficult to assess in precise or absolute terms independently of specific objectives of analysis. Nevertheless, it is a fundamental requirement for any data to be used in multi-population comparisons and contrasts, and is the major rationale in the launching surveys such as the ECHP and EU-SILC.

While comparability may defy precise definition, it is an important and useful concept. It is a relative concept: we can only have 'degrees of comparability', not absolute comparability. Furthermore, the same data may be sufficiently comparable for some purposes, but not so for others.

By comparability we mean that data (estimates) for different populations (whether countries or different groups within the same country) can be legitimately (i.e. in a statistically valid way) put together (aggregated), compared (differenced), and interpreted (given meaning) in relation to each other and against some common standards. Comparability permits results for these subpopulations to be (1) aggregated to construct the total picture; (2) contrasted to study differentiation; and (3) even for individual subpopulations, given meaningful interpretation, which can be done only on the basis of shared concepts, definitions and classifications. A degree of comparability is the essential basis for these purposes.

In an intuitive sense, comparability between different data sets (or different sources of information) implies that, in certain essential respects, the data measure the 'same thing'. The data sets lack comparability if they provide measures of different things (different concepts, phenomena, objects, reference periods, etc.) – even if each source provides the most accurate measure of the thing *it* refers to. Hence comparability as a dimension of quality is distinct from data accuracy.⁸ Yet two data sets cease to be comparable if they are subject to measurement errors of different magnitudes and types, and in any case if such errors are large in either. Hence an 'adequate' level of accuracy is essential for comparability.

⁸ Of course, it is appropriate that EU-SILC encourages the Member States to use the best national sources in collecting income data, whether they be from surveys or registers. But for the reason noted above, sometimes this choice may not be the best for comparability!

2.2 Why comparability?

Internationally comparable information is needed because of the countries' need to:

- o assess their place in relation to other countries;
- o increase the scope for learning from others' practices and join in co-operative ventures; and take advantage of financial and technical support from international programmes;
- o international and bilateral agencies need comparable information for their programmes and policies;
- o a major impetus is also provided by researchers who are increasingly looking for internationally comparable datasets.

2.3 How comparability is achieved

In relation to the basic requirements for generating comparable data, a distinction can be drawn between the measurement and estimation aspects of a data generation system.

Measurement aspects. These concern obtaining information on the particular set of units included in the study, such as a given sample of households and persons. These aspects include definition of concepts, variables and survey population; methods of measurement and data collection; and the related substantive analysis. These should be strictly standardised so as to control (make similar) biases of measurement in the comparisons.

Estimation aspects. These concern drawing conclusions about the population which the observed units are meant to represent. These include sampling frames, sample size and design, many operational aspects, as well as weighting, estimation and other aspects of statistical analysis. Generally, these have to be chosen flexibly to suit the conditions and requirements of individual populations in the comparison. *What is required is not identical procedures, but the common standards to be followed.*

Comparability requires control of the measurement aspects so as to ensure that the same type of information is obtained. In principle, the estimation aspects can be chosen flexibly without affecting comparability, as long as valid and common standards are followed.

Standardisation. In addition, there are in practice often powerful reasons for aiming at standardisation and control of many details in systems aimed at generating comparable data, going well beyond the development and provision of common concepts, definitions, statistical instruments and procedures, and the main statistical outputs. This is especially useful when existing technical capability of co-operating institutions is uneven, especially if some of the institutions have inadequate capacity. Standardisation is a useful tool for ensuring that conditions for comparability are actually met. There is often also a considerable economy of effort in adopting a common package of procedures and tools, in contrast to custom-designing for each case.

3 Inter-country comparability of EU-SILC data

3.1 Lessons from ECHP

A comparison with the situation of ECHP is very instructive in appreciating the issues relating to comparability as they are likely to arise in EU-SILC.

In ECHP comparability was achieved through a standardised design and common technical and implementation procedures, with centralised support and co-ordination of the national surveys by Eurostat. This included: (1) common concepts, definitions and classifications; (2) the use of a common 'blue-print' questionnaire, which served as the point of departure for the national surveys; (3) a common survey structure and procedures; (4) common sampling requirements and standards, coupled with flexibility in the actual designs to suit national conditions; (5) common standards and arrangements for data processing and statistical analysis; (6) the creation of standardised microdata sets - this was a crucial element of data comparability in practice; and (7) achieving the above in practice through centralised support and co-ordination of the national surveys by Eurostat.

Concerning (5), common standards and arrangements for data processing and statistical analysis, we should comment on two important aspects, namely imputation and weighting.

Imputation for missing income components:

Household income is composed of diverse components, and incomplete information for the construction of total income is a major problem in surveys. Complex and comprehensive procedures are required for good quality imputation of missing income components. In ECHP imputations were confined to missing income components and were done *centrally by Eurostat*. For this purpose a comprehensive methodology was adopted, following discussions at the ECHP Working Group and Task Forces. The method used was a variant of the estimation-maximisation (EM) algorithm, with the algorithm and programs developed at the University of Michigan.

Sample weights:

In complex panel surveys like the ECHP susceptible to high rates of non-response, the sample data have to be appropriately weighted. The weighting procedures can be elaborate and complex. In ECHP standard procedures were developed and implemented for the computation of sample weights, generally again *centrally at Eurostat*. Starting from ECHP wave 2 weights were developed on the basis of weights from preceding waves, modified to take into account unit non-response between the waves and adjustment of the achieved sample to external control distributions by various person and household characteristics.

Of course, there were limitations to comparability across countries in ECHP data. Nevertheless, each of the features (1)-(7) contributed much to the achievement of a fairly high degree of comparability across countries.

In order to adequately address the comparability issues in EU-SILC, it is necessary to identify how the differences between the EU-SILC and ECHP structure and arrangements make the problem of comparability more complex and acute in EU-SILC. For this purpose it is necessary to begin by clarifying the basic characteristics and requirements of EU-SILC in their effect on comparability.

3.2 Characteristics and requirements of EU-SILC affecting comparability

Flexibility is an essential feature of EU-SILC. This means that the EU-SILC dataset may comprise different types and combinations of data sources, with different designs.

Cross-sectional and longitudinal components

Data are required in both cross-sectional and longitudinal dimensions. Both cross-sectional and longitudinal micro-data sets need to be updated on an annual basis.

The first priority is given to the production of comparable, timely and high quality *cross-sectional* data. The cross-sectional component covers information pertaining to the current and a recent period such as the preceding calendar year. It aims to provide estimates of cross-sectional levels as well as estimates of net change from one period (year) to another.

The longitudinal component covers information compiled or collected through repeated enumeration of individual units, and then linked over time at the micro-level. It aims at measuring gross (micro-level) change and elucidating the dynamic processes of social exclusion and poverty. Longitudinal data are to be limited to income information and a limited set of critical qualitative, non-monetary variables of deprivation, aimed at identifying the incidence and dynamic processes of persistence of poverty and social exclusion among subgroups in the population. The longitudinal component is also more limited in sample size. Furthermore, for any given set of individuals, micro-level changes needs to be followed up only for a limited duration, such as a period of four years.

Diverse data sources

The cross-sectional and longitudinal data can come from separate sources, i.e., the longitudinal dataset does not need to be “linkable” with the cross-sectional dataset at the micro-level. Of course, such linkage is normally present in so far as the two types of data come from the same source. In principle, depending on the country, micro-data could come from:

- (1) one existing national source (survey or register);
- (2) two or more existing national sources (surveys and/or registers) directly linkable at micro-level;
- (3) one or more existing national sources combined with a new survey – all of them directly linkable at micro-level;
- (4) a *new harmonised survey* (termed an 'integrated survey') to meet all EU-SILC requirements.

To-date, the integrated survey is the most common option adopted. This is because a majority of the countries are starting new surveys for EU-SILC.

Varied structures

Flexibility of EU-SILC means that EU-SILC dataset may comprise different types and combinations of data sources, with different designs. A *typology* has been developed of the structure and design of EU-SILC data sources, describing in particular aspects pertaining to sampling.⁹ It identifies the following possibilities:

⁹ Verma, V. (2001) *EU-SILC Sampling Guideline*. Manual developed for Eurostat.
Verma, V., and Betti, G. (2006). EU statistics on income and living conditions (EU-SILC): choosing the survey structure and sample design. *Statistic in Transition*, 7(5), pp. 935-970.

- [A] a single integrated source covering all components – cross-sectional and longitudinal, income and social; or
- [B] two separate surveys, one cross-sectional and the other longitudinal, each covering both income and social (non-income) variables; or
- [C] two separate sources, one covering income variables and the other covering social variables, both cross-sectional and longitudinal in each case.

It notes two other possible arrangements, though likely to be rare:

- [D] a single ECHP-type panel survey, providing all cross-sectional and longitudinal data, but primarily focused on the latter; or
- [E] a separate source/arrangement for each component, cross-sectional vs. longitudinal, and income vs. social variables.

More complex structure(s)

In practice, the most commonly used design is the integrated design, [A], as it is generally the most suitable one, especially for countries starting a new EU-SILC survey. In the integrated design, the cross-sectional sample at any year consists of four panels, one having been introduced afresh that year, and the others introduced, respectively, 1, 2 and 3 years ago. Clearly, this structure is more complex than that of ECHP. At any one time, panels of different ages constitute the total sample, which is likely to increase the complexity of both the cross-sectional and the longitudinal weighting and estimation procedures.

Another major source of complexity (and also diversity) arises from the choice between using a sample of 'complete' households (i.e., taking all members of a selected household into the sample), and using a sample of persons (essentially, selecting only one adult per household for the survey). This choice applies only to interview surveys aimed at collecting more complex non-income variables in countries where income information can be obtained from registers. Procedures need to be developed also for such samples of persons.

Income and social variables

In relation to the required survey structure the data covered in EU-SILC can be classified into the following types, for both the cross-sectional and the longitudinal components:

- o household variables, covering variables measured at the household level;
- o household member variables providing information on basic characteristics of household members;
- o income variables, covering the set of target variables on income, income sources and related aspects; these are relatively complex variables measured primarily at the personal level, but aggregated to construct household-level variables; and
- o social variables covering a range of target variables on living conditions, activities, attitudes and other non-monetary indicators, and may also include some closely related income variables; these are relatively complex variables collected and analysed at the person-level.

Variables concerning the household and household members mainly include straightforward items, collected or compiled at the household level. The information may come from an interview survey with a single respondent in the household, from registers and other administrative sources, or from some combination of the above. These sets of

variables require the same type of survey structure in all countries: a representative sample of households, covering all members of each household in the sample.

The choices concerning income and social variables are more complex and inter-related.

Income variables must be obtained for a sample of complete households, i.e. covering all income recipients (adults aged 16+) in each household. The information is too complex to be obtained by proxy, and must either: (i) be collected through personal interview with all adults in the sample households; or (ii) compiled from registers, thus replacing the interview survey altogether for these variables. All persons in the initial sample households need to be followed up (even if they move to a new address) over the duration of the panel for obtaining longitudinal data on income based on the full sample.

Social variables are also too complex or personal in nature to be collected by proxy. However, in contrast to income variables, these are generally not available from registers or other administrative sources, and must be collected through direct personal interview. Another crucial difference is that, from the substantive requirements of EU-SILC, it is not essential that these variables be collected for all persons in each sample household. It is possible to do this collection on a representative sample of persons (adult members aged 16+), such as by selecting one such person per sample household. Hence within EU-SILC objectives, the choice is between covering this set of variables: (i) on a sample of complete households, i.e. covering all adult members of each sample household; or (ii) on a subsample of adults, such as by selecting one adult per sample household.

In practice, the choice depends on the source of information for the income variables. The normal choice is to collect social variables on a sample of complete households if income variables are collected through personal interview, since the latter would then already involve detailed interviewing of all adults in the household. However, where income data are obtained from registers and involve no personal interviewing, it is more convenient and economical to collect social variables from a subsample, such as one adult per sample household.

Expanded coverage

In addition to all the above factors, the problems of comparability are made more complex and acute in EU-SILC simply as a result of the fact that more diverse conditions are to be covered. While ECHP was confined to the old EU15, EU-SILC covers the expanded EU25, and possibly also Norway, Iceland, Switzerland, and the Candidate countries including Bulgaria, Romania, Croatia and Turkey. Survey instruments and procedures have to be adapted such that comparable data are produced despite major differences in the prevailing structures and circumstances of the countries involved.

4 Assessment of achieved comparability

How may we assess the extent to which comparability has or has not been achieved across countries in the implementation of EU-SILC?

As noted, comparability is a complex concept and not easy to quantify. In order to assess the degree to which different bodies of data are 'comparable', it is necessary to examine them both from the input side and from the output side.

This parallels the distinction between 'process' and 'product' indicators of the broader concept of data quality.

- o By examination of comparability from the input side, we mean an analysis of the methodology and implementation of the process of production of the data. From where and how the data were collected, statistically treated, processed, and analysed?
- o By examination from the output side, we mean a comparison of the substantive results actually obtained, with reference to appropriate standards such as alternative data sources, prior knowledge, logical expectations, etc. How meaningful are the substantive results being compared in relation to each other and to appropriate external standards? How far can the estimates for different countries be put together, compared, and interpreted in relation to each other and against some common standards?

Both these aspects in the assessment are important, and must be taken into account in the implementation of the comparability assessment process.

4.1 Comparison of the production processes

A thorough examination of similarities and differences in the methodology of the production process of the data can provide very reliable indicators of the degree of comparability which may be expected. Even more strongly, it can indicate where the results are most likely to *lack* the required degree of comparability.

Frequently, reliable comparisons from the "output" side are simply not possible, and methodological comparisons from the "input" side is all that can be achieved.

Two important points need to be kept in view in comparisons of the data production methodologies. One is the distinction noted above between the *measurement aspects* which need to be the same or comparable, and the *estimation aspects* which can be chosen flexibly without affecting comparability. There are many examples of confusion arising as a result of unnecessarily requiring standardisation of not only the measurement aspects, but also of the estimation aspects.

Secondly, it should be noted that sometimes even the measurement aspects need to be *different* to obtain comparable results. A good illustration of this point is provided by the use of country-specific questionnaires to obtain the required information, as opposed to the use of common 'blue-print' questionnaires in all countries. Of course, a common questionnaire can ensure a common operationalisation of the concepts and content for the surveys. However, the requirement of comparability of the information generated does not necessarily imply the need to use identical questionnaires in all countries. On the contrary, because of differing legal and institutional frameworks, *different questions* are sometimes required in different countries to obtain the *same information*. An example is the enumeration of income from the diverse social protection schemes in different countries.

Hence it is important to compare EU-SILC data production methodologies among countries in a context-sensitive manner, rather than mechanically. Also, account must be taken not only of formal differences, but of actual differences in the implementation. Real differences can be much more important than differences in the formal procedures adopted; the converse is also possible – sometimes formal procedural differences having little consequence in their actual implementation.

4.2 Comparison of the substantive results

In principle, it is the comparison of the substantive results actually obtained (the "output") which is of interest in the assessment of comparability. Sometimes, comparisons from the "input" side may point to differences which appear serious, but

the effect of which on the actual results obtained turns out to be unimportant. Equally, comparisons from the "input" side may fail to identify differences, which in fact damage the comparability achieved in significant ways.

However, direct comparison of the substantive results is sometimes too difficult or even impossible. Hence it must be complemented by comparison from the "input" side.

For the same reason, it is often necessary to look for relatively large differences, for patterns which appear *implausible* in the light of all the available information. The dividing line as to which differences are "large" is, to a considerable extent, a matter of the analyst's judgement. In any case, to be of interest in the evaluation of comparability, the observed differences should be significantly larger than sampling and non-sampling errors involved in both the sources, and also larger than the differences between them expected on the basis of existing knowledge.

Hence, comparisons of the substantive results is also not a mechanical task. Researchers would need to use their analytical skills and subject-matter specific knowledge in this exercise.

Intermediate indicators between the input and output sides may also be useful in the assessment of comparability.

Consider for instance differences in the definitions of household and household membership. From the 'input' side, presence of conceptual and definitional differences point to their potential effect on comparability. As 'intermediate' indicators, we can examine the *extent* of the differences, for instance in terms of the numbers of individuals who get classified differently because of the conceptual and definitional differences. On the 'output' side we may examine, for instance, the *impact* of these differences on the resulting indicators of poverty and inequality.

5 Potential sources of non-comparability in EU-SILC

In this section we identify a number of specific aspects where problems of comparability are likely to arise, or where at least an investigation is called for. We also try to elaborate on possible directions of investigation for a few selected ones.

In practice, one would begin from the development a of *comprehensive framework* for such analysis and assessment. Then on that basis, individual sources of non-comparability can be examined in a systematic and consistent way.

5.1 Some potential sources of non-comparability in EU-SILC

The following is a tentative list of sources of non-comparability, which may be modified or extended as a result of further study and feedback from those implementing EU-SILC. It is possible to identify many other topics of interest in the study and assessment of comparability in EU-SILC data.

1. Detailed analysis of the comparability of income distribution by component, with particular attention to self-employment income, imputed rent and housing costs, and sources of non-monetary income.¹⁰

¹⁰ **There is a large number of technical documents prepared by EU-SILC Task Force on Methodological Issues, covering diverse income components. All these documents are also very relevant to most other sections of this paper.**

2. Analysis of differences in how income taxes are treated in different countries; assessment of the impact on comparability of the Gross-Net conversion procedures used; examination of how these different procedures can fit into the general micro-simulation model SM2, adopted by Eurostat; application of SM2 to check and replicate the gross-net conversion.¹¹
3. Assessment of the effect on comparability of the choice between annual vs. current income concepts.¹²
4. Study of comparability of non-income items defining living conditions, deprivation and social exclusion, examining in particular the type of supplementary variables developed in Eurostat *Social Report* on income poverty and social exclusion.¹³
5. Analysis of major differences in structure of the SILC instrument: fundamental difference between (i) use of registers for income, with a sample of persons for complex social variables; and (ii) income from the interview survey, with a sample of household for complex social variables.¹⁴

Some other relevant references on this topic include the following:

Church J., Verma V. (2001). *Methodological Manual on Income Statistics for EU Member States*. Manual prepared for Eurostat.

Verma V. (2003). *Preparation, Follow-up and Evaluation of the Pilot Experiment for the EU-SILC Project*. Report prepared for Eurostat.

¹¹ An essential reference to this is the following:

Verma V., and co-researchers (2004). *Income in EU-SILC: Net-Gross-Net Conversion; Common Structure of the Model; Model Description; and Application to ECHP Data for France, Italy and Spain*. Eurostat doc. EU-SILC 133/04. This provides a comprehensive description of the Siena Micro Simulation Model (SM2) and its application for the conversion of household and personal income data, as collected in diverse forms in different Member States under EU-SILC project, to the standard target variables on gross and net income required at the EU level.

See also:

Verma V. (2002). *Income in EU-SILC: Net/Gross conversion. An application to ECHP data*. Eurostat EU-SILC doc 113/02.

Verma V., Betti G., Ballini F., Natilli M., Galgani S. (2003). The Siena Micro-Simulation Model (SM2) for Income Data and Statistical Imputation in Conjunction with Micro-Simulation, presented to the *Workshop on Income and Labour Dynamics, WILD*, Siena.

¹² see Church and Verma (2001) mentioned above.

¹³ Giorgi L., Verma V. (editors; 2002). *European Social Statistics: Income, Poverty and Social Exclusion: 2nd Report*. Luxembourg: Office for Official Publications of the European Communities.

See also:

Commission Regulation (EC) No 1983/2003 of 7 November 2003 concerning EU-SILC as regards the list of target primary variables.

¹⁴ These aspects are examined in some detail in:

Verma V. (2001). *EU-SILC Sampling Guidelines*. Manual developed for Eurostat.

See also:

Commission Regulation (EC) No 1982/2003 of 21 October 2003 concerning EU-SILC as regards the sampling and tracing rules.

6. Assessment of limitations of comparability between income data from registers and from interview surveys. Are there systematic differences in the measured levels, composition and distribution of income between registers and surveys?¹⁵
7. Examination of the extent and impact of national differences in the basic concepts for data collection and analysis, such as definition of the household, and household membership.¹⁶
8. Documentation of the variations in modes of data collection (for instance, the use of CAPI, CATI and PAPI; different fieldwork and interviewing procedures), and studies of their impact on data comparability.
9. Comparison of the national survey questionnaires, including an overview of and commentary on coverage and content of the questionnaires.¹⁷
10. Quantifying the effect of different rates of cross-sectional non-response, and different attrition rates of the panel components. As in the case of the ECHP, different countries may achieve very different response rates. The problem may be smaller in EU-SILC because of the limited duration of its panels. On the other hand, the overall impact on comparability may be increased due to very different data collection situations in countries using registers and countries using the personal interview survey for the collection of income data.
11. Study of comparability of imputation procedures: Eurostat had developed an elaborate procedure for the imputation of missing items on income in ECHP; in principle, similar procedures have been recommended for EU-SILC.¹⁸ However, it is expected that generally detailed procedures will be developed and implemented only at the country level. This raises the issue of comparability of the results of imputations.
12. Study of differences in the weighting procedures used, and an assessment of the effects of such differences on comparability of the results. (As such, comparability does not require identical weighting procedures, but it is essential to have *common standards*.)¹⁹

¹⁵ See for instance:

Epland J. (2006). Challenges in income comparability: Experiences from the use of register data in the Norwegian EU-SILC. *VII International Meeting on Quantitative Methods of Applied Sciences*, University of Siena, Italy.

Rendtel et al (2004). *Quality of Income Data*. Report in CHINTEX project.

¹⁶ See Commission Regulation (EC) No 1980/2003 of 21 October 2003 concerning EU-SILC as regards definitions and updated definitions.

¹⁷ Verma V. (2002). *Pilot Experiment first EU synthesis-Overview of coverage and content of the questionnaires*. EU-SILC Doc 95/02. This is an extensive report examining the national questionnaires used in the various Member States for EU-SILC project for completeness and comparability, and makes recommendations to Eurostat for improvement of the survey instruments.

¹⁸ For the description of the fundamental requirements, see Commission Regulation (EC) No 1981/2003 of 21 October 2003 concerning EU-SILC as regards the fieldwork aspects and the imputation procedures.

See also:

Verma V., Betti G., Ballini F., Natilli M., Galgani S. (2003) quoted above.

¹⁹ The standard recommended procedures for cross-sectional and longitudinal weighting are developed in Verma V. (2006). *EU-SILC Weighting Procedures: An Outline*, document prepared for Eurostat. Individual countries may adopt or depart from the standard recommendations to varying degrees.

13. Comparative analysis of the incidence of negative, zero and small values of the total disposable household income, how these are treated in the national data sets, and the magnitude of the impact of these differences on the poverty, inequality and other indicators produced from the data.

5.2 Example of studies in comparability

(1) Assessment of the comparability of income distribution by component

Income of households is made up of diverse components received by different individuals in the household. Its elements may be compiled from different types of sources, which may differ in concepts and definitions and may not refer to exactly the same reference time. The different sources may be subject to differing patterns of response and recording errors, sampling errors, inconsistencies and incompleteness etc. All this affects comparability.

The assessment of comparability would require detailed analysis of the income distribution by component. The recommended definition of income for use in EU Member States specifies that gross income should include all regular receipts such as wages and salaries, income from self-employment, interest and dividends from invested funds, pensions or other regular receipts from social protection schemes, and any other current transfers received in cash which are regular rather than one-time. Income should not include any large, one-time or irregular receipts from inheritances and the like, which should be regarded as capital transfers since they are unlikely to be spent immediately upon receipt.

The EU income concept – definition²⁰

1	Employee income
2	Income from self-employment
3	Imputed rent of owner-occupiers and others
4	Property income
5	Current transfers received
6	Interest payments
Gross income (1+2+3+4+5 less 6)	
7	Current transfers paid
Disposable income (1+2+3+4+5) less (6+7)	

Possibilities

1. As concerns problems of non-comparability, perhaps the main issue here concerns differences among Member States in the classification used for sources of income. A given type of income may not be classified under the same heading. While the major components may be similarly defined, differences often emerge when components are disaggregated more finely. That is, differences in classification tends to become more problematic when the major income components listed in the table above are disaggregated further. Generally, more

²⁰ Based on Table 2.1, Church and Verma (2001) quoted above.

detailed are the income components, the lower is the degree of expected comparability among the countries. Of course, larger components are also affected if their composition in terms of the set of the constituent smaller components is not the same in all countries.

Furthermore, systematic differences may be expected between countries compiling information from registers and those collected it through personal interviews. Even if the disaggregation of income by components in registers is very precise or detailed, this is not necessarily exactly the same as other registers in other countries, or the same as recommended by the EU-SILC Income Manual, or the same as implemented in interview surveys.

Assessment of comparability would require comparison of the income distribution by component in order to identify major and unexpected differences, followed by a careful study of the sources, concepts and definitions used by each Member State in order to identify the main sources of non-comparability.

Particular attention would need to be paid to self-employment income. In the context of enlarged EU, non-cash income - including imputed rent, housing costs, and other sources of non-monetary income – has also become more important. Another related issue is the effect on comparability of the choice between the annual income and the current income concepts, though so far the latter has been used only rarely in EU-SILC (Ireland; possibly the UK).

2. At a later stage, EU-SILC will include imputed rent of owner-occupied or rent-free dwelling as a target variable, and as a component to be included in the total disposable income. This is a major component and can change the relative situation of countries differing greatly in the proportion of households living in owner-occupied or rent-free accommodation. Excluding this components damages comparability. But on the other hand, its inclusion can also introduce lack of comparability, which may arise from the use of different procedures for imputing rent. While Eurostat may recommend particular procedures (or a strategy) for the purpose, individual countries may not be willing to – and what is more likely, may not be *able* to – use identical procedures.
3. A related issue is the assessment of housing costs. EU-SILC is not concerned with consumption patterns, but the importance of housing costs in the analysis of *income* arises from the presence of *housing assistance* as an important component of income, in particular at the bottom end of the distribution. Increases in housing assistance may simply reflect increases in housing costs (such as increased rent of accommodation provided by public authorities), rather than a real difference in the level of living of the concerned households. It becomes a moot point whether the income distribution is examined before or after deducting housing costs. The significance of the issue varies from one country to another, making it a source of non-comparability.
4. Income-in-kind can only be covered partially in any income survey. Furthermore, EU-SILC begins with a restricted coverage for practical reasons: including only some main and feasible components such as private use of the ‘company car’. It is expected that the importance of income-in-kind – in particular of ‘auto consumption’, i.e., consumption of own production – varies greatly across countries especially with the enlargement of the EU.

For the analysis and evaluation of comparability, it will be necessary to carefully study the *extent* of the differences across countries in the distribution of income by component in order to identify the main sources of non-comparability. The next step would be to try

and quantify the *impact* of these differences on the inter-country comparability of EU-SILC data. Such assessment of comparability is an essential aspect in the implementation of EU-SILC, albeit a difficult and challenging one.

In conclusion, I would like to emphasise the following point.

An examination of the distribution of income by component is an extremely important step in the evaluation of the data quality, especially data comparability across countries. This can begin from comparing simple indicators such as the proportion of households receiving income from each component, and the share of that component in the total income received by households. This has to be followed by the identification of any large or unexpected differences in the patterns across countries, and search for plausible explanations for those. In my view, *such an examination and investigation is a minimum requirement before public release of micro data from EU-SILC*, a requirement which must supersede considerations even of timeliness and punctuality. Without it there is a danger of the data losing their credibility in the eyes of the potential users.²¹

(2) The incidence and treatment of negative and zero incomes

In principle, certain components making up the total household income can have negative values. These may also result in negative values for total net (disposable) household income, or even of total gross household income.

Many individual components of course have a zero value, simply implying that the household does not receive any income from that source. However, a proportion of households may report their total net income or even total gross income as zero.

Genuine negative or zero incomes can arise for a number of reasons, while negative or zero incomes may also result from errors either in what the respondent reports, how the information is recorded and processed, or what the information pertains to. In our view, it is likely that in a majority of the cases, negative incomes result from data errors of this type. This probably also applies to a large proportion of zero incomes.

Negative and zero incomes often result from under-reporting of income. Such under-reporting can lead to biased estimates of income levels, and also to biased estimates of the inequality.²²

²¹ EU-SILC is still in the process of establishing its credibility.

As noted by Brackstone:

“There is another aspect of data that may be more important (even) than accuracy. That is credibility. Credibility, i.e., the degree to which data are trusted by users may be partly a function of accuracy, but even more a function of the producing agency’s reputation, which in turn may be partly dependent on its ability to produce accurate data over a long period of time. (Brackstone G., 2001. How important is accuracy? *Proceedings*, Statistics Canada Symposium “Achieving data quality in a statistical agency: a methodological perspective”).

Fellgi identifies two ‘survival issues’ beyond quality: *respect for respondents* (privacy, confidentiality, data security, management of respondent burden); and *credibility of information* (accuracy, transparency, non-political objectivity, relevance). See Fellgi (2001). Comment. *Journal of Official Statistics*, 17, pp. 43-50.

²² Jenkins, S.P. (2000). The distribution of income by sector of population. ISER Working Paper 18.

Rigg, J.A. (1999). Income shares and income inequality in OECD countries since the late 1970s. Unpublished PhD thesis, University of Cambridge, UK.

Clearly, incomes reported as negative (or zero) purely in error cannot be accepted as such. However, the important point is that there are strong arguments for adjusting cases with negative (also zero) incomes, even where the negative amount is thought to be genuine. This is the case when incomes are used as a *proxy for living standards*, which can only be quantified as a positive number. People with genuinely negative incomes still have to maintain a 'positive' living standard by drawing on past income (in the form of savings and other capital) or future incomes (in the form of loans and other debt). Negative or zero 'total net disposable income of a household' cannot serve a meaningful proxy for living standards.

Consider total gross household income to begin with. All components making up the total gross must be positive or zero, except for the following two, for which negative values cannot be ruled out:

- (a) Gross self-employment income. This is defined as receipts *minus* expenses, before taxes and other deductions;
- (b) Interest paid on mortgage (EU-SILC variable HY100G)

Negative values in net income can arise from a number of additional sources:

- (c) The deduction of taxes and social insurance contributions from the gross amounts.²³
- (d) Regular inter-household cash transfers paid ('private transfers out'). Total net is defined as the sum of net income from all sources, *minus* private transfers out.
- (e) The deduction of certain costs in defining a particular concept of income. For example, income 'after housing costs' (AHC), much used in the UK for instance:

income AHC = net disposable income, *minus* housing costs.

- (f) The exclusion of certain components of income, but without taking into account the effect of that exclusion on various deductions related to the component concerned.

Two important examples of (f) may be noted.

- (f.1) Not including receipts from private pensions as components of income, but at the same time deducting any tax on those receipts from the gross.
- (f.2) In constructing income 'before social transfers' (BST):

income BST = total net income, *minus* social transfers received,

failing to take into account that certain private transfers out are sustained only from social transfers, and must therefore be disregarded in constructing the income variable if their source, namely social transfers, is disregarded.²⁴

Country data differ in the extent to which negative and zero values appear for particular components and for total net and gross incomes. While in part these no doubt reflect genuine differences in national situations, they may also be the result of factors which damage inter-country comparability of the data – factors such as the use of different

²³ We may also include here regular taxes on wealth (HY120). Employer's social insurance contribution (PY030) pose no problem in the present context, since they are first added into the total gross, and then exactly the same amount is deducted in computing the net – hence they cannot contribute to a negative value.

²⁴ The same applies to variable 'total disposable household income before social transfers except old-age and survivors' benefits' (HY022).

types of data sources (e.g., registers versus personal interviews), different extents and patterns of reporting errors, and also different treatments applied to the collected data.

*The last mentioned source – different data treatments – is an important one because its influence on the resulting data can be large, yet the source itself is largely controllable by countries adopting uniform procedures in the treatment of income data.*²⁵

A study on comparability of these aspects may investigate areas such as the following:

- the extent of variation across countries in the incidence of negative and zero values for total (net and gross) household income;
- whether or not these differences appear plausible on the basis of comparison with existing knowledge or logical considerations;
- if not, what are the possible sources of these differences; what are the main contributing components; whether these differences arise largely from reporting errors, or from differences in data treatment;
- what procedures can be adopted for dealing with negative and zero values so as to make the data more comparable;
- what is the impact of this type of adjustments on poverty, inequality and income distribution measures computed from the data.

(3) Impact on comparability of differences in the definition of household and household membership

The basic units of data collection and analysis in EU-SILC are the household and its members. How the household and household membership is defined is important for two reasons.

Firstly, as a unit for selection of the sample, the definition adopted influences the coverage of the population in the survey. The objective is to define the households such that each person in the study population belongs to one and only one household, so that a sample of households properly covers the entire population of interest.

The second consideration, particularly important in the case of EU-SILC, is the definition of the household as a substantive unit. Income of an individual person is defined on the basis of the total income of all members in the person's household (with that income equivalised to take into account the household size and composition). Hence how individuals are grouped into households determines the central variable – namely income – measured in EU-SILC.

The definitions of household and household membership, and how these definitions have been implemented in EU-SILC in different Member States, has consequences for the comparability between countries.

EU-SILC framework regulation provides a general definition of private household to mean “a person living alone or a group of people who live together in the same private

²⁵ Verma V. and co-researchers. “Some proposals on the treatment of negative incomes”. Report to Eurostat, March 2006. “Treatment of negative income: empirical assessment of the impact of the method used”. Report to Eurostat, June 2006.

dwelling and share expenditures, including the joint provision of the essentials of living”.²⁶

This general definition is supplemented in EU-SILC commission regulations by clarifying the treatment concerning household membership of certain special categories of persons in particular circumstances.²⁷

The special categories considered include: (1) persons usually resident, related to other members; persons usually resident, not related to other members; (2) resident boarders, lodgers, tenants; visitors; live-in domestic servants, au-pairs; (3) persons usually resident, but temporarily absent from the dwelling (for reasons of holiday travel, work, education or similar); (4) children of the household being educated away from home; persons absent for long periods, but having household ties (persons working away from home); and (5) persons temporarily absent but having household ties (persons in hospital, nursing homes or other institutions).

Such persons are included as household members if they share expenses and also satisfy certain additional conditions as follows. No additional conditions apply in the case of category (1). Concerning category (2), such persons must currently have no private address elsewhere; or their actual or intended duration of stay must be six months or more. In category (3), the persons must currently have no private address elsewhere and their actual or intended duration of absence from the household must be less than six months. In (4), irrespective of the actual or intended duration of absence, such persons must currently have no private address elsewhere, must be the partner or child of a household member and must continue to retain close ties with the household and must consider this address to be his/her main residence. Finally, in category (5), the person must have clear financial ties to the household and must be actually or prospectively absent from the household for less than six months.

The standard definition of household and household membership is not, or cannot be, followed exactly in all countries.²⁸ While the recommended formal definition of household has been followed by most countries, there are some cases of departure. For instance, Austria has used the dwelling-unit concept: a private household is defined as a dwelling unit with at least one person that has his/her principal residence in this dwelling, without reference to the sharing of expenses. This means that some households as defined in the survey would have been divided into more than one smaller households if the standard definition were used. In Finland, the private household is constructed of persons residing permanently in Finland in the end of the year (31.12.), who live alone, or who, related or not, reside and have their meals together or otherwise use their income together. In Italy, a household can only be composed of related persons in the following sense: “cohabitants related through marriage, kinship, affinity, adoption, patronage and affection”.

²⁶ REGULATION (EC) No 1177/2003 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC).

²⁷ Commission Regulation (EC) No 1980/2003 of 21 October 2003 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 definitions and updated definitions.

²⁸ Though the EU-SILC Framework Regulation (article 2) gives a definition of household, article 16 of this regulation notes that “small departures from common definitions, such as those relating to private household definition, shall be allowed ... provided they affect comparability only marginally”.

Even when the formal definition of household is the same as the recommended standard, there are departures in the treatment of particular categories of persons and circumstances. For instance, France notes that the actual composition of households in the field was determined by the respondent and the interviewer without reference to too formal a definition of membership.

In particular, there are differences among countries both in the reality of the living patterns of students and the statistical treatment of such persons in the survey. Special issues arise in relation to the position of students living away from home. If students are regarded as separate households they are likely to represent a substantial group of poor households. However, the fact that many students have very low incomes may reflect the reality of the situation. Patterns differ by country.

There are also other special groups the condition and treatment of which needs to be compared – groups such as domestic servants, boarders and lodgers. Are they covered within households where they live and work? Or are they treated as separate households living at the same address? Or are they simply ignored (or covered only partly) in the survey?

Such comparative research would involve at least three aspects: identification of the conceptual and operational differences of the survey units; determining their extent (i.e., the number of persons in the population affected); and estimating the impact of these differences on poverty, inequality and other substantive measures.

Vijay Verma
Siena, October 2006.