Harmonising socio-demographic information in household surveys of official statistics: experiences from the Federal Statistical Office Germany

Körner, Thomas; Meyer, Iris

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Terms of Use:
This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to use this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.
HARMONISING SOCIO-DEMOGRAPHIC INFORMATION IN HOUSEHOLD SURVEYS OF OFFICIAL STATISTICS

Experiences from the Federal Statistical Office Germany

THOMAS KÖRNER & IRIS MEYER

In order to provide the European Union with comparable statistical information, the European Statistical System (ESS) has developed different strategies of cross-national harmonisation. These strategies range from a complete (input) harmonisation of concepts and survey methodology to output harmonisation approaches mainly taking place on the level of the aggregated data. Such strategies challenge statistical agencies in different ways and they have specific strengths and weaknesses. The paper focuses on the harmonisation of socio-demographic information in household surveys of official statistics. We discuss the approaches used in the new Community Statistics on Income and Living Conditions (EU-SILC).

1 Introduction

The availability of harmonised data on income and living conditions in the member states of the European Union is on top of the political agenda. Decisions in the field of social policy in the 25 member states are highly dependent on an accurate and comparable data basis. The Lisbon European Council of March 2000, for example, defined the eradication of poverty as one of the highest political priorities. The political strategy chosen necessitates reliable statistical information: According to the heads of government of the member states the national policies for combating poverty and social exclusion shall be coordinated on the basis of the so called open method of co-ordination “combining common objectives, national action plans, common indicators with the aim of promoting more ambitious and effective policy strategies for social inclusion” (Commission of the European Communities, 2003: 4). Official statistics constitute the basis for the definition of these objectives and the monitoring reports.
With the Community Statistics on Income and Living Conditions (EU-SILC) – the successor of the European Community Household Panel (ECHP) – a new EU-wide data source has been created which feeds a whole system of reports and indicators. These reports include the annual Report to the Spring Council, the annual Social Situation Report, the periodic Joint Social Inclusion and Social Protection Report, and others.

While the political importance of harmonised data on income and living conditions is evident, the approaches towards the production of harmonised data are complex. A precise measurement not only of the survey variables but also of the socio-economic background variables is necessary. Harmonised measurement depends on various factors, such as a common definition, taking into account cultural differences affecting the meaning of variables, as well as an operationalisation of the variables respecting the specific national contexts. Harmonisation thus requires a careful consideration of both survey concepts and survey methods.

This paper focuses on problems of the harmonisation of socio-economic variables in household surveys of official statistics. We first summarise the main strategies of harmonisation in the European Statistical System (ESS). In a second step, we outline the harmonisation efforts in the field of surveys on income and living conditions, where a change from input harmonisation to ex-ante output harmonisation can be stated. Finally, we take a closer look at the harmonisation of socio-economic variables in the context of the EU-SILC which is carried out in Germany beginning in 2005. The harmonisation strategy for selected socio-demographic variables is analysed in detail.

2 Harmonisation Strategies in the European Statistical System

Providing harmonised European statistics is an essential function of the European Statistical System. The notion of harmonisation appears in almost all basic legal documents referring to official statistics. The Quality Declaration of the European Statistical System heralds improving “a programme of harmonised European statistics that constitutes an essential basis for democratic processes and progress in society” (Eurostat, 2002: 21) as the central “vision” of the ESS. However, any approach towards cross-national harmonisation is limited by two features of the ESS institutional environment: The subsidiarity principle and the national statistical systems. Both bring about specific challenges and opportunities.

Firstly, subsidiarity is a key principle of the EU administrative structure. The way European official statistics are produced is heavily relying on the existing structures in the national statistical systems. They encompass far more than only technical or purely administrative features. In the long tradition of the national statistical systems fundamental methodological traditions have been formed. In order to harmonise statistics at EU level successfully, these experiences and traditions must be taken into consideration.
Secondly, official statistics constitute more than a bunch of isolated surveys. In each member state, there are statistical systems which enable users to combine statistical information originating from different surveys. Long before harmonisation became an issue at the European level, these national information systems have taken shape. In order to get used properly, statistics have to fit into this environment. As Mejer (2003: 83) notes, the poor integration of the ECHP in the existing national statistical systems was one of its major drawbacks, thus creating a “potential for fragmented figures showing differing results”. Hence, harmonisation in official statistics has at least two dimensions which could be referred to as “cross-national harmonisation” and as “cross-survey harmonisation”. Successful harmonisation in one of these dimensions often counter acts harmonisation efforts in the other one. Figure 1 illustrates the problem of both types of harmonisation or selected household surveys. The horizontal bar represents the problem of cross-survey harmonisation in member state 1.

**Figure 1   Cross-National and Cross-Survey Harmonisation**

![Diagram](image.png)

**MS**: Member state  
**EU-SILC**: Community Survey on Income and Living Conditions  
**LFS**: Labour Force Survey  
**HBS**: Household Budget Survey

The vertical bar represents the problem of cross-national harmonisation in the case of the EU-SILC.
The current level of harmonisation of official statistics on the European level has been achieved by adopting a variety of different harmonisation strategies, tailored to each individual case. Evidently, these harmonisation strategies will not lead to a complete harmonisation of the entire set of concepts and methods used. Scholars like Günther (2004) or Grais (1999) distinguish input harmonisation, ex-ante output harmonisation and ex-post output harmonisation as strategies towards a harmonisation of survey data in the European Statistical System (Figure 2):

For **input harmonisation** all methods, concepts and procedures relevant for the production of statistical results are standardised from scratch. Therefore, input harmonisation is necessarily applied ex-ante and does in principle not take into account statistical systems already existing in member states. For this reason, substantial changes of the methods, definitions, and processes used in member states become necessary which in turn often lead to increased cost and could hamper the usability of the data on the national level. It is not surprising that there are only few examples for the use of this strategy, such as the European Community Household Panel (ECHP) and the Community Innovation Survey (CIS).

In contrast, with **output harmonisation** standardisation is limited to the statistical results. While the definitions of the statistical results are binding, the National Statistical Offices are (within certain limitations) free to choose the methodology to be applied. One can either implement a new survey in order to meet the requirements of the statistical results; we refer to this strategy as ex-ante output harmonisation. It is used for the EU-SILC. Or one can use pre-existing national data sources and convert the results into a harmonised format. This strategy is referred to as ex-post output harmonisation. Ex-post output harmonisation is supposed to cause lower costs than any other strategy of harmonisation, but its applicability has strict limitations and preconditions fairly difficult to fulfil (Minkel, 2004).

It has to be noted that these harmonisation strategies have to be regarded as ideal types which will not be purely found in the empirical reality of surveys. Even the ECHP has not been fully harmonised as some degrees of freedom were left to the National Statistical Institutes, e.g. regarding specific aspects of the sampling technique. For the EU-SILC, here classified as ex-ante output harmonised, quite a number of common rules for the methodologies to be used have been put in place. Consequently, the EU-SILC could still be regarded as a partly input harmonised survey (Mejer, 2003: 72).

---

1 One could argue that in different cultural contexts different methodological approaches might, however, be considered functionally equivalent. The same survey mode, e.g. mail surveys, might lead to very different mode effects in different cultural contexts. Such considerations suggest that problems of harmonisation have to be reconsidered in a more differentiated manner.
The harmonisation of socio-economic variables – which is a special case of the harmonisation of statistical surveys in general – can use similar strategies (Hoffmeyer-Zlotnik & Wolf, 2004). In this case, input harmonisation denotes that the harmonised concept is included already in the measurement instrument (e.g. the questionnaire). In the case of output harmonisation, measurement is carried out with reference to a non-harmonised (national or regional) concept, and the harmonisation will take place at a later stage by transforming the non-harmonised results. One would refer to such a strategy as “ex-ante” if the harmonisation is already planned in the design stage of the survey and as “ex-post” in the case of different surveys on the same topic which have been designed independently from one another. These strategies have to be applied complementarily within a given survey, as for each variable an appropriate approach must to be tailored. Some variables (like education) cannot reasonably be measured in an input harmonised ways whereas for others the detour via the measurement of national concepts can lead to important “losses”. Section 3 will illustrate this process with reference to selected socio-demographic variables in EU-SILC.
3 Harmonising Socio-Demographic Variables in Surveys on Income and Living Conditions

3.1 From ECHP to EU-SILC

Initially, the ECHP was carried out as an input harmonised survey. The so called “blueprint method” using uniform questionnaires, detailed definitions as well as rules and procedures was applied (Günther, 2004: 17). German official statistics conducted the ECHP from first to the third panel wave (1994 to 1996) in this input harmonised way. From 1997 to 2001, together with Luxembourg and the United Kingdom, Germany changed to ex-post output harmonisation. In this period, the data have been converted from the German Socio-Economic Panel (SOEP) into the format of the ECHP.2

In 1999, the ESS decided to change from input to output harmonisation and to replace the ECHP by the EU-SILC. The reasons for this decision were the need for updating the content according to the new political demands and requests for operational improvements, mainly the timeliness of the data (Mejer, 2003: 72). Thus the National Statistical Institutes were given more flexibility in the choice of survey methods. The use of existing national data sources was explicitly encouraged in order to achieve a better integration into the national statistical systems and to make sure that the “best national source” for information on income and social exclusion is used (Eurostat, 2001: 7).

Especially the Nordic countries use register information on income and further selected socio-demographic information. Most of those member states in which register information is not available, decided to implement a new survey for the EU-SILC. This might be interpreted as an indicator for the difficulties that arise when trying to integrate a new and extensive survey like the EU-SILC into an existing national survey (Meyer, 2004).

3.2 Harmonising socio-demographic variables in EU-SILC

While the ECHP was carried out on the basis of a “gentleman’s agreement”, in the case of the EU-SILC, the harmonisation of socio-demographic variables is prescribed by a number of EU regulations. The regulations provide, e.g. common definitions, the use of standard classifications for all National Statistical Institutes as well as the entire set of target variables and items. They incorporate nearly all units, variables, and classifications which have been defined in the Eurostat recommendations on the harmonisation of “core variables” (Eurostat, 2000).

2 In this contribution we will not focus on the problems of ex-post output harmonisation. The implications of this strategy have been investigated in detail in the CHINTEX project (The Change from Input Harmonisation to Ex-post Harmonisation in National Samples of the European Community Household Panel – Implications on Data Quality; see Ehling & Rendtel et al., 2004).
Given this solid legal framework, the cross-national harmonisation of socio-demographic and other variables might be considered close to an ideal solution as there is a legal obligation to adopt harmonised concepts throughout the entire survey. However, the situation is more complex. As Hoffmeyer-Zlotnik & Wolf (2004: 6) point out, it is only a first step to agree on common definitions. In addition further points, like the cultural and national background of the variables, the operationalisation of the common definition and the wording and design of the measurement instrument (are the concepts easily understood by all respondents?) have to be taken into account. Thus, successful harmonisation has to be prepared carefully already during the survey planning and questionnaire design phases.

The common definitions prescribed by regulations in some cases already have some bearing on the operationalisation (e.g. the household definition), resulting in input harmonised approaches. In other cases, only the target variable is defined, i.e. the breakdown according to which data on a given variable has to be delivered. In these cases the operationalisation is largely left to the National Statistical Institute: Can the target variable be measured as worded in the list of target variables (input harmonisation), or do we have to find a national operationalisation (output harmonisation)? Of course, the national operationalisations of the common definitions have to be functionally equivalent.

In the following we will present a number of selected examples of how the harmonisation of socio-demographic variables has been dealt with in the case of the implementation of the EU-SILIC in Germany.

**Private household and relationship of household members**

Private households are at the main population unit of the EU-SILC. The concept of household is defined in the regulations in detail, i.e. that input harmonisation has to be applied at least to some extent. The definition of a private household used for the EU-SILC combines the aspects of (a) living together and (b) sharing expenditures. This definition is in line with the one recommended by Eurostat and applied by the member states in the Household Budget Surveys (Eurostat, 2003a: 15). The regulation further specifies this quite broad and general definition in order to have unequivocal criteria of which individuals belong to a household: In addition to the general definition, household members are identified by a full set of conditions for inclusion, for instance

- having no private address elsewhere,
- having spent most of the daily night-rest in the household over the past six months, and
- intending to stay in the household for a period of at least six months.
For specific categories of persons further criteria apply. For example are considered as household members “persons temporarily absent [e.g. in hospital] but having household ties”, i.e. having financial ties and being absent for less than six months. However, if people are temporarily absent in private accommodation, they “may be included as household members irrespective of the length of absence, provided they are not considered members of another private household” (Commission Regulation (EC) 1980/2003).

Such a complex definition might be helpful to make sure that the risk of double-counting or under-coverage in the sampling frame are minimised. However, the definition is rather difficult to operationalise, at least if using self-administered questionnaires (as practiced in the German implementation of the EU-SILC). It is hardly possible to inform respondents on all the conditions for being regarded as a household member so that there is a risk that respondents simply rely on the definition on household membership they have in mind from their everyday knowledge. Response will be given (at least in part) according to the culturally shaped conception of private household. In order to solve this problem in the case of Germany, we used the general definition (residence and sharing expenses) and simplified its further specification for the questionnaire by illustrating it by different examples.

Concerning the household structure, in the regulation, there was a change from input to output harmonisation. In the ECHP, household structure was measured by the help of a relationship matrix in which respondents had to name the relationship of every single household member with every other member of the household. In the EU-SILC, this matrix is no longer required; the target variables can also be obtained with an alternative operationalisation. One of the reasons might have been that the relationship matrix can be difficult to respond at least in the case of large households or in self-administered questionnaires. Our experience from the EU-SILC pilot study as well as the pretests for the EU-SILC questionnaires have shown that completing a relationship matrix is a complex and cognitively demanding task which is prone to response errors, item nonresponse, or even a break-off of the interview.3

For these reasons the Federal Statistical Office Germany decided to no longer to use the complete matrix of all household members in the EU-SILC. Alternatively, we ask for the relations of a reference person only to each of the household members and add few complementary questions (similar to the concept used in the Labour Force Survey). The complete relationship matrix, in the sense of the ECHP, can then be obtained by calculating the complementary relationships based on the relationships filled in for the reference

---

3 More generally, recent research has again shown that particular groups of respondents have difficulties to fill in matrices or tables in survey questionnaires (see, e.g. Timm, 2004).
person. The change towards output harmonisation clearly gave the National Statistical Institutes greater flexibility to adopt their “best national practice”. Furthermore, given the German experience, a complete household relationship matrix in all different national contexts and traditions had probably not helped to achieve a higher degree of harmonisation. The harmonised method would have been prone to various kinds of errors, including not only measurement errors, but also processing errors (due to the fact that the analysis of the data from the matrix is a time consuming and complex task).

**Occupation and employment**

For a harmonised measurement of occupations in cross-national surveys, the International Standard Classification of Occupations of 1988 (ISCO-88) is a standardised and widely accepted instrument. Not surprisingly, also the EU-SILC relies on ISCO-88 (on the two-digit level) in order to obtain background information on the current main job. As the ISCO-88 is a highly complex instrument, it is still not considered possible to categorise occupations according to ISCO-88 during the data collection phase (Hoffmeyer-Zlotnik, Hess & Geis, 2004). In principle, input harmonisation could be adopted: Respondents could be asked to provide information on their current job in a set of open questions. Subsequently, the information has to be recoded according to the international classification.

However, in the German case there is a further complication. Besides the ISCO-88, there is also a national standard classification on occupation, the “Klassifizierung der Berufe 1992 (KldB)”, published by the Federal Statistical Office. The KldB is still a necessary information resource within the German statistical system. Due to the differing classification principles of the ISCO-88 and the KldB, a conversion from ISCO-88 to KldB is not possible, whereas within certain limitations, KldB information can be converted into the ISCO-88 format. Unfortunately, also the conversion from KldB to ISCO-88 leads to some loss of information which could only be avoided by adopting a strategy of input harmonisation (i.e. phrasing the respective questions according to the requirements of ISCO-88 and coding directly from this information). However such an option would heavily restrict the use of the results within the national statistical system. As different

---

4 ISCO-88 categorises occupations on the basis of the level and similarity of the skills required. It uses the two dimensions of (in a first step) “skill level” (range and complexity of a task) and (subsequently) the “skill specialisation” (reflecting the type of knowledge applied, tools and equipment used, materials worked on, or with, and the nature of the goods and services produced). ISCO-88 distinguishes nine major groups, which are further subdivided into 390 unit groups on the four-digit level (ILO 1990: 2). The KldB, in contrast, distinguishes occupations according to the occupational title, starting from six “occupational areas” (like “occupations in agriculture, forestry and gardening”) all the way down to 29527 concrete occupation titles. On the three-digit level, 369 orders of occupations are distinguished (Statistisches Bundesamt, 1992).
survey questions are required also parallel coding according to both classification systems is not an option. Finally, it should be mentioned that coding according to ISCO-88 is by far more resource consuming.

As for the EU-SILC ISCO-88 codes are required on the two-digit level only the loss of information during the recoding can be restricted to an acceptable level. Thus, a conversion table can be used, which has already been applied in the German Labour Force Survey since 1996 (Emmerling & Riede, 1997; Schwarz, 2001).

This case shows that in official statistics the requirements of international and national users have to be reconciled. There is not a major problem as long as national and international classification systems are identical, at least on a broad level. For example, for the coding of economic activities the Statistical Classification of Economic Activities in the European Community (NACE rev. 1) is identical with the national classification “Wirtschaftszweige 2003” (WZ 2003) on the two digit level. In contrast, the case of the measurement of occupations makes clear that in case of conflicts between national and international classifications systems solutions have to be found which take into consideration the requirements of both national and international users.

**Education**

As for the occupation, also for education, an international standard classification is available and adopted in EU-SILC – the International Standard classification of Education (ISCED 1997). As the ISCED categories are quite remote from the everyday life of respondents, output harmonisation is the only feasible harmonisation strategy: Respondents are asked to tick the category which applies to them out of a list of national educational levels. In Germany, this information can subsequently be recoded according to ISCED 97 without major problems as a standard conversion table exists.

However, despite the fact that for education the national and international classification systems are convertible, some questions arise with respect to the harmonisation. The national education systems greatly differ across nations, which is very difficult to grasp by an international classification. In contrast to the situation with ISCO-88, for ISCED 97 this does not conflict with a measurement which is also instructive for national purposes. However, ISCED fails to represent cases of horizontally differentiated education programmes. For example, vocational training in Germany cannot be represented appropriately side by side with the general educational levels. Nevertheless such a request would possibly overburden an international classification system.
Income

The main purpose of the EU-SILC is to become the new reference source for EU comparisons on poverty, living conditions, and income distribution. As basis for the so called “open method of co-ordination” the EU-SILC is also one of the major inputs for decisions on social policy within the member states of the EU. This aim necessitates first of all a highly precise and extensive measurement of the income. The major income information required for the EU-SILC includes:

- the total gross household income,
- the total disposable household income, and
- the total disposable household income before social transfers including/excluding old-age and survivor’s benefits.

This income information will be compiled on the basis of a broad variety of income components (gross as well as net) on household and personal level. Income reference period is the previous year. The retrospective measurement is necessary as some income components are only available after the reference year and only on a yearly basis (like tax repayment). According to the regulation, a multitude of income components have to be provided in equally high precision: income from self-employment, from employment, from social transfers and from private pension plans. Additionally, income from capital as well as near-cash income have to be reported. The German implementation of EU-SILC comprises 89 types of income information which are required in order to provide the 23 target variables laid down in the regulations.

This brief overview makes clear that for the EU-SILC, income is far more than a socio-demographic background variable. It is the one of the main focus areas of the survey. As a simple background variable, one would not have to measure income in such a differentiated way. For this reason, the problem of a harmonised measurement of income differs in some respect from other surveys. If, e.g. the net household income was required, only two or three questions might have been sufficient and the questionnaire would be fairly different from that of the EU-SILC.

The features of the measurement of income in the context of the EU-SILC are based on the recommendations of the Expert Group on Household Income Statistics (2001), the so called “Canberra Group” and the ESSPROSS manual for social transfers received which have been adopted in the target variables. The measurement itself has to rely on the more differentiated national categories, which could in principle be grouped according to the international conventions without major problems. In order to assure precise income information on household level, the person who knows best about household affairs shall
answer the questionnaire. This rule shall help to get the best income information from each private household. For personal income information, according to the framework regulation of EU-SILC, the mode of data collection is a personal interview. This will be realised in the German case by the use of separate self-administered personal surveys for every respondent in order to prevent an underestimation of the personal income.

However, for the international harmonisation there are numerous difficulties which arise from differences in the national systems of income and social security. Just to name a few examples: In some countries it is not usual for employees to get information on their gross income. Sick leave payment might be part of the income or of the social security system in different countries. Tax reimbursements have a very different timing and volume in different tax systems. Some income components might fit in different income categories according to the national structures (e.g. subsidies for housing cost, meals, transport which can be included in the income or paid separately). Finally, the retrospective income measurement is a further source of error. Respondents have to recall the income from the last year as for many income components no information available before the spring of the following year (or even later).

4 Conclusion

Realising cross-national harmonisation is a challenging process for statistical agencies in which no ideal solution is at sight. Instead, feasible solutions have to be looked for in some kind of an optimisation process. Suitable approaches have to take into account the requirements of cross-national and cross-survey harmonisation within the national statistical systems as well as the restrictions related to possible survey designs and data collection modes. Although the specific situation of each variable has to be analysed individually, some general conclusions can be drawn.

From the point of view of cross-national harmonisation, theoretically, one could think of input harmonisation as an ideal solution. At first sight, harmonising all concepts and methods seems to the best way to obtain harmonised results. The problem is that while the survey itself can be harmonised to a large extent, the national and cultural environment of the survey remains quite heterogeneous which leads to a number of problems. Problems arise from the fact that most variables cannot be removed from their national setting (e.g. the income variables referred to above). Other harmonised variables cannot be measured appropriately as respondents do not have the international concepts available in their everyday knowledge.
With regard to the requirements of the users of the national-level data, output harmonisation seems to be the preferable strategy. Input harmonised surveys often lack a foundation within the respective national statistical systems. This potentially leads to survey results which could be used only partially on the national level (as shown in the case of occupation variables). Further possible consequences include the longer time span required for data processing, increased survey costs, and lacks of coherence with other statistics from the same socio-economic area. Numerous problems of this kind have been experienced in the context of the European Community Household Panel (ECHP).

However, ex-post output harmonisation cannot be regarded as the best solution either. Although offering the possibility to keep the national statistical systems unchanged (and thus avoiding breaks in time series etc.) the extensive research carried out in the context of CHINTEX as well as the German pilot study for the implementation of the EU-SILC (in which the data provision from an existing national data source also was tested) made clear that ex-post output harmonisation could not be applied as a standard procedure. As CHINTEX showed (Minkel, 2004), ex-post output harmonisation is subject to an important number of preconditions which are given only in a fairly small number of cases and only for a restricted set of survey variables.

Against this background, ex-ante output harmonisation seems to be the most promising way to enable international harmonisation, at least in the case of EU-SILC. This strategy gives flexibility to use the “best national practices” and to reconcile national and international information requirements. Complete harmonisation probably is only a theoretical option.

References


