



EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics and Information Society
Unit F-3: Living Conditions and Social Protection Statistics



DOC LC-HBS/07/09/EN

WORKING PARTY
LIVING CONDITIONS

10-12 June 2009

EUROSTAT-LUXEMBOURG

Room: Ampère

**IMPROVING DATA COMPARABILITY FOR THE NEXT HBS
ROUND (2010)**

Issues on data comparability in the EU Household Budget Surveys (HBS)

Introduction

The 2005 round of data collection of the Household Budget Surveys (HBS) in the European Union (EU) is coming to an end. Micro-data files and/or aggregated tables have been collected by Eurostat for all the EU-27 countries plus Norway and Croatia thus far. A series of indicators derived from the HBS 2005 data is now available on Eurostat's website¹ (NewCronos Database) for all these 29 countries.

One of the primary goals of the next HBS wave, which is due to take place in 2010, is to strengthen data comparability between the countries as well as with other Eurostat data sources (EU-SILC, LFS...). The purpose of the present document is to draw up a list of problems which arose in HBS 2005 in relation to data comparability, with the prospect of issuing recommendations for further improvements of the instrument.

The HBS 2005 experience has shown there are three areas where actions might be taken in order to improve the comparability of the HBS data:

- The survey methodology, that is, the various statistical aspects in relation to the way the survey is designed and conducted. For instance, the survey methodology comprises the following key elements: the sampling design, the method of data collection (use of diaries), the basic concepts and definitions used (consumption expenditure, household's reference person...) or post-survey adjustments like, for instance, weight adjustments for non-response or calibration to external data sources.
- The structure of the HBS data files, that is, the list of variables for HBS micro-data files and/or the list of aggregated tables.
- The HBS metadata reporting, that is, descriptive information or documentation provided about HBS micro-data, aggregated tables, or about other HBS metadata. Statistical Metadata facilitates sharing, querying, and understanding of statistical data over the lifetime of the data.
- The consistency of the HBS variables compared to the core variables.

Comparability issues in the HBS methodology

1. The achieved sample sizes in the HBS countries

A prerequisite for greater comparability of the HBS data is to satisfy a minimum level of accuracy in each country. A possible solution for better accuracy in each country is to set out minimum household sample sizes that should be attained by the countries for the next HBS wave. Given the EU-HBS is run under a Gentlemen's agreement, the determination of minimum sample sizes for HBS 2010 should take into account both the need to ensure a minimum level of accuracy at EU and national levels, and the need for the minimum sample sizes not to be too burdensome for the countries. The latter means the achieved sample sizes for HBS 2005 should be taken into account somehow when determining the minimum sizes for HBS 2010.

¹ Population and social conditions/Living conditions and welfare/Consumption expenditure of private households

Having said this, one has to note that from an EU perspective the relative sample size is less balanced than in other surveys. Table 1 compares sample sizes in HBS and EU-SILC. Another point worth nothing is that all the countries having a sample size lesser then 6000 in HBS, have a higher sample size in EU-SILC, with Netherlands having an EU-SILC sample which is 5 times then that in HBS.

Table 1: Achieved sample sizes (in households) – HBS 2005² and EU-SILC 2005³

Country	HBS	EU Statistics on Income and Living Conditions (EU-SILC)
Austria (AT)	8400	5148
Belgium (BE)	3550	5166
Bulgaria (BG)	2870	-
Croatia (HR)	2727	-
Cyprus (CY)	2990	3746
Czech Republic (CZ)	2965	4351
Denmark (DK)	2449	5957
Estonia (EE)	3432	4208
Finland (FI)	4007	11229
France (FR)	10240	9775
Germany (DE)	52217	13111
Greece (GR)	6555	5568
Hungary (HU)	9058	6927
Ireland (IE)	6884	6085
Italy (IT)	24107	22032
Latvia (LV)	3774	3846
Lithuania (LT)	7586	4441
Luxembourg (LU)	3202	3622
Malta (MT)	2586	3459
The Netherlands (NL)	1570	9562
Norway (NO)	3376	5996
Poland (PL)	34767	16395
Portugal (PT)	10403	4615
Romania (RO)	33066	-
Slovakia (SK)	4710	5414
Slovenia (SI)	3725	8287
Spain (ES)	8881	13027
Sweden (SE)	2079	6133
United Kingdom (UK)	6785	10826

² Source : Quality Report of the Household Budget Surveys – 2005

³ Source: EU-SILC Comparative Final EU Quality Report 2005 (Version 2 – September 2008)

Two different approaches for determining minimum sample sizes might be used. The idea of the first approach is to set out a minimum level of accuracy that all the countries should achieve in 2010. The minimum sample size needed to meet the precision requirement would then be derived for each country separately using analytic variance formulae.

The second approach considers it as a pure problem of sample allocation among countries under the constraint of a fixed size at EU level. The latter problem is similar to a problem of sample allocation in case of stratified sampling.

These two approaches are described in detail in a separate document entitled "Determination of minimum sample sizes for the 2010 round of data collection of the Household Budget Surveys (HBS) in the EU"

It is expected that all possible efforts are done to improve the sample size, especially in those countries which have a very small sample

2. The non provision of Imputed Rent by certain countries

According to the European System of Accounts (ESA 95), which is the reference for the HBS, the purchase of the dwelling as such is regarded primarily as capital formation (investment) and not consumption expenditure. However, the ownership of a dwelling is considered to produce a service – a shelter –, which is actually consumed over time by the households. As a consequence, ESA requires the estimation of the price of the shelter, by imputation of a rent, since no monetary transaction is involved. This imputed rent is part of household consumption expenditure. So, for the HBS to be consistent with the ESA principles, it has been recommended to exclude the acquisition of dwellings, whereas the consumption of the service of the dwelling should be included. Different methods can be used in order to estimate the imputed rent. The choice of a method generally depends on the size and the structure of the national rental housing market⁴:

- Self assessment: the self-assessment approach is based on information provided by the homeowners on the market rent they would pay if they were to rent their accommodation.
- Stratification: the sample of dwellings is divided into homogeneous groups. The mean of the rent values actually paid within each group is then given to the owned dwellings.
- Regression based approaches, consists of fitting a linear model to the rents actually paid by renters using auxiliary variables which are assumed to be correlated to the amount of rent paid. This model is used to estimate the values for owners, reduced rent and rent free. In particular, Heckman regression, which allows the researcher to correct for selection bias (by using renters to calculate the value for non renters) by using a two-step regression approach,
- User Cost method : User cost consists of the depreciation on the asset or durable (measured at current prices and not at historic cost) plus the capital, or interest, cost.

ESA 95 also takes into account dwellings which are rent free or have a reduced rent. In HBS 2005, all the countries provided the imputed rent variable, with the exceptions of CZ, MT, RO, IE and UK. For the two latter countries, in order to be consistent with the 1999 HBS data and as micro-data were available, the imputed rent was estimated by Eurostat using post-stratification.

⁴ For a more comprehensive analysis one can consult the document on CIRCA entitled : Meeting of the Working Group on Living Conditions (HBS, EU-SILC and IPSE) 15-16 May 2006 : HBS and EU-SILC Imputed rent

Table 2: Method of estimation of Imputed Rent, HBS 2005⁵

Country	Self-assessment	Stratification	Regression	User-cost
Austria (AT)		X		
Belgium (BE)			X	
Bulgaria (BG)		X		
Croatia (HR)	X			
Cyprus (CY)	X			
Czech Republic (CZ)	Imputed rent not provided			
Denmark (DK)				X
Estonia (EE)		X		
Finland (FI)		X		
France (FR)			X	
Germany (DE)		X		
Greece (GR)	X			
Hungary (HU)				
Ireland (IE)	Imputed rent not provided			
Italy (IT)	X			
Latvia (LV)			X	
Lithuania (LT)			X	
Luxembourg (LU)				
Malta (MT)	Imputed rent not provided			
The Netherlands (NL)				
Norway (NO)		X		
Poland (PL)				
Portugal (PT)	X			
Romania (RO)	Imputed rent not provided			
Slovakia (SK)				X
Slovenia (SI)		X		
Spain (ES)	X			
Sweden (SE)		X		
United Kingdom (UK)	Imputed rent not provided			

For the next HBS round, all the countries should provide the imputed rent variable, as per ESA 95 guidelines and as already agreed too for the HBS 2005 wave. This is even more possible as from 2007 onwards that variable is mandatory for EU-SILC (EU Statistics on Income and Living Conditions). Priority is to be given to:

- (1) Stratification or regression;
- (2) over self-assessment and user-cost

⁵ Source : Quality Report of the Household Budget Surveys – 2005

3. The use of non-random sample selections by certain countries

Most of the HBS participating countries drew a sample of households in a way that the probability of a household being selected is known (technically known as a probability design). In this way, the results can be reliably projected from the sample to the household reference population with known levels of certainty/precision, i.e. standard errors and confidence intervals for survey estimates can be constructed.

On the other hand, non-probability schemes (Quota selection) were implemented in Czech Republic and Germany. Although this type of sampling is generally quicker and cheaper, there is no assurance that the selection of households is not biased and is representative of the whole population. This error can be reduced if the enumerators are knowledgeable enough to choose alternative households with the same characteristics as the ones which are not available. Unfortunately, not enough information is available to guarantee that this has actually happened and/or to what degree.

For the next HBS round, Eurostat should recommend all the countries to implement probabilistic sample selections. As regards Czech Republic and Germany, Eurostat should enquire whether or not this proposal can be implemented in these two countries within their own national statistical systems

4. The treatment of unit non-response

High rates of non-response are a common and major problem in Household Budget surveys. In a number of surveys the sample initially selected is substantially, even several-fold, larger than the completed number finally required. At worst, the sample may become essentially self-selected and hence quite unrepresentative of the population of private households. It is therefore important to keep a track of the response rates achieved. Table 3 below shows the response rates achieved in HBS 2005.

Table 3: Household response rates, HBS 2005⁶

Country	Response (%)
Austria	41.9
Belgium	5.9
Bulgaria	65.1
Croatia	71
Cyprus	88.9
Czech Republic	Unknown
Denmark	54.9
Estonia	49.4
Finland	52.6
France	56.8
Germany	Unknown
Greece	60.3
Hungary	50.3
Ireland	-
Italy	75.2
Latvia	51.9
Lithuania	69.8

⁶ Source : Quality Report of the Household Budget Surveys – 2005

Luxembourg	-
Malta	-
The Netherlands	-
Norway	52
Poland	55.2
Portugal	62.3
Romania	90.3
Slovakia	74.7
Slovenia	74
Spain	65
Sweden	52
United Kingdom	57.0

Response rates are unknown in Germany and Czech Republic because of the non-random (quota) nature of the procedure used for sample selection. The lowest response rates are obtained in Belgium where the survey puts a heavy response burden on the household (actually, the Belgian HBS is combined with a Time Use Survey - TUS). On the other hand, relatively high response rates were attained in Italy (75.2%), Cyprus (88.9%) and Romania (90.3%).

In order to make the treatment of nonresponse in the Household Budget Surveys more efficient, national good practices in dealing with nonresponse should be shared with all the countries. National practices may comprise preventive measures (better training of the interviewers, information letter sent to the households before data collection...), or weight adjustments (post-stratification, logistic regression...) to reduce bias. Eurostat invites the countries which achieved more than 70% response, namely Croatia, Cyprus, Italy, Romania, Slovakia and Slovenia to prepare a short paper describing their experiences, and this will be circulated to the other countries for their consideration.

5. The concept of household's reference person

In each household it has been necessary to identify as its head (or reference person) a particular individual whose personal characteristics can be used in the classification and analysis of information on the household. The social group, occupation and employment status, income, sex and age etc. of the head are often used to classify the sample households for presentation of the results and for weighting classifications used in the derivation of the survey estimates. The identification of such a person can also be relevant in the collection of the data, for instance in determining the appropriate respondent for certain items of information if not all respondents are asked about the same individual information.

The next table lists the main criteria used by the countries in the definition of household's reference person. Although there have been suggestions in using an objective definition for the household's reference person, that is, the person aged 16 or more who most contributes to the household income, some countries keep using subjective criteria:

- The person who is designated as such by the other members (Ireland, Greece, Cyprus, Luxembourg, Romania and Croatia)
- The husband or the male partner (Czech Republic and Hungary)
- The householder (United Kingdom)
- The oldest active male (Greece)

From the EU Comparability point of view, it is important to identify and apply a consistent definition of a reference person which can be used in the classification and analysis of information on the household. Leaving the choice of the reference person to the household itself gives too much space for arbitrary decisions that will damage the comparability of data afterwards. By assigning the person who contributes most to the household income, there should be a clear connection between the income of the household and its expenditure patterns.

Table 4: Household's reference person, HBS 2005⁷

<p>Belgium In households with at least one active member, the reference person is defined as the member who contributes the most to the household financial resources. In households without any active members, the reference person is defined as the member of the household who has the highest replacement income</p> <p>Bulgaria The adult (16+) contributing most to the household income</p> <p>Czech Republic In complete (husband-wife, unmarried-couple) families, the reference person is always the adult male (husband). In single-parent families, the reference person is always the parent if he/she is economically active. If he/she is inactive, he/she can still be the reference person provided the child/children is/are inactive too. In non-family households, this is the person with the highest income</p> <p>Denmark The person having the largest gross taxable income</p> <p>Germany The person who contributes most to the household net income</p> <p>Estonia Person with the highest income</p> <p>Ireland The person chosen by the household members</p> <p>Greece As reference person was considered:</p> <ul style="list-style-type: none"> • The head of the household, if he/she was economically active (having or searching for a job). When the head of the household was neither working nor searching for a job, in order him/her to be considered as reference person, there shouldn't be any other economically active person in the household. • The spouse of partner of the head if he/she was working or searching for a job, if the person declared as head was not economically active. • The oldest member of the household, being economically active, if the head/s spouse/partner was not economically active. • The head of the household if noone else in the household was economically active. As head of the household was considered the person being approved by the other household members, in the sense that this person has the responsibility for the decision making concerning the household's management, in general. <p>Spain The reference person is defined as the member of the household whose regular (not occasional) contribution to the common household budget is intended to meet a greater proportion of the household expenditure than the contributions made by the other members</p> <p>France Person with the highest income</p> <p>Italy The household reference person is the holder of the household register sheet</p> <p>Cyprus As reference person of the household was considered the person who had, in the opinion of the other household members, the responsibility for all the major decisions concerning the household's affairs and the household's management in general, and / or is the principal economic supporter</p> <p>Latvia The reference person is person who has the highest contribution to the common household budget (the main bread winner) not only during the survey month, but also during the last 3-6 months. For farmer households this period is 12 months</p> <p>Lithuania The person with the highest income in the course of the year in the household</p>

⁷ Source : Quality Report of the Household Budget Surveys – 2005

Luxembourg

Self-defined (*main decision maker*)

Hungary

For one-family households, the reference person is the husband or the male partner. In case of multiple family households, this is the oldest family head

Malta

Person contributing mostly to the budget of the household

Netherlands

Principal wage earner (*member with the highest personal income*)

Austria

The reference person is the one who contributes most to the total income of the household

Poland

Person aged 16 and over with the highest income

Portugal

The reference person is the person aged 16 or more and with the highest income

Romania

The adult person declared and recognised as such by the other members of the household, usually the husband. The decision is exclusively up to the household members and takes into account some personal characteristics, such as: authority, age, occupation, income, ownership of the household, etc.

Slovenia

The reference person is the person with the highest income

Finland

The person contributing most to the income of the household (according to interview)

Sweden

Person with the highest income

United Kingdom

The Household Reference Person is the householder, that is, the person who:

- owns the household accommodation, or
- is legally responsible for the rent of the accommodation, or
- has the household accommodation as an emolument or prerequisite, or
- has the household accommodation by virtue of some relationship to the owner, who is not a member of the household

If there are joint householders, the HRP is deemed to be the one with the highest income. If the income is the same, then the eldest householder is taken

Croatia

In practice, the reference person is either the person under whose name a particular household is registered or a person who is declared as such by the members of the household

Norway

Person with the highest income

The Head of household/reference person is to be nationally defined, i.e. it can differ from one country to another. But Eurostat recommends that all relevant information about the adult (16+) with the largest income is collected, irrespective whether this person is used as the reference person at National level. In the aggregated information computed and disseminated on the Eurostat website, the reference person will be defined as the person with the highest income in the household.

6. The recording of health and education expenditures

Previous HBS experiences⁸ shown the field of health is certainly one of those where the comparability of Household Budget surveys' data is least good owing to the difference of the social protection systems in the Member States as regards both the level and the method of implementing the social welfare effort. In certain countries, the public acceptance of responsibility is such that the households' expenditure is mainly limited to the payment of taxes and social security contributions. In other countries, households have substantial expenditure e.g. recourse to private complementary insurance or the direct payment of medical

⁸ See for instance European Commission, Household Budget Surveys in the EU, methodology and recommendations for harmonization , 2003

costs. The consumption heading of health is of great importance in determining the living standard of households, thus differences in treatment can skew international comparisons. There are two possibilities of treatment of health expenditures. The first is fairly close to the concepts of national accounts; the second, although it does not lead to perfect comparability of the data in terms of 'consumption' has at least the merit of allowing a comparison of 'expenditure'.

The measurement of real consumption makes it possible to mitigate the problem of the differences between social protection systems. It consists in adding to households' net effective outgoings (i.e., real cost of health born by households after refunds) the value of the goods and health services obtained free of charge (public transfers). This value can be regarded as a benefit in kind granted by the state for the households' benefit. However such a practice gives rise to two major objections:

- Initially, it poses a problem of feasibility since it appears very difficult to consider public transfers in health at the level of the individual. This would initially require a detailed description of medical consumption in the survey questionnaires, and subsequently the use of sufficiently reliable external sources to carry out estimates. But it is probable that a large number of Member States have no such external data.
- Moreover it introduces a bias into the consumption of a particular category of the population, the elderly, who have greater need for health services without benefiting from a consequently higher standard of living.

It is primarily for reasons of feasibility that Eurostat recommends an approach based on the household's expenditure for the Household Budget surveys. This solution does not permit a very satisfactory measure of the standard of living, but it at least offers harmonization of the health expenditure concept between Member States. Various possibilities of treatment can be envisaged, approaching the concept via either (i) gross expenditure, or (ii) net expenditure.

The components of the expenditure effected by the households in the field of health are:

- expenditure refunded at a later date by the Social Security organisms or by private (complementary) insurance;
- non refunded expenditure;
- Social Security contributions;
- premiums for private insurance.

(i) The 'gross expenditure' approach would consist in totaling what households actually pay, without subtraction of any possible reimbursement by the social welfare organizations or by private insurers. In this hypothesis, social security contributions and private insurance premiums are not included since reimbursements are not deducted. On this precise point of health contributions and insurance, the national accounts nomenclature (COICOP) records only service remuneration (transaction cost) and not the whole premium.

(ii) The 'net expenditure' approach stresses the household's actual contribution, and the record covers the households' effective expenditure minus later reimbursements. Here, on the other hand, it is best to bring contributions and insurance premiums into account.

Each of these two approaches has arguments in its favor:

- The 'gross expenditure' approach is better adapted to the collection of data in this kind of survey. Frankly, it appears difficult to subtract reimbursements which are not always known at the time of data collection. In addition, measuring gross expenditure ensures greater coherence between income and consumption. Indeed, the measurement of disposable income recommended by Eurostat corresponds to the income after deduction of taxes and social security contributions. But in order to ensure a minimum of coherence between the measurement of consumption and that of income, it is advisable not to record insurance contributions and premiums as expenditure (which would correspond to the 'gross expenditure' approach).
- The 'net expenditure' approach ensures for its part a better degree of comparability between the various surveys since it aims to record the share of the expenditure which remains the household's responsibility after deduction of the benefit derived from all types of coverage (public or private). Nevertheless, this has two distinctive characteristics: first, the share of the budget devoted to health will in certain countries be very small (except for the 'public and private insurance' headings); secondly, as has already been said, it can cause practical data-collection problems.

A third 'mixed' approach could be to consider gross expenditures and the complementary private insurance premiums. This approach is used in certain countries.

Having said all this, one has to remember that there is the Health Accounts system, which when dealing with out of pocket expenditure in this field the official recommendation is to collect the net expenditure to avoid double counting.⁹

For the sake of consistency, Eurostat therefore recommends to use the net expenditure for both of these cases

Comparability issues in the structure of the HBS data files

1. Different HBS reference years

Actually, the HBS data that were transmitted to Eurostat pertain to different years depending on the countries. A distinguishing feature of EU-HBS is country data are output harmonized by inflating all the expenditures so to make them refer to Eurostat's reference year (1988, 1994, 1999, 2005 and 2010). While most of the countries actually transmitted 2004 or 2005 data, the HBS data of CY, DE and MT are older (reference years 2003, 2003 and 2000, respectively)

⁹ The SHA manual : <http://www.oecd.org/health/sha>

ESTAT data according SHA methodology:
http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1996,45323734&_dad=portal&_schema=PORTAL&screen=welcomeref&open=/health/hlth/hlth_sha&language=en&product=EU_MASTER_health&root=EU_MASTER_health&scrollto=0

Table 5: HBS reference years and date of transmission¹⁰

	Reference Year 2005 wave	Date of Transmission 2005 wave	Reference Year 2010 wave	Planned date of Transmission 2010 wave
Belgium	2005	6/07		
Bulgaria	2005	12/07		
Czech Republic	2005	6/08		
Denmark	2003, 2004 and 2005	10/07		
Germany	2003	12/07		
Estonia	2005	1/07		
Ireland	2004	2/08		
Greece	2004	1/08		
Spain	2004	1/08		
France	march 2005-march 2006	12/07		
Italy	2005	3/07		
Cyprus	2003	12/07		
Latvia	2005	2/08		
Lithuania	2005	12/07		
Luxembourg	-	5/08		
Hungary	2005	1/08		
Malta	2000	5/08		
Netherlands	2004	2/07		
Austria	September 2004 - 25 September 2005	1/07		
Poland	2005	4/08		
Portugal	Oct 2005 – Oct 2006	5/08		
Romania	2005	3/08		
Slovenia	2003, 2004 and 2005	6/07		
Slovakia	2005	3/08		
Finland	2006	3/08		
Sweden	2005	3/08		
United Kingdom	2005-2006	3/08		
Croatia	2005	3/08		
Norway	2005	4/08		

¹⁰ Source: Quality Report of the 'Household Budget Surveys' round of 2005

The use of different HBS reference years from one country to another is expected to create comparability problems, both between the HBS countries and with other statistical domains like EU-SILC or LFS, where micro-data are being collected on a more regular frequency (annual for EU-SILC, quarterly for LFS).

Table 6: HBS frequency¹¹

Belgium Annual
Bulgaria continuous
Czech Republic continuous
Denmark Annual
Germany every five years
Estonia Annual
Ireland every five years
Greece Annual
Spain Annual
France every five years
Italy annual
Cyprus Every five years
Latvia annual
Lithuania Since 2003 quarterly and annual data produced, only annual data published
Luxembourg annual
Hungary annual
Malta every five years possibly annual
Netherlands annual
Austria Every five years
Poland annual
Portugal Every five years
Romania annual
Slovenia continuous
Slovakia annual
Finland Every five years
Sweden annual
United Kingdom annual
Croatia annual
Norway annual

Besides, the provision every five years by the countries of all their annual data will increase the achieved HBS sample size, thereby making the accuracy of estimates better.

Since HBS is carried out either on an annual or a continuous basis in most of the countries EUROSTAT would like participating countries to state whether they are ready to send the data on an annual basis.

Besides this, countries are to declare when they will be ready to send the data for the next wave by completing table 5.

¹¹ Source: Quality Report of the 'Household Budget Surveys' round of 2005

2. The provision of aggregated tables by certain countries

The most commonly used approach to the transmission of the HBS 2005 data was for the countries to send a household dataset having as many rows as responding units, and containing both 'basic' and 'derived' household variables. The latter variables were in fact calculated using 'basic' variables at household and individual level.

However, five countries (Czech Republic, Italy, Poland, Portugal and Malta) were authorized to provide their HBS data in the form of aggregated tables. Even though that decision was probably justified by confidentiality issues relating to the dissemination of micro-data files, it has serious implications on EU comparability: since there is no way to create other aggregated tables than the pre-defined ones for these five countries, those are generally not accounted for when HBS aggregates are calculated. This is even more a problem as the HBS data are used to be quarried by a wide range of users.

Besides, the actual HBS documentation does not appear to be quite accurate as to how the HBS target indicators must be worked out: among other issues, how to define the income quintile groups? How to handle the indicators 'per adult equivalent'? In the absence of strict definitions of the HBS target indicators, discrepancies are likely to arise between the countries which sent micro-data files and those which provided aggregated tables.

Eurostat should get back to the five countries which sent aggregated tables for HBS 2005, asking them whether they could make micro-data files available subsequently.

3. No micro-information available at individual level

As mentioned in the above, no micro-information at individual level was to be made available to Eurostat by the countries, though the latter collect individual data on all household members. Likewise, this has implications on the comparability of the HBS data with other statistical domains like, for instance, EU Statistics on Income and Living Conditions (EU-SILC), where Eurostat does collect micro-data both at household and individual level.

The non-availability of individual information also makes the HBS data less relevant to a wide range of user needs. For instance, in Market Research, detailed analysis of household consumption patterns would demand a great deal of auxiliary information at individual level for efficient household segmentation. In the absence of data at individual level, the actual HBS data could not respond such a need.

Eurostat would like to ask participating countries to provide the list of variables that pertain to the individuals making up the family and to state whether they are in a position to send it to EUROSTAT or not.

4. The treatment of missing values in the expenditure data

The HBS expenditure data are liable to no or under-reporting errors, especially the most “delicate” items (e.g., Alcohol, Tobacco, Narcotics, Prostitution...). It is important in the micro-data files to make clear distinction between the expenditure items the households would not report and those for which no expenditure was reported during the reference period.

Eurostat recommends that 0 is used for the expenditure categories for which there were no expenditure reported, and “.” (*missing*) for the categories that were not reported.

5. The scaling of the weights

For clarity’s sake, a common practice in certain countries (BG, CY, LU, NL, NO, SI and UK) was to scale the sample weights so they average to 1. Even though this makes no difference when mean expenditures are calculated, it makes the estimation of population totals (e.g., size of the household/individual population, total consumption expenditure ...) impossible using merely the HBS micro-data.

Eurostat recommends that the household sample weights are not scaled, that is their sum should remain equal to the household population size.

Comparability issues in the HBS metadata reporting

Even though no EU Regulations exist regarding the transmission of quality reports to Eurostat and the content of such reports, most of the HBS countries (with the exceptions of IE, MT, LU, NL and TR thus far) actually delivered to Eurostat a document providing meta-information relating to their national HBS instrument. Most of the quality reports appear to have been structured according to the six dimensions as identified in Eurostat’s definition of quality, namely, Relevance, Accuracy, Timeliness and Punctuality, Accessibility and Clarity, Comparability and Coherence.

Nonetheless, the HBS national quality reports remain much less harmonized than, for instance, their EU-SILC (EU Statistics on Income and Living Conditions) counterparts, where the provision of quality reports by the countries was made mandatory (EU-SILC Commission Regulation N°28/2004 of 5 January 2004). In addition to IE, MT, LU, NL and TR which have not sent any documentation yet; the quality information provided by certain countries (PT, NO) is rather limited.

In preparation for the next round of data collection, it is important some minimum meta-information on survey quality be shared by all the HBS countries. For EU comparability’s sake, this information should focus more on the different HBS national concepts and definitions (e.g., Consumption expenditure, Imputed rent...) than on other quality dimensions (e.g., Accuracy), even though the latter is interesting too for comparability purposes.

A ‘minimum’ quality report should be prepared and made available to the HBS Countries for the next HBS wave. For simplicity sake, it should be composed of a series of summary tables, each corresponding to a given domain of study. A proposed list of summary tables is given in annex.

Eurostat is asking the countries to accept the attached list, and to convey the information requested, together with the data files.

The HBS and Core Variables

All of the proposed core variables are considered to be directly relevant to the aims of the HBS (the delivery of comparable information at EU level on consumer expenditure and consumer behaviour). All of them are feasible from the instrument as it is currently designed.

1. Variables not yet available

Currently, the following information is not collected in HBS:

- Country of birth;
- Country of citizenship;
- Economic sector in employment.

2. Variables available but not compliant with the TF's recommendations regarding the definition

Currently, the following information is collected in HBS but the definition of the corresponding variables is not compliant with the TF's recommendations:

- De facto marital status (consensual union) and de jure marital status (legal marital status): both are mixed in HBS;
- Household type (different definitions);
- Region of residence;
- Net monthly income of the household.

The table below summarizes the differences between the TF's recommendations and the variables as implemented within HBS.

Variables	TF's recommendations (variable and proposed coding)	HBS
Marital status (de jure/de facto)	Legal marital status 1. Unmarried (i.e. never married) 2. Married (including registered partnership) 3. Widowed and not remarried (including widowed from registered partnership) 4. Divorced and not remarried (including legally separated and dissolved registered partnership)	Marital status of the reference person (Var HC05) 0. Never married 1. Married or in cohabitation (consensual union) 3. Widowed 4. Divorced or separated 9. Not specified
	De facto marital status (consensual union) 1. Person living in a consensual union 2. Person not living in a consensual union	Consensual union of the reference person (Var HC05.1) 0. No 1. Yes, on a legal basis (in relation to the marriage laws of the country) 2. Yes, without a legal basis 9. Not specified

<p style="text-align: center;">Household type</p>	<p>Household size Total number of persons in household Number of persons aged less than or equal to Number of persons aged from 5 to 13 Number of persons aged from 14 to 15 Number of persons aged from 16 to 24 of which, number of students Number of persons aged from 25 to 64 Number of persons aged more than or equal to 65</p>	<p>Household size (numerical variable) (variable HB05)</p> <p><u>Calculation rule:</u> <i>HB05 = Sum of household members.</i></p> <p>Equivalent size (OECD scale) (variable HB06.1)</p> <p><i>This variable is established by allocating weighting coefficients to the household's members according to their demographic characteristics. Given the existence of big differences in the sizes and structures of households, comparability can be improved by using expenditure or income by adult equivalent.</i></p> <p><i>The OECD scale consists in allocating the following weightings to persons in the calculation of the "equivalent household's size".</i></p> <ul style="list-style-type: none"> •first adult in the household = 1.0 •each adult thereafter (aged over 13) = 0.7 •each child (13 or under) = 0.5 <p><u>Calculation rule:</u> $HB06.1 = 0.3 + (0.7*A) + (0.5*B)$</p> <p><i>A = Sum of household members where age (in completed years) of household members > 13</i> <i>B = Sum of household members where age (in completed years) of household members < 14</i> <i>(Please note that the first adult of the household counts by 1 because of the addition of the constant term 0.3, assuming that each household must have at least one adult).</i></p> <p>Modified equivalent size (OECD scale) (variable HB06.2)</p> <p><i>The modified scale was developed on the basis of the argument that the original scale of the OECD gave relatively too much weight to additional persons.</i></p> <ul style="list-style-type: none"> •first adult in the household = 1.0 •each adult thereafter (aged over 13) = 0.5 •each child (13 or under) = 0.3 <p><u>Calculation rule:</u> $HB062 = 0.5 + (0.5*A) + (0.3*B)$</p> <p><i>A = Sum of household members where age (in completed years) of household members > 13</i> <i>B = Sum of household members where age (in completed years) of household members < 14</i> <i>(Please note that the first adult of the household counts by 1 because of the addition of the constant term 0.5, assuming that each household must have at least one adult).</i></p>
--	---	---

<p>Household type 1 One-person households 2 Multi-person households 2.1. Lone parent with child(ren) aged less than 25 2.2. Couple without child(ren) aged less than 25 2.3. Couple with child(ren) aged less than 25 2.4. Couple or lone parent with child(ren) aged less than 25 and other persons living in household 2.5 Other type of household</p>	<p>Household type Categorisation of the household's members (variable MB09.1) 1 young child from 0 to 13 years old 2 child aged from 14 to 15 years old 3 'older child' 4 'adult child' 5 adults not living in union 6 adults living in union 9 not specified</p> <p>Categorisation of the household's members (variable MB09.2) 1 young child from 0 to 17 years old 2 adult child 3 adults not living in union 4 adults living in union 9 not specified</p>
<p>Economical activity Number of persons aged 16-64 in household who are at work Number of persons aged 16-64 in household who are unemployed or are economically inactive</p>	<p>Current activity status of household members (variable ME01) ⇒economically <i>active</i> 1 working 2 with employment but temporarily absent 3 unemployed ⇒economically <i>inactive</i> 4 retired 5 student or in national service 6 non economic activity, housewife 7 unable to work 8 not applicable (legal age to work unfulfilled) 9 not specified</p> <p>Usual activity status of household members (variable ME07) ⇒economically <i>active</i> 1 working (incl. Temporarily away from job) 3 unemployed ⇒economically <i>inactive</i> 4 retired 5 student or in national service 6 non economic activity, housewife 7 unable to work 8 not applicable (legal age to work unattained) 9 not specified</p>

Region of residence	Region of residence Use NUTS 2	Region (variable HA08) use NUTS 1 00 For the countries with no division for NUTS 1 99 Not specified
Net monthly income of the household	Net monthly income of the household (value in national currency) Unequalised household total net monthly income (value in national currency) Decile 1-10 cut-off points Equalised household total net monthly income per capita (value in national currency) Decile 1-10 cut-off points	Income in kind from employment (wages and salaries in kind) (variable HH01.2) Income in kind from non-salaried activities (variable HH02.3) Imputed rent (variable HH03.2) Monetary net income (total monetary income from all sources minus income taxes) (variable HH09.5) Net income (total income from all sources including non-monetary components minus income taxes) (variable HH09.9 = HH09.5 + HH01.2 + HH02.3 + HH03.2)

3. Variables available but not compliant with the TF's recommendations regarding the technical recommendations

Currently, the following information is collected in HBS but the TF's recommendations are not followed for these variables:

- Age (*in HBS, no question is asked regarding the data/year of birth (the collection of such a data is recommended by the TF)*);
- Highest level of educational attainment (in HBS, levels ISCED1 and 2 are aggregated, levels 3 and 4 are aggregated and levels 5 and 6 are aggregated);
- Current self-declared labour status (the categories 'students' and 'in national service' are aggregated and no information is collected on whether the profession is full-time or part-time, the categories "permanently disabled" and "other inactive persons" are missing).

It is being proposed that all countries introduce all the core variables from 2010 onwards
--

Conclusion

The Working Group is invited to accept the recommendations made in the end of each section for this paper, namely:

- It is expected that all possible efforts are done to improve the sample size, especially in those countries which have a very small sample
- For the next HBS round, all the countries should provide the imputed rent variable, as per ESA 95 guidelines and as already agreed too for the HBS 2005 wave. This is even more possible as from 2007 onwards that variable is mandatory for EU-SILC (EU Statistics on Income and Living Conditions). Priority is to be given to:
 - (1) Stratification or regression;
 - (2) over self-assessment and user-cost
- For the next HBS round, Eurostat should recommend all the countries to implement probabilistic sample selections. As regards Czech Republic and Germany, Eurostat should enquire whether or not this proposal can be implemented in these two countries within their own national statistical systems
- In order to make the treatment of nonresponse in the Household Budget Surveys more efficient, national good practices in dealing with nonresponse should be shared with all the countries. National practices may comprise preventive measures (better training of the interviewers, information letter sent to the households before data collection...), or weight adjustments (post-stratification, logistic regression...) to reduce bias. Eurostat invites the countries which achieved more than 70% response, namely Croatia, Cyprus, Italy, Romania, Slovakia and Slovenia to prepare a short paper describing their experiences, and this will be circulated to the other countries for their consideration

- The Head of household/reference person is to be nationally defined, i.e. it can differ from one country to another. But Eurostat recommends that all relevant information about the adult (16+) with the largest income is collected, irrespective whether this person is used as the reference person at National level. In the aggregated information computed and disseminated on the Eurostat website, the reference person will be defined as the person with the highest income in the household
- For the sake of consistency, Eurostat recommends to use the net expenditure for both Health and Education expenditure.
- Since HBS is carried out on, either an annual or a continuous basis in most of the countries EUROSTAT would like participating countries to state whether they are ready to send the data on an annual basis.
Besides this, countries are to declare when they will be ready to send the data for the next wave by completing table 5
- Eurostat should get back to the five countries which sent aggregated tables for HBS 2005, asking them whether they could make micro-data files available subsequently
- Eurostat would like to ask participating countries to provide the list of variables that pertain to the individuals making up the family and to state whether they are in a position to send it to EUROSTAT or not.
- Eurostat recommends that 0 is used for the expenditure categories for which there were no expenditure reported, and “.” (*missing*) for the categories that were not reported.
- Eurostat recommends that the household sample weights are not scaled, that is their sum should remain equal to the household population size
- Eurostat is asking the countries to accept the attached list, and to convey the information requested, together with the data files.
- It is being proposed that all countries introduce all the core variables from 2010 onwards.

Annex1 : HBS “minimum” quality report

This annex presents the tables which could be provided to Eurostat by the countries as minimum quality information on HBS.

Table 1: Overall information

Title of the survey	Name of the household budget survey used at national level
Reference years	Reference year of the HBS data as collected by the countries. In case HBS data pertaining to several years are provided, all the reference years must be reported
Frequency	Frequency at which HBS is conducted at national level: continuous, annual, every five years or other
Achieved sample size	Total number of households which were successfully surveyed (interviews+diaries)
EUR/NAC conversion factor	For the euro zone countries, the rate is 1. Otherwise, HBS data on income and expenditure are collected in NAC and has to be converted into Euros for dissemination at Eurostat level by applying a conversion factor
NAC/PPS conversion factor	Since most of the HBS expenditure aggregates are expressed in PPS, a NAC/PPS conversion factor must be applied to the expenditure data

Table 2: Main sampling characteristics

Ultimate sampling unit	The Household Budget surveys collect information on households and persons as well. In many countries, the household itself constitutes the ultimate sampling unit (USU) but, on the other hand, households are only “built” around the selected addresses, dwellings or persons in some other countries. The latter involves enumerating all the households occupying an address or a dwelling, or having at least one sample person as current member. Consequently, the ultimate sampling unit is the address, the dwelling or the person
Probability sampling	Any HBS sample should meet the basic requirement of probability sampling in design. By probability sampling is meant a sampling scheme in which each unit in the study population is given a known, non-zero probability of selection into the sample
Number of sampling stages	Most HBS sample designs involve the selection of the sample in multiple stages. A common practice is to use a two-stage selection. First, a sample of geographical areas is selected, typically with probabilities proportional to size. The second stage consists of the simple random selection, within each sample area, of households or addresses for inclusion in the survey. On the other hand, some countries draw their HBS samples using a single-stage sampling scheme involving direct selection of addresses or households. Finally, non-probabilistic schemes (quotas) were also met in previous HBS rounds
Stratification criteria	Most countries stratify the sample according to certain criteria, thus hoping to make it more accurate. Common stratification criteria are region, socio-economic status of reference person, household size or type
Over-sampling of special domains	In most cases, the entire population is sampled uniformly and the sample is distributed proportionately across different areas and population groups in the country. However, some population groups of special interest may happen to be over-sampled

<p>Survey population: main exclusions</p>	<p>All Household Budget surveys are confined to the population residing in private households. Collective or institutional households (old persons' homes, hospitals, hostels, boarding houses, prisons...) are excluded, as are generally homeless people. As to geographical coverage, most HBS cover the entire population residing in private households in the national territory. However, for reasons of cost and accessibility, some remote areas with very small populations may be excluded. Occasionally, some more important groups have also been excluded</p>
<p>Sampling frame</p>	<p>Basically, there are three main sources for sample selection: <u>Registers</u>: registers can provide up-to-date lists of households or individuals, with many relevant characteristics useful for stratification and efficient selection of the sample. <u>Use of an area frame</u>: another way to select a sample is to obtain a sample of area units from a suitable source such as population census or a master sample of areas. In the areas selected, lists of addresses, households or persons may be prepared or updated from other sources in order to complete the process of sample selection. <u>Use of an existing larger sample</u>: if the larger sample obtains relevant auxiliary information, that information can be used to improve the quality of the sample (stratification, weighting for non-response...). On the other hand, using households which have been already included in another survey increase the respondent burden. It may also damage the representativity of the sample, because it is certainly difficult to include households which have already refused to co-operate in an earlier survey.</p>
<p>Whether substitutions are allowed</p>	<p>Because of high rates of non-response, it is common in Household Budget Surveys to substitute for non-respondents in the sample that was originally selected. Substitutions are usually made after matching with characteristics of the non-respondent to the extent possible.</p>

Table 3: Sample size and non-response errors

<p>Gross sample size</p>	<p>By "gross" sample is meant the sample that is initially drawn from the sampling frame. Generally, a gross sample contains both eligible units and units which are not eligible for the survey (dwellings which do not exist, vacant...). At this stage, sampling units may be households, dwellings, addresses or persons</p>
<p>Number of eligible units</p>	<p>It refers to the total number of sampling units initially selected minus those which are not eligible for the survey (dwellings which do not exist, vacant...)</p>
<p>Number of units successfully contacted – BEFORE SUBSTITUTION</p>	<p>This is the total number of eligible sampling units which have been successfully accessed</p>
<p>Number of units successfully contacted – AFTER SUBSTITUTION</p>	

Number of responding households – BEFORE SUBSTITUTION	This is the total number of households which have successfully cooperated to the survey, both to the interviews and the completion of the diaries. The practice of making substitutions for non-responding households further complicates the picture concerning response rates. Simply computing response rate as the ratio of the number of households completed to the number contacted will provide an over-estimate in the presence of substitution
Number of responding households – AFTER SUBSTITUTION	

Table 4: Weighting

Calculation of the household design weights	The household design weights lead to unbiased values for totals estimated from the gross sample. They are generally calculated for each sampling unit as the reciprocal of the probability of selection of the unit
Weight adjustments for non-response at household level	Description of any weight adjustments that are done to correct non-response at household level
Weight adjustments to external data sources (calibration)	Description of any weight adjustments to external data sources. In most cases, the calibration technique makes the accuracy better, therefore, is widely used to improve the quality of survey data
Any other weight adjustments	For instance, trimming or top(bottom)-coding of the weight distribution so to adjust for outliers

Table 5: Estimated standard errors, confidence intervals and design effects

Indicator	Name of the target HBS indicator for which the precision is wanted At least, estimated sampling errors should be provided for the mean total household consumption expenditure as well as for the following breakdowns: One-digit COICOP categories; age of the household's reference person: less than 30, 30-44, 45-59 and 60+ years; household type: single person, two adults, three adults or more, single parent with dependent children, two adults with dependent children, three or more adults with dependent children; Employment status of the household's reference person: manual worker in Industry and Services, non-manual worker in Industry and Services, self-employed, unemployed and other inactive
Achieved sample size	Achieved sample size (in number of households) for the indicator
Estimated value	Estimated value (weighted) of the indicator
Estimated coefficient of variation (%)	Estimation coefficient of variation (%) of the indicator
95% Confidence interval – lower bound	Lower and upper bounds of the confidence interval at 95% level of the indicator
95% Confidence interval – upper bound	
Estimated design effect (Deff)	By definition, the design effect (Deff) is the ratio of the variance under the actual sampling design to the variance that would be obtained under simple random sampling of same size

Table 6: Household, household membership and head of household

Household defined as persons sharing	Accommodation	(Y/N)
	Expenditure	(Y/N)
	Income	(Y/N)
	Family emotional ties	(Y/N)
	Other (indicate)	
Household membership	Usually resident, related to other members	(Y/N)
	Usually resident, not related to other members	(Y/N)
	Resident boarder, tenant, lodger	(Y/N)
	Visitor	(Y/N)
	Live-in domestic servant, au pair	(Y/N)
	Resident, absent from dwelling in the short-term	(Y/N)
	Children in household, in education away from home	(Y/N)
	Long-term absence with household ties: working away from home	(Y/N)
	Temporary absence with household ties: in hospital, nursing home or other institution	(Y/N)
Head of household (indicate)	<p>In each household it is necessary to identify as its head a particular individual whose personal characteristics can be used in the classification and analysis of information on the household. The social group, occupation and employment status, income, sex and age...of the head are often used to classify the sample households for presentation of the results and for weighting classifications used in the derivation of survey estimates.</p> <p>A common practice is to consider as the head the person designated as such by the household concerned. More objective and specific criteria can be used, such as the person contributing most to the income of the household, the person owning or renting the household accommodation, or the oldest active male.</p>	

Table 7: Consumption expenditure approach

Consumption expenditure approach	Actual final consumption	(Y/N)
	Final consumption expenditure	(Y/N)
	Monetary consumption expenditure	(Y/N)
	Other (indicate)	
	Consumption expenditure approach: Main exclusions	
	Reference periods for expenditure	
Borderline cases: recording and valuation	Goods or services for own final consumption	
	Leasing and hire purchases	
	Health and education expenditures	
Estimation of imputed rentals: population	Owner-occupiers: Principal dwellings	(Y/N)
	Owner-occupiers: Secondary dwellings	(Y/N)
	Tenants: reduced or provided rent-free	(Y/N)
Estimation of imputed rentals: method	Self-assessment	(Y/N)
	Stratification	(Y/N)
	Log-linear regression	(Y/N)
	Heckman regression	(Y/N)
	Other (indicate)	
	List of the variables used in the model:	
Salaries and wages in-kind: which benefits are evaluated	Free or reduced cost housing	(Y/N)
	Private use of company car	(Y/N)
	Gas, electricity or water	(Y/N)
	Telephone	(Y/N)
	Other (indicate)	

Table 8: Data collection

Diaries	Recording unit	The period for which a diary is maintained is called the recording period and its duration and distribution over time is the primary determinant of the structure of the survey. The recording period may be of 'fixed' type, i.e. defined in terms of the same calendar period for all households in the sample. However, in most cases a so-called 'moving' reference period is used, the exact timing of which is different for different households in the sample. The recording periods are staggered over the survey period to even out the effect of seasonal and other temporal variations for the sample as a whole. This is done by dividing the sample into a number of sub-samples, and distributing these sub-samples in terms of the recording period uniformly over the survey period
	Recording period	
	Distribution of recording periods over the survey year	
	Items covered in the diary	
Collection of substantive information	Instrument	A sample household may be subject to more than one interview during the time it is in the survey. The common pattern in Household Budget surveys is to have two interviews per household arranged around the recording period. Mostly these consist of an initial interview before the recording period and a final interview after the recording period. Generally, background characteristics are covered in the first interview and income questions in the second. However, other instruments may be used for the collection of substantive information, mostly registers and/or questionnaires (household/individual)
	Recording unit	
	Items covered	

Table 9: Summary statistics on the net household income / comparison between HBS and EUSILC

	HBS	EU-SILC (EU Statistics on Income and Living Conditions)
Mean		
Min		
Max		
Median		
CV (%)		

Table 10: Distribution of the population by age group and gender / comparison between HBS, EUSILC and LFS

Age Groups	HBS	EU-SILC (EU Statistics on Income and Living Conditions)	LFS (Labour Force Survey)
15-24			
25-34			
35-44			
45-54			
55-64			
65+			