

**Expert Group Meeting on
Setting the Scope of Social Statistics**
United Nations Statistics Division
in collaboration with the Siena Group on Social Statistics
New York, 6-9 May 2003

Harmonisation of Concepts, Definitions and Classifications*

by

Howard Meltzer**

* This document is being issued without formal editing.

** Office for National Statistics, United Kingdom. The views expressed in this report are those of the author and do not imply the expression of any opinion on the part of the United Nations Secretariat.

Harmonisation of concepts, definitions and classifications

Howard Meltzer, Office for National Statistics, UK

1. Standards of harmonisation

The draft protocol on Statistical Integration as part of the National Statistics Code of Practice promotes standards for the harmonisation of:

- Classifications – geographical, social and economic.
- Statistical units – business (enterprise, company, establishment), social (family, census family, economic family, household, dwelling).
- Definitions – standard concepts and variables.
- Harmonised questions and question modules.
- Frameworks – common frameworks.
- Variable names.

The aim of harmonising data over time, within sources and across sources is to minimise:

- Time and costs in developing data collections.
- Respondent load.
- Unnecessary duplication across sources.

and maximise:

- Quality and value of information gained from any single source.
- Opportunities for the cross-analysis, exchange and re-use of data.
- Users' understanding of statistical information.

Where the harmonisation process has been initiated, those responsible for standards and classifications will need to:

- Provide support and advice to users on the standards and classifications for which they are responsible.
- Undertake regular reviews of the standards and classifications for which they are responsible and integrate these reviews with the National Statistics Quality Review process. The aim of these reviews will be to ensure that standards are still relevant and to achieve a balance between reflecting contemporary circumstances and maintaining consistency over time.
- Publicise and disseminate any updates or revisions as soon as possible.
- Develop and promote further standards in order to facilitate a harmonised approach to business and social surveys, and, where possible, to administrative sources.

2. Harmonisation across social surveys

The United Kingdom has a wide range of Government surveys of persons and households which provide sources of social and economic statistics. The decennial Census of Population is the largest and best known but in addition Government departments commission continuous household surveys on a range of topics. These include economic activity (*Labour Force Survey*), income (*Family Resources Survey*), expenditure (*Expenditure and Food Survey*), health (*Health Survey for England*), housing (*Survey of English Housing*) and transport (*National Travel Survey*), as well as the multi-purpose *General Household Survey* which links many of these topics and others such as education. There are also several large-scale surveys which are repeated regularly, such as the *British Crime Surveys*, the *Dental Surveys* and the *House Conditions Surveys*. The Government also commissions single surveys from time to time on subjects of national importance such as the prevalence of disability and psychiatric morbidity.

These surveys were designed at different times, to meet different needs, and have been commissioned by a range of ministries. Consequently, the surveys were developed to a significant degree in isolation from each other. This resulted in a lack of cohesion, with differences arising in concepts and definitions, in design, in fieldwork and processing practices and in outputs. This lack of cohesion was a source of frustration for many users. A major factor behind the creation of National Statistics in 1996 was the desire to improve the service being provided to users of statistics, and the availability and accessibility of those statistics.

Development plans for the Office for National Statistics (ONS) include improving access to statistics across Government by establishing a central database of key economic and social statistics, produced to common classifications, definitions and standards; and improved comparability of social statistics. Harmonised concepts represent a step towards this goal by making the interpretation and analysis of these data easier, so that they allow users of published sources to see a more coherent picture of British society and trends within it, and to help users plan surveys (e.g. at a local level) that can provide data comparable with national surveys.

ONS is increasingly positioning itself to make greater use of administrative data. In order to exploit such data it will be important to extend the principles of harmonisation beyond surveys and the Census. This is likely to be the focus of new harmonisation activities over the next few years.

Other factors influencing harmonisation activities may be summarised as follows:

- Eurostat's requirements – the desirability for the UK to be working with Eurostat in developing harmonised Key Social Indicators (KSIs, which are at the heart of Europe-wide social reporting).

- Ever-increasing emphasis on coherence and priority given to the need to develop harmonised concepts for topics such as households, income and urban/rural, sub-groups within households.
- Emerging social topics, such as social capital, e-society and cultural identity.
- The development of an Integrated Social survey which will pull together several continuous surveys

3. Principles of harmonisation in social surveys

Harmonisation concerns concepts which are *inputs* (i.e. interview questions and answer categories) or *outputs* (i.e. analysis variables derived from the inputs) or both (e.g. the question on sex).

3.1 Inputs

Different surveys have different purposes and hence cover topics in different depth. Harmonised questions are designed to provide the recommended minimum information to allow common classifications and facilitate the analysis of data from different surveys in combination. Not all surveys will include questions on all topics or in every year, but the recommendation is that where a topic is covered, harmonised questions should be included wherever possible.

Some surveys will require further detail on topics than can be obtained from the harmonised questions alone. It will normally be the case that such surveys already ask for that detail. The harmonised questions have been designed so that the surveys which ask for more detail can either derive them, without asking them directly, or combine them with the further detail, without adding to the length of interview.

The idea that a survey for which it is appropriate to derive an equivalent variable to a harmonised question should additionally ask the harmonised question directly has been considered and rejected. To do so would duplicate effort and seem inappropriate to respondents in the context of the interview.

Harmonisation involves some compromises, since surveys' prime concerns vary so widely. For example, surveys vary in the extent to which they allow information to be given by one respondent on behalf of another who is absent at the time of interview. It would be unrealistic to expect ministries which currently accept proxy data as adequate to their needs on certain of their surveys to find the resources to harmonise on data given in person.

Harmonised questions are intended to fit flexibly into the designs of different surveys. There is no intention that they should form a unified sequence within a questionnaire. Questions and groups of questions are intended to be placed in existing questionnaires in the most appropriate places. This will often mean substituting a harmonised question for an existing one on the same topic.

Minimising respondent burden has been a major consideration in designing harmonised questions. Some of the harmonised questions have more detailed sets of answer categories than some of the current surveys use for these topics. However, classifying respondents' answers to a more detailed set does not necessarily increase the time needed to answer a question. More detailed categories have been included only where they will not add to interview length. There is no intention to probe for detail which is not volunteered, unless a question specifically demands such probing. The aim has been to save time by providing clear categories for the rarer answers, where these are of interest for analysis. However, account has also been taken of the need to make it easy for interviewers and respondents to find the major answer categories and not to lose them in a host of details. Finding the right balance in such compromises has an important bearing on survey quality.

3.2 *Outputs*

While it is neither feasible nor desirable to produce outputs from different surveys in a completely harmonised way, it is hoped that unless there are strong reasons for doing otherwise, Government surveys will adopt the harmonised outputs, as well as questions, so far as possible. This will not only allow users to interpret data more easily but also help those outside Government better to plan their own data collection and analysis. However, as noted above, it will be for the commissioners of individual surveys to decide how outputs should be presented, bearing in mind the needs of their users.

Of course, not all surveys will be able to provide information in full detail or for all tabulations, owing to the limitations of sample size, but it is hoped that so far as possible analyses will use a standard hierarchy, so that results can be compared. For some outputs, where the categories are likely to vary considerably in size (such as ethnic groups), the commentary suggests how groups might be further amalgamated on a harmonised basis. It is suggested that information included in the notes attached to the harmonised output categories be included in each publication where the categories are used (though not necessarily in every table); and that those publications also refer to this website entry for further clarification. Where output categories differ from the harmonised ones, it is also hoped that this will be made clear.

The output categories may be applied to households or individuals as appropriate, or to households containing one (or more) individuals with the characteristic specified. It should be made clear in the table heading or documentation what the unit of analysis is.

Harmonised outputs are not intended to restrict published outputs; where more detail is currently published from a particular survey, it is expected that this will continue, though including some aggregate data on a harmonised basis. Users who have access to micro data (for example, through the Data Archive) should not generally be affected by the harmonisation of survey outputs, except where it impacts on classification systems themselves, or where it is suggested that electronic databases should include harmonised categories.

4 Practical assumptions for harmonised questions

While it might be desirable that harmonised questions should be asked in exactly the same ways on different surveys, it is recognised that this may not be achievable. The emphasis in the harmonisation project is on harmonising question wording, answer categories and the subsamples to whom the questions are addressed. Matters such as question sequence (for the factual type of question involved), the use of proxy respondents, and (at a rather more minor methodological level) use of specific kinds of grids were regarded as secondary in the sense that achievement of the main type of harmonisation would be worthwhile even if the second were not practicable. Documentation of the questions used, and their routing, can be found in the published reports of surveys and may also be available from commissioning ministries.

These differences in the way in which data are collected, even using harmonised survey questions, are extremely important. ONS research has pointed to a variety of factors which lead to different estimates for harmonised questions – factors including:

- question wording and context effects
- definitive differences
- non-response bias
- geographical coverage
- sampled population
- mode effect
- acceptance of proxy information
- treatment of multi-households
- unit of analysis
- field procedures
- item non-response
- time period
- organisational effects

5. The scope of harmonised concepts/questions in social surveys

Harmonisation which extends to all or nearly all major Government household surveys, can be thought of as covering a *primary set* of concepts and questions. Concepts and questions which apply only for a selected group of surveys, can be thought of as belonging to a *secondary set*.

5.1 *The primary set of harmonised concepts and questions*

Common definitions of person and household response units are vital steps towards harmonisation. For Government surveys, there is already a standard definition of adults as persons aged 16 years or more. Definition of the household response unit has differed

between surveys. Most use the household definition which was adopted in the 1981, 1991 and 2001 Censuses of Population, which focuses on shared living accommodation.

The topics covered by harmonised outputs and inputs in the primary set are:

- survey reference period
- geography
- demographic information, household composition and relationships
- national identity and ethnic group
- tenure
- economic status
- industry, occupation, employment status and socio-economic class
- full or part-time work (self-assessed)

5.2 The secondary sets of harmonised concepts and questions

The secondary sets of concepts and questions have been based on the shared interests in particular topics of different groups of surveys.

The topics covered in the secondary sets, and the surveys on which harmonised questions are asked, are:

- social security/benefits
- consumer durables (derived directly from harmonised inputs)
- detailed income for substantive analysis
- selected job details
- accommodation type of the household
- housing costs and benefits
- length of residence
- household motor vehicles
- general health
- carers
- educational attainment
- qualifications