

**WORKING PAPER No 6
1 June 2005**

ENGLISH ONLY

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR
EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES
(EUROSTAT)**

**CONFERENCE OF EUROPEAN
STATISTICIANS**

**FOOD AND AGRICULTURAL
ORGANISATION (FAO)**

**Joint UNECE/EUROSTAT/FAO/OECD
Meeting on Food and Agricultural Statistics
in Europe
(Rome, 29 June-1 July 2005)**

**ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT
(OECD)**

CHAPTER 6

OF THE HANDBOOK ON RURAL HOUSEHOLD, LIVELIHOOD AND WELL-BEING: STATISTICS ON RURAL DEVELOPMENT AND AGRICULTURE HOUSEHOLD INCOME.

Paper submitted by the Task Force
on Statistics for Rural Development and Agriculture Household Income*

* The Task Force is comprised of experts from the following national agencies, universities and international organizations: Statistics Canada, Hungarian Central Statistical Office, National Statistical Institute of Italy (ISTAT), Swedish Board of Agriculture, Dept. for Environment, Food and Rural Affairs (United Kingdom), Economic Research Service (United States), Imperial College (United Kingdom), University of Verona (Italy), University of Pescara (Italy), Food and Agriculture Organization of United Nations (FAO), World Bank, Statistical Office of the European Communities (Eurostat), Organization for Economic Co-operation and Development (OECD) and the United Nations Economic Commission for Europe (UNECE).

VI DATA SOURCES, METHODOLOGIES AND MEASUREMENT ISSUES

VI.1 Introduction

Rural development statistics is concerned with the measurements of a multitude of heterogeneous variables, in the areas of demography, geography and territorial descriptions, environment and above all the whole spectra of economic and social well-being, from a multitude of different sources. In fact, rural development statistics draw from almost all statistical fields. Yet, few of the agencies responsible for these statistics have rural areas in mind when designing their statistical systems. This creates two major problems: first, the **various data sources are often not designed in such a way that they can accommodate the various requirements of rural development statistics**, in particular sufficient geographical background. Labour and educational statistics, to take a few examples, might be very well be timely and have detailed breakdown on types of employment and education but only on a national level or in the best case major regions and not on smaller administrative levels such as communes, which are the basic units in rural development statistics.

The second problem is that rural development statistics has to **combine data from different sources**, derived from differences in definitions, coverage, precision, measurement period and periodicity. Statistics on employment may differ depending on the source: household surveys, establishment surveys, or administrative records. The problem is compounded when data from two or more sources are combined. Only to combine national accounts data with employment and foreign trade data is problematic despite the close complementary nature of these statistics. When constructing rural development indicators, in which the nominator and the denominator are based on different data sources, much attention must be paid to explaining the consequences of this on the results of the indicator. The combination of register data and survey data is an example of what was mentioned above.

A good illustration of this problem is the ILO's new agricultural labour productivity indicator (KILM 18e), which is largely derived from the Groningen Growth and Development Centre (GGDC) Database at the University of Groningen (the Netherlands).¹

For OECD economies, most of which are included in the KILM 18 tables under the headings of "major Europe" and "major non-Europe", GDP and labour compensation are mainly obtained from OECD: *National Accounts*, Volumes I and II (annual issues). Employment estimates for the aggregate economy are mostly taken

¹ The GGDC has long-standing expertise in development and analysis of data on productivity performance, in particular on comparisons of levels of productivity by sector and industry. A full documentation of sources and methods by economy and underlying documentation on the use of PPPs, etc., can be downloaded from the Internet site of the GGDC. Website: <http://www.eco.rug.nl/ggdc/>. See also B. van Ark and E.J. Monnikhof: *Productivity and unit labour cost comparisons: A data base*, Employment Paper 2000/5 (Geneva, ILO); website: <http://www.ilo.org/public/english/employment/strat/publ/ep00-5.htm>.

from OECD: *Labour Force Statistics* (annual issues), and for the individual sectors – as far as available – from OECD: *National Accounts*, Volume II (annual issues) and OECD: STructural ANalysis (STAN) database. The estimates available from the OECD, which originally were obtained from national statistical offices and, where possible, harmonized for differences in concepts and industry classifications, have been supplemented, where necessary, with national accounts statistics obtained directly from the individual economies. For some economies, the database of the United States Bureau of Labor Statistics (BLS) was used, in particular for estimates on employment and for manufacturing.

For non-OECD economies, the national accounts and labour statistics publications of individual economies were often taken as the point of departure. The statistics from these sources were used to supplement statistics from international organizations such as the World Bank, the Asian Development Bank, the ILO and the United Nations Statistical Office. In the estimations for agriculture, forestry and fisheries, international sources served as the point of departure. Intensive use was also made of the FAO Database from the Food and Agriculture Organization of the United Nations (FAO). However, where it was possible to disaggregate the group to provide estimates for agriculture and forestry (without fisheries) and for agriculture (without forestry and fisheries), data came mostly from national sources. In addition, Angus Maddison (1995) provides benchmark estimates of annual hours worked for a significant number of non-OECD economies.

The estimates for agriculture, forestry and fisheries (AFF), which are of an experimental nature and include as many as 112 economies, are based on measures of PPP taken from an FAO database of prices received by farmers for about 180 products in 1995. These prices refer to farm-gate prices or first-point-of-sale prices, and in principle do not include transport costs or the profit margins that generally accrue to intermediaries. The PPPs for the farm sector are assumed to be representative of the fisheries and forestry industries. As for the aggregate economy, the PPPs for AFF are multilateral, using a Geary-Khamis weighting system. In contrast to the estimates for the aggregate economy and for the other sectors, the estimates for AFF are not backdated to provide them at a 1990 price level.²

When comparing rural indicators between countries other problems are added. Countries tend to use different definitions of rurality and degree of rurality as well as of key concepts such as household. Even among rather homogenous countries, for instance in the European Union, there is a wide span in the definition of household: people living under the same dwelling whether or not there are family ties, sharing of expenditures, pooling of income, the existence of family or emotional ties or a

² Sources: KILM, 2004. "7. Labour productivity and unit labour costs indicator (KILM 18)." 3rd Edition; http://www.esds.ac.uk/international/support/user_guides/ilo/kilm18EN.pdf; and Bart van Ark and Erik Monnikhof, 2000. "Productivity and unit labour cost comparisons: a data base." ILO EMPLOYMENT PAPER 2000/5 <http://www.ilo.org/public/english/employment/strat/download/ep5.pdf>

combination of one or several of these characteristics. There might also be big difference in the child-adult definition and how much weight is given in the household aggregate (see chapter ...). Also when it concerns the variables for which the household is characterized there is a great span in the definition among countries. This speaks for being very careful when comparing indicators levels between countries and suggests a **focus on comparing changes in indicator levels between countries**, in particular if there is relative national stability in the calculation of indicators over time. To take a comparable example. While price statisticians might be rather confident in presenting estimates of price changes for country comparisons, they are not that keen in comparing levels because equals are not compared to equals.

VI.2 Data sources

VI.2.1 Population and housing censuses

Censuses covering the total population and housing stock are indispensable for providing statistics on the population, family, household and housing situation on a uniform basis for small areas and sub-groups. The characteristics of the population include geographic, demographic, economic, educational and household and family characteristics. For many countries, a traditional census is vital for providing such information and often there are no viable alternatives. Registers and other administrative sources are an alternative to the traditional census as far as they contain the relevant topics, definitions and classifications and cover the entire population. Sample surveys used alone cannot provide equivalent data for small areas, but they can be used in combination with a census or to supplement census information on specific topics.

For rural development statistics, censuses are an invaluable source of data because of their complete coverage of the population, at least for basic demographic statistics, which permits a breakdown in small geographical units. The main disadvantage is the long time periods between the individual censuses, normally five years. Another disadvantage is that they are usually not sufficiently detailed as concern expenditures, consumption and income patterns. This can be attenuated, however, by conducting household budget surveys linked as sub-samples to censuses (see section below).

In view of the importance of censuses as data source for rural development statistics, a brief account of internationally recommended standards is given below.

For censuses there are a number of international recommendations, which provide guidance and assistance to countries in planning the content of their census; and to facilitate and improve international comparability through the harmonisation of data, definitions and classification of topics. To this end readers may wish to consult:

Recommendations for the 2000 Censuses of Population and Housing in the ECE Region. Jointly prepared by the United Nations Economic Commission for Europe and the Statistical Office of the European Communities, United Nations, New York and Geneva, 1998. Statistical Standards and Studies – No.49.

Principles and Recommendations for Population and Housing Censuses, Statistical Papers, Series M, No.67/ Rev.1, United Nations.

The scope of the UNECE/Eurostat joint recommendations is limited to the following three elements: (a) a list of “core” topics which countries should cover in their censuses, and recommended definitions and classifications for each of the core topics; (b) a list of “non-core” topics which countries may wish to consider including in their censuses, and suggested definitions and/or classifications for some of the non-core topics, and (c) a basic set of tabulation programmes.

The data collection could involve both a short form (with selected questions) and a long form (with more questions). The long form is completed for a sample of households or people. Alternatively, one form could be used, but when this is done, a sample is often selected for processing certain questions: for example, those which could be costly to process, such as industry and occupation.

Some countries are using registers and other administrative sources, together with information from sample surveys, to provide census-type statistics. Other countries are able to collect all relevant information by combining data from different registers. Still other countries can get part of the information from registers and other administrative sources - often information on persons - but have to supplement these by using questionnaires, particularly as there are difficulties in putting new items required for statistical purposes into administrative registers.

Six units of enumeration may be used in population and housing censuses tables:

- (a) persons;
- (b) private households;
- (c) institutional households;
- (d) family nuclei;
- (e) living quarters (i.e. housing units and living quarters other than housing units, such as institutions); and
- (f) buildings.

In statistical terms, these concepts are clearly distinguishable, and the terms are not themselves interchangeable. Some households contain more than one family, several households may live together in a housing unit and, exceptionally, a single household may occupy more than one housing unit as its usual place of residence. Similarly, a building may contain several housing units and a housing unit may, exceptionally, be located in more than one building, e.g. in the case of dwellings comprising a main building and a room or rooms above a detached garage that are clearly designed to be used as part of the dwelling.

The UNECE/Eurostat Recommendations contain the following 29 core topics, listed below, and a few of them of particular importance for rural development statistics are briefly annotated (there are also 52 non-core topics recommended for countries that have the means of covering them).

Geographic characteristics of persons

- 1. Place of usual residence**
- 2. Place of usual residence one year prior to the census**

Demographic characteristics of persons

- 3. Sex**
- 4. Age**
- 5. Legal marital status**
- 6. Country/place of birth**
- 7. Country of citizenship**

Economic characteristics of persons

8. Current activity status

This is the current relationship of a person to economic activity, based on a brief reference period, preferably the previous week. The use of the "current activity" is considered most appropriate for countries where the economic activity of people is not influenced much by seasonal or other factors causing variations over the year.

The "currently active population" (the "labour force") comprises all persons who fulfil the requirements for inclusion among the employed or the unemployed. "Employed" persons comprise all persons above a specified age who, during the reference period, performed some work for pay or profit, in cash or in kind, or were temporarily absent from a job in which they had already worked and to which they had a formal attachment, or from a self-employment activity such as a farm, a business enterprise or a service undertaking. The census documentation and tabulations should clearly describe the time limit chosen as cut-off for considering persons to be "at work". According to the present international recommendations, the notion of "some work" should be interpreted as work for at least one hour during the reference period. The one-hour criterion is an essential feature of the labour force framework embedded in the international definitions of employment and unemployment, and a prerequisite for the consistency of employment statistics with national accounts data on production. Countries concerned about the usefulness of the one-hour criterion for other users of census results should also collect data on "time worked".³

9. Time usually worked

10. Occupation

"Occupation" refers to the type of work done in a job. "Type of work" is described by the main tasks and duties of the work.⁴

³ "Labour force, employment, unemployment and underemployment" were subjects considered at the 13th International Conferences of Labour Statisticians in 1982. It resulted in a resolution, which represents the current ILO recommendations on the subject (source: <http://www.ilo.org/public/english/bureau/stat/techmeet/icls/subjects.htm>).

⁴ An original ILO (LABORSTA) database was compiled from responses to a questionnaire sent to countries that we knew had coded the variable "occupation" using more than 20 occupational groups in their last Population Census or Labour Force (or Household) Survey. ILO only included in the original database those countries that provided adequate data by sex for at least 15 occupational groups. This reduced the number of countries included to 40: six that provided data for only one year, 20 for two years and 14 for three. Geographical representation was unequal: only five African countries, two Latin

11. Industry (branch of economic activity)

"Industry" (branch of economic activity) refers to the kind of production or activity of the establishment or other unit in which the job(s) of the economically active person was located (or, in the case of unemployed, last located). For purposes of international comparability, it is recommended that countries compile the industrial characteristics of active persons according to the latest revision of the International Standard Industrial Classification of All Economic Activities (ISIC) available at the time of the census.

12. Status in employment

"Status in employment" refers to the type of explicit or implicit contract of employment with other persons or organizations, which the person has in his/her job. The basic criteria used to define the groups of the classification are the type of economic risk, an element of which is the strength of the attachment between the person and the job, and the type of authority over establishments and other workers, which the person has or will have in the job. Care should be taken to ensure that an "economically active" person is classified by "status in employment" on the basis of the same job(s) as used for classifying the person by "occupation", "industry" and "sector".

It is recommended that the economically active population be classified by status in employment as follows:

1. "Employees", among whom it may be possible to distinguish "employees with stable contracts" (including "regular employees");
2. "Employers";
3. "Own-account workers";
4. "Contributing family workers";
5. "Members of producers' co-operatives";
6. "Persons not classifiable by status".

It is also recommended to identify separately "Owner-managers of incorporated enterprises," who normally will be classified among "employees," but

American and two Caribbean countries provided data, in contrast with nine Asian and 18 European countries.

The updated ILO database was compiled from responses to a similar questionnaire sent to all countries. With this strategy ILO managed to increase the total coverage to 85 countries and territories of which: 11 in Africa, 15 in the Americas, 4 in the Arab region, 13 from other Asian countries, 18 from Transition Economies, 20 from other European countries and 4 in Oceania, for years around 1970 (58 countries), 1980 (52), 1990 (58) and 2000 (47). Most of the statistics that were provided were based on census results. A first presentation of results from the latest updating of the database can be found in ILO (2003). "Global report under the Follow-up to the ILO Declaration on Fundamental Principles and Rights at Work." Report 1(b) to the 91st Session of the International Labour Conference. June 2003. International Labour Office, Geneva (source: ILO, 2003. "SEGREGAT database Data on employment by sex and detailed occupational groups Documentation May 2003. <http://laborsta.ilo.org/>).

whom one may prefer, for certain descriptive and analytical purposes, to group together with "employers."⁵

13. Place of work

Place of work is the location in which a "currently employed" person performs his/her job, and where a "usually employed" person currently performs or last performed the job. While the information on place of work can be used to develop area profiles in terms of the employed labour force (as opposed to demographic profiles by place of residence), the primary objective is to link the place of work information to the place of residence. Therefore, the place of work should relate to the smallest civil division in which the economic activity is performed in order to establish commuter flows from the place of usual residence to the place of work.

A non-core topic related to place of work is **mode of transport to work**, which relates to the daily journey made. For people making several journeys or using more than one mode of transport, the mode of transport used for the greatest distance in the journey should be indicated.

Another non-core topic, which is important for rural development statistics, is **length (in distance and time) and frequency of journey to work**.

Educational characteristics of persons

14. Educational attainment

For purposes of international comparisons, it is recommended that countries compile their data in accordance with the latest revision available of International Standard Classification of Education (ISCED).

"Educational attainment" refers essentially to the highest level successfully completed in the educational system of the country where the education was received. If relevant, "educational attainment" should take into account all deliberate, systematic and organized communication, designed to bring about learning, even if these were provided outside schools and universities.

⁵ The ILO also provides an indicator of status in employment, which distinguishes between three important and useful categories of the total employed. These are: (a) wage and salaried workers (also known as employees); (b) self-employed workers; and (c) contributing family workers (also known as unpaid family workers). These three groups of workers are presented as percentages of the total employed for both sexes and for males and females separately. Information on the subcategories of the self-employed group – self-employed workers with employees (employers) and self-employed workers without employees (own-account workers) – is not available for all economies but is presented wherever possible.

The indicator on status in employment is available for most developed (industrialized) and transition economies, as well as for many Eastern Asian, Latin American and Caribbean economies. Unfortunately, there are only a few sub-Saharan African economies for which this indicator is available and, where coverage does exist, extensive time series are lacking. Currently, information is also unavailable for some large developing economies, such as China and India. Information for the indicator, at least to some extent, is available for 129 economies (source: <http://www.ilo.org/public/english/employment/strat/kilm/kilm03.htm>).

At least four levels of education should be normally distinguished: primary; secondary - first stage; secondary – second stage; and post secondary. Persons who have received no formal schooling should also be identified.

Ideally, educational statistics should be based on the working age population.⁶

Household and family characteristics of persons

Countries are recommended to use the place of usual residence as the basis of household membership. If only *de jure* information is available (e.g. from registers) on place of residence, i.e. no information is available on usual place of residence, then that information can be used (alone or in combination with other information from other sources) provided that it is judged to reflect the usual residence situation sufficiently accurately.

The concept of “usual residence” is not without ambiguities. Are children away at school or post-secondary school still considered residents of their parents’ locality or are they residents of place where they are schooled? In general, the latter is preferred as that is the residence at the time of the census or survey.

A private household is either:

(a) a one-person household, i.e. a person who lives alone in a separate housing unit or who occupies, as a lodger, a separate room (or rooms) of a housing unit but does not join with any of the other occupants of the housing unit to form part of a multi-person household as defined below; or

(b) a multi-person household, i.e. a group of two or more persons who combine to occupy the whole or part of a housing unit and to provide themselves with food and possibly other essentials for living. Members of the group may pool their incomes to a greater or lesser extent.

This concept of a private household, referred to as the housekeeping unit concept, requires some clarification and explanation on the distinction between

⁶ The ILO covers two relevant indicators respectively, KILM 14 on Educational attainment and illiteracy and KILM 11 on Unemployment by educational attainment. The latter indicator – KILM 11 – focuses on unemployment among workers categorized by their level of educational attainment. Specifically, the indicator is the percentage distribution of an economy's total unemployed according to five levels of schooling - less than one year, less than primary level, primary level, secondary level and tertiary level. Information for the indicator is given for 105 economies, to some extent. Coverage for the indicator is sparse for sub-Saharan Africa, the Middle East and North Africa. Many economies of Asia and the Pacific are missing as well. KILM 14 reflects the levels and distribution of the knowledge and skills base of the labour force. The indicator includes two measures pertaining to educational level of the labour force, and a third measure estimating illiteracy in the adult population. The indicator covers the educational attainment of both women and men in the entire labour force and shows the distribution of the educational attainment of the labour force aged 15 years and above for 104 economies (sources: <http://www.ilo.org/public/english/employment/strat/kilm/kilm14.htm> <http://www.ilo.org/public/english/employment/strat/kilm/kilm11.htm>).

"boarders" and "lodgers". Boarders take meals with the household and generally are allowed to use all the available household facilities. Lodgers, however, are sub-tenants who have hired part of the housing unit for their exclusive use.

Some countries use a different concept of the private household in which the private household is equated with the housing unit. This concept of the household is referred to as the household-dwelling concept, and is defined as the aggregate number of persons occupying a housing unit.

The household-dwelling concept does not provide direct information on the number of housekeeping units sharing housing units. It is recommended that countries applying the household-dwelling concept give an estimate of the total number of housekeeping units in the census report. If the difference between the number of household-dwelling units and the number of housekeeping units is significant, these countries should also endeavour to analyse the occupants of housing units in such a way that they are able to compile the recommended basic tabulations relating to private households on the basis of the housekeeping unit concept as well.

Countries should specify in their census reports whether they used the "housekeeping unit" or the "household-dwelling" concept of a private household.

15. Relationship to reference person

Information should be collected for all persons living in private households on their relationship to the reference member of the household. Data on this topic are needed for use in (i) identifying households and family nuclei; and (ii) compiling tabulations in which households are classified according to characteristics of the reference member. It is left to countries to decide whether the reference member should be:

- (a) the head of household or one of the joint heads;
- (b) the person (or one of the persons) who owns or rents the housing unit or in whose name the housing unit is occupied under some other form of tenure or in whose name part of the housing unit is rented or occupied under some other form of tenure;
- (c) an adult person selected with a view to facilitating the determination of family relationships; or
- (d) a person selected on the basis of other criteria.

It is important that countries describe clearly in the census report the concept of the reference member that has been adopted and the definition that has been used.

16. Tenure status of households

Non-core topics which are important for rural development statistics in this context are durable consumer goods possessed by the household, number of cars available for the use of the household and access to telephones, computers and Internet.

Characteristics of housing units and other living quarters

17. Type of living quarters

18. Type of ownership

19. Location of living quarters**20. Occupancy status****21. Number of occupants****22. Number of rooms****23. Kitchen****24. Water supply system****25. Toilet facilities****26. Bathing facilities****27. Type of heating****Characteristics of buildings containing dwellings****28. Type of building****29. Period of construction**

As for other issues with special importance for rural development statistics, it is recommended that information on place of usual residence should be collected in enough detail to enable tabulations to be made for the smallest geographic or administrative subdivisions required to meet users' needs for information on this topic.

In the UNECE/Eurostat Recommendations, it is suggested that countries which tabulate statistics only for civil divisions should, as a minimum, endeavour to compile data on the total population of each part of a civil division containing a population cluster, or part of a population cluster, of at least 2,000 inhabitants and so provide a basis for making a more clear-cut distinction between urban and rural areas and populations. Specific size classes are also suggested.

The division in urban and rural areas is a derived non-core topic in the UNECE/Eurostat Recommendations. The most appropriate unit of classification for distinguishing urban centres from rural areas is the locality. While it is left to countries to decide whether to use the locality or the smallest civil division as the unit of classification, countries which use the smallest civil division as the unit are encouraged to endeavour to obtain results which correspond as closely as possible with those obtained by countries which use the locality as the unit.

It is suggested that localities or similar units be grouped into the following five categories:

- Less than 2,000 inhabitants
- 2,000 to 9,999 inhabitants
- 10,000 to 99,999 inhabitants
- 100,000 to 999,999 inhabitants
- 1,000,000 or more inhabitants

Countries are also encouraged to develop typologies of localities or similar areas based on additional criteria that could be used to distinguish different types of areas within particular categories of the suggested classification. Some countries may wish to subdivide one or more of the intermediate categories to distinguish market towns, industrial centres, service centres, etc. or to subdivide the large urban agglomerations in order to distinguish various types of central and suburban areas.

Extensions of the classification in these and other ways would enhance its analytical usefulness.

It is also suggested that for purposes of international comparisons, countries define urban centres as localities with a population of 2,000 or more, and rural areas to include localities with a population of less than 2,000 and sparsely populated areas. However, for some purposes, a threshold of 10,000 residents may be more appropriate.

It is the recommendation here that analyses of rural development use regions as units of analysis rather than localities and open country areas. In industrialized countries, many people who now live in the countryside commute to major towns and cities for work. Their livelihoods are essentially urban livelihoods and they generally have access to a full range of urban services. At the same time, many small towns have lost urban functions as services have become more concentrated in urban agglomerations. Regions with only small towns may be considered rural regions even though some of the region's residents live in localities defined as urban on the basis of locality size. Urban regions typically have a large urban center (50,000 or more population), while rural regions do not. However, one may also want to define intermediate regions on the basis of a city size of 10,000 or more and its commuting shed.

To enhance comparability across member countries, the OECD Territorial Indicators Project developed internationally comparable sets of regions based on population density of sub-regions and this is an alternative approach.

Some countries might also wish to consider defining urban areas in other ways (e.g. in terms of administrative boundaries, of built-up areas, of the area for which services such as shops, educational facilities, recreational facilities, employment, etc., are provided, or in terms of functional areas). Whatever approach is taken should be clearly described in the census report.

Agricultural censuses and surveys

In many developing countries, most of the rural population depend directly or indirectly on agriculture for their livelihood. Censuses and surveys on the agriculture sector are major sources of data on the agricultural production units which are mainly household based. As for the population and housing census, agricultural censuses are comprehensive investigations providing structural information on the agriculture sector for small geographical units. They have the same limitations as the population censuses, mainly the length of time separating two censuses (10 years as recommended by FAO) and the lack of coverage of expenditures, consumption and income patterns. During the interval separating two censuses, more frequent agricultural surveys (using the census data as benchmark and sampling frame) are usually conducted to up-date census results.

In many countries, despite the limitations indicated above, censuses of agriculture remain a major source of data on rural populations. FAO, in consultation with other International Organisations prepares every ten years the "World Programme

for agricultural census” which provides up-dated guidelines on methods, concepts and definitions for the organisation of agricultural censuses.

The new “2010 World Programme of Agricultural Census” being prepared by FAO, adopts a modular approach which aims at helping countries to meet the need for a wider range of data from the agricultural censuses, while minimizing the cost of census-taking :

- A core census module, to be conducted on a complete enumeration basis⁷, will provide a limited range of key structural items of importance for national policy-making, making international comparisons, constructing sampling frames, and analysing data at detailed geographic or other levels. The core module is similar to conventional agricultural censuses in the past, but with a much more restricted range of items.
- One or more census supplementary modules, to be conducted on a sample basis at the same time as, or immediately after, the core census module to provide more detailed structural data or data not required at lower administrative levels. The sample for the census supplementary modules will be selected based on sampling frames from the core census module

Below are the items recommended for core and supplementary modules

Items recommended for the core module⁸

- 0001 Identification and location of agricultural holding
- 0002* Sector of agricultural holding
- 0003 Sex of agricultural holder
- 0004 Age of agricultural holder
- 0005 Household size
- 0006 Main purpose of production of the holding
- 0007 Area of holding according to land use types
- 0008 Total area of holding
- 0009 Land tenure types on the holding
- 0010 Whether holding is irrigated
- 0011 Presence of temporary crops on the holding by crop type
- 0012 Presence of permanent crops on the holding by crop type and whether in a compact plantation.
- 0013 Number of animals on the holding by livestock type
- 0014 Presence of aquaculture on the holding
- 0015* Presence of forest and other wooded land on the holding
- 0016 Other economic production activities of the holding’s enterprise

Items for consideration for the supplementary modules

Theme 01 – Land

(Reference group: holdings with land in Item 0008)

Theme 02 – Irrigation and water management

(Reference groups: Items 0201-0205 – holdings with irrigation in Item 0010; Item 0206

⁷ For countries where a complete enumeration is not possible, the core module can be conducted on a large sample basis

⁸ If the agricultural census is conducted in conjunction with the aquacultural census, an additional core item on area of aquaculture by type of site is included

– holdings with temporary crops or permanent crops in Items 0011 and 0012; Item 0207 – holdings with land in Item 0008).

Theme 03 – Crops

(Reference groups: Items 0301-0303 – holdings with temporary crops in Item 0011; Items 0311-0314 – holdings with permanent crops in Item 0012; Items 0321-0327 – holdings with temporary crops or permanent crops in Items 0011 and 0012).

Theme 04 – Livestock

(Reference group: holdings with livestock in Item 0013)

Theme 05 – Agricultural practices

(Reference group: all holdings)

Theme 06 – Agricultural services

(Reference group: holdings in sector “single-holding household” in Item 0002)

Theme 07 – Demographic and social characteristics

(Reference group: holdings in sector “single-holding household” in Item 0002)

Theme 08 – Farm labour

(Reference group: Items 0801-0814 – holdings in sector “single-holding household” in Item 0002; Items 0821-0823 – all holdings)

Theme 09 – Household food security

(Reference group: holdings in sector “single-holding household” in Item 0002)

Theme 10 – Aquaculture

(Reference group: holdings with aquaculture in Item 0014)

Theme 11 – Forestry

(Reference group: holdings with forest and other wooded land in Item 0015)

Theme 12 – Management of the holding

(Reference group: holdings in sector “single-holding household” in Item 0002)

An important feature of the new approach for generating data on rural populations and areas is the provision made for the collection of infrastructure data at the community level in addition to holding level data. Emphasis is also given to integrating the agricultural and population censuses, not only through the use of standard concepts and definitions and sharing field materials, but also coordinating the two data collection activities, adding agriculture-related questions to the population census, and linking of data from the two sources. Also, recommendations are made for countries that wish to broaden the census to all rural households. The new approach will make agricultural censuses a much more relevant source of data on rural areas and population. Since many farm households depend in part on off-farm income or income from family members living elsewhere, this approach will also aid in the assessment of the well-being of farm households.

VI.2.2 Household budget surveys⁹

There are two main objectives with household budget surveys (HBS). The first is to obtain weights, which are used for elementary aggregates in the construction of consumer price indices or in cost of living indices. The international recommended classification of the items of goods and services in the HBS is the United Nations Classification of Individual Consumption According to Purpose (COICOP).¹⁰ The HBS is also used as an input to the building of the national accounts for measuring the household final consumption at an aggregate level.

The second major objective, which is the main interest in the present Handbook, is to provide detailed information of the **living conditions of private households** in defined areas and time. The surveys are meant to give a precise picture of private households total consumption and expenditures broken down in various details as a function of household characteristics such as income, socio economic characteristics, size and composition, degree of urbanization, region, patterns of consumption between different types of households as well as information levels of income, wealth and expenditures. Specific sub-populations such as the elderly, the young, rural populations etc. might be studied. In many countries the distribution of income and consumption is studied from the poverty perspective. Another usage in some countries is for the studying of nutritional patterns of households.

It is very common that special surveys are attached to the regular HBS, e.g. the use of ICT or Internet to take one example.

HBS are hence very multi-purpose in nature and cater for a large number of users and uses. For rural development statistics, it is an invaluable source of data although in some countries it does not permit sufficient breakdown by regions and socio-professional categories, e.g. farmers. However, for improvements of rural development statistics it is clear that HBS are one of the major vehicles. It is here that demands for special rural statistics and its required breakdown must be formulated. If this is done it is, however, in competition with many other statistical areas of interest, and there is of course a limit to how large the HBSs can be made. This implies that the demands from rural statistics must be well formulated, consistent over time, be of interest to many users and be cost effective.

As for the frequency of HBS it varies greatly between countries, from annual survey to surveys every five years or even every 10 years. In many countries the survey period is 12 months, that is, the survey is carried out continuously over the whole year.

As the basic unit for data collection is the household, even if dwellings or addresses happen to be the sampling units, it is important how households are defined

⁹ Main sources: Household Budget Survey in the EU. Methodology and recommendations for harmonisation - 2003. European Commission, Luxembourg, 2003.
Household Surveys in Developing Countries and Transition Countries. Implementation and Analysis. United Nations, New York.

¹⁰ Classification of Expenditure According to Purpose. United Nations, New York, 2000.

(see also chapters V2.1 and ...). Normally the definition of a household is based on persons sharing an accommodation or address and/or sharing expenditures or/and income to various degrees. The family bond is another possible criteria. Even within homogenous groups of countries, e.g. the European Union, different definitions might be applied. Whatever definition reached, it has to be elaborated in details concerning which categories of persons are included and which are excluded. The reference person of the household should also be defined as well as the child-adult setting.

The main characteristics of HBS are briefly summarized in bullet form below:

- ◆ The HBSs are normally confined to the population living in **private households**.
- ◆ The **sample size** varies from a few thousands households to 50,000 in large European countries and to more than 100,000 people in the largest countries.
- ◆ **Probability sampling** is the recommended method, which is also used in most European countries.
- ◆ The most common practise is to use a **two-stage design** for the sampling. First, a stratified sample of area units is selected by using probabilities proportional to size after stratification by geographical areas or other by other variables. The second stage consists of the selection of households within each sample area. In countries with full-cover registers, a single stage sampling is used.
- ◆ As for the **sampling frames** there are a number of different approaches in countries.
- ◆ One approach is to **base the HBSs as a sub-sample on another survey**, such as the labour force surveys, or drawn from a master sample. This means that the HBSs can be linked to other surveys and that a wider range of date can be imputed and estimated. There is also an economic advantage of using such a sub-sample. There are, however, two major drawbacks. First, the response burden increases for the selected households. Second, and which is related to the first drawback, is the effect on the response rates. It is generally necessary to restrict the HBS to those households that previous surveys successfully completed. A second approach is to use **registers**, for those countries that have sufficiently well-covered registers, as the sampling frame. A third approach is to use **area frames** such as a sample of areas from the population census or from a master sample, often constructed from the population census. By sharing between different surveys, the cost of developing and maintaining sampling frames might be reduced. There is also the advantage of operational linkage between different surveys.
- ◆ Normally, the **stratification criteria** are region, socio-professional categories of the reference person and household type or size.

- ◆ One of the major problems with HBS is high **non-response**, which might have serious implications on the reliability of the results. With high rate of non-response it is common to undertake substitution.
- ◆ The recording of household data is usually done by a combination of **interviews** (in the EU usually two per household), **questionnaires** and **diaries**, in which households record all items of daily consumption and expenditures.
- ◆ The objective of HBS is to determine the level and structure of private households' expenditures or consumption or both. **Final consumption expenditure of households** is defined as:

Monetary expenditures intended for consumption + Consumption of own production + benefits in kind + imputed rents for owner occupied housing.

Another, and possibly a better concept, is the **actual final consumption expenditure of the household** which also includes acquisitions from governments and non-profit institutions servicing households (NPISH).

- ◆ As for measuring various types of income, disposable income, net income etc. and their components, see part II of the present Handbook.
- ◆ The list of variables in a HBS usually contain the following items listed in an aggregated level:
 1. Demographic and geographical characteristics;
 2. Characteristics of the household, household members and the reference person;
 3. Education;
 4. Insurance;
 5. Indebtedness;
 6. Economic activity;
 7. Housing;
 8. Possession of consumer durables;
 9. Income;
 10. Individual consumption by purpose.

HBS are, as has been shown above, essential for many important statistical areas: national accounts, CPI, poverty statistics, rural statistics, just to mention a few. They are also, however, rather costly which put focus on cost-effectiveness. To this end there are two approaches to the design of the surveys: larger surveys every five years with say 5,000 households or surveys conducted on a continuous basis with say 1,000 households per year. In the latter model, data are cumulated over several years. For instance, data from three subsequent years might be aggregated and converted into annual results when having adjusted for price changes. The advantage is that the workload is more evenly distributed and that a smaller number of staff can be permanently engaged and thereby increasing their competence level.

An Introduction to Living Standards Measurement Study (LSMS).

The overall objective of the LSMS survey is to measure and study the determinants of living standards in developing countries, especially the living standards of the poor. To accomplish this objective, LSMS surveys must collect data on many aspects of living standards, on the choices that households make, and on the economic and social environment in which household members live. Much of the analysis undertaken using LSMS survey attempt to investigate the determinants of living standards – which requires more sophisticated analytical methods than simple descriptive tables.

LSMS surveys have several characteristics that distinguish them from other surveys. One of the most important is that they use several questionnaires to collect information about many different aspects of household welfare and behaviour. These consist of a household questionnaire, a community questionnaire, a price questionnaire, and, in some cases, a facility questionnaire.

Another characteristic of LSMS surveys is that they typically have nationally representative, but relatively small, samples – usually between 2,000 and 5,000 households. This will yield fairly accurate descriptive statistics for the country as a whole and for large subareas (such as rural and urban areas or a few agroclimatic zones), but usually not for political jurisdictions (such as states or provinces). The surveys' sample sizes are generally adequate for the regression methods often used for policy analysis of LSMS survey data.

Because of the complexity of most LSMS surveys, these surveys have rigorous quality control procedures to ensure that the data they gather are of high quality. This procedure minimize any errors caused by respondent fatigue or the use of proxy respondents. The interviewer make multiple visits to households to find any members who were not home during the interview's earlier visits - which also reduces the need to use proxy respondents.

There is one supervisor for every two or three interviewers. The supervisors must revisit a significant percentage (often 25%) of the sampled households to check on the accuracy of the interviewer's data. They must directly observe some interviews, and they must review each questionnaire in detail. Supervisors performance of these procedures is documented. Data entry and editing are done as soon as each interview is over. A data entry programme carries out a large number of quality checks to detect responses that are out of range or inconsistent with the other data from the questionnaire. Any problem this programme detects can be verified or corrected in a subsequent visit to the household by the interviewer (Grosh and Glewwe, 2000).

Components of a Typical LSMS Survey

One distinguishing characteristics of LSMS surveys is that they are both multi-topic and multi-level: they use several questionnaires to study many different aspects of household welfare and behaviour: The largest LSMS questionnaire is the household questionnaire. The LSMS household questionnaire always collect detailed

information to measure household consumption, which is the best monetary indicator of household welfare. The household questionnaire also collects information on: income; transfer income; and income from wage employment are collected in almost every LSMS survey; and many LSMS surveys also collect data on income from agriculture, household enterprises, and miscellaneous sources.

LSMS household questionnaires always record information on a variety of other dimensions of welfare and the use of social services; housing and related amenities; the level of education of adults. A typical household questionnaire collects more information than this, in order to expand the range of living standards indicators that can be studied and allow researchers to model the choices households make. Some of the information (consumption, housing quality, agricultural production) is collected only at the household level, but much of it (employment, education, health) is collected at the individual level (Grosh and Glewwe, 2000:12).

International Household Survey Network

The demand for household-based economic and social data by national as well as international agencies has been constantly growing. In developing countries, HBS are the most effective way of obtaining most of the information relating to well-being and poverty. However, surveys are complex and expensive undertaking, requiring skilled staff to design and manage the surveys. In many poor countries surveys can only be undertaken with significant outside assistance, as in the case of the LSMS surveys. This means that the surveys are donor driven and often of an *ad hoc* nature – surveys are conducted when donor's funding are available. The most serious problem is that the activities among different donors are often not coordinated. There are many examples of duplicated or conflicting data collection activities, which result in huge waste of funds and put a high burden of scarce personnel resources in the statistical offices in the receiving countries. To this should also be added the need to improve the quality of household survey data.

In this context, the initiative of the World Bank to establish an **International Household Survey Network**, which will ensure better coordinate between international and national donor agencies, on the one hand, and the receiving agencies, on the other. The aim is that this will lead to the collection of more and better data, better use is made of available data, increased frequency, strengthened national capacities in receiving countries and reduced costs. This comes at a price of less freedom of survey design by a particular donor and the prestige that this implies. It ought to be seen that this price is realistic to pay.

Master sampling frames and master samples

Ad hoc scheduling of surveys has now been replaced in many NSOs by long-range plans in which surveys covering different topics are conducted

continuously or at regular intervals. The United Nations National Household Survey Capability Programme has played an important role in this process.¹¹

Sharing of survey personnel and facilities among the surveys will secure effective use of staff and facilities. The development of a master sampling frame (MSF) and a master sample (MS) for the surveys is often an important part of an integrated household survey programme.

For each unit there may be information on urban/rural classification, identification of higher-level units. For each area unit there must also be information on the boundaries of the unit.

The most common type of MSF is one with census enumeration areas as the basic frame units. Usually there is information for each unit that links the unit to higher-level units (administrative subdivisions).

An up-to-date master sampling frame with built-in flexibility has advantages apart from the cost and quality aspects discussed above. It facilitates quick and easy selection of samples for surveys of different kinds and it could meet different requirements on the sample from the surveys.

The MS makes it possible to have overlapping samples in two or more surveys. This permits integration of data at the micro level by linking of household data from the surveys. There is, however, a risk for adverse effects on the quality of survey results when sample units are used several times. Households participating in several rounds of a survey or in several surveys may become reluctant to participate or be less inclined to give accurate responses in the later surveys.

The master sample design always represents a compromise between different requirements on the design from the surveys in the programme. The design chosen for the master sample will usually suit most of the surveys in the survey programme fairly well but none perfectly well.

The most important source of data and materials will usually be the latest population census. That is obvious in the case where the NSO intends to use census enumeration areas as frame units but even if other (administrative) units will be used there is usually a need for population or household data from the census for these units.

The quality of the master sample deteriorates over time. The measures of size used for assigning selection probabilities become out-of-date as population changes take place. That would not be a problem if the population change were a more or less uniform growth in all units in the master sampling frame. However, this is usually not the situation. Population growth and migration occur at varying rates in different

¹¹ See latest report on Household Sample Surveys in Developing and Transition Countries (source: <http://unstats.un.org/unsd/hhsurveys/>).

areas; often there is a low growth, or even decline, in some rural areas and a high growth in some suburban areas in the cities

The master sampling frame is normally completely revised after each population census, usually every ten years.

In some situations, however, it may be possible to use the master sample for surveys concerned with the study of characteristics of economic units, such as household enterprises, own-account business and small-scale agricultural holdings.

In most developing countries a large proportion of the economic establishments in the service, trade and agricultural sectors are closely associated with private households. Those establishments are typically many and small and they are widely spread in the population. There may often be a one-to-one correspondence between such establishments and households, and households rather than the establishments themselves may serve as the ultimate sampling units.

A note on surveys

To be useful for rural development statistics, surveys must have the rural-urban code that is felt most appropriate for analyzing rural development or contain codes with sufficient geographic specificity so that an analyst can create such a code. This is generally less of a problem where the statistics for rural development are analyzed by the statistical agencies that carry out the survey, as confidentiality is not an issue. However, where rural analysts use publicly release versions of data and geographic specificity is limited, survey data may be made largely useless for rural analysis if the relevant geographic codes are not part of the data set. Thus, if the survey defines “rural” as outside of urban localities of 2,000 or more inhabitants, this cannot be used where the analyst is concerned with rural regions—areas outside the immediate influence of major urban centres.

Whatever definition of rural is used, the actual territory classified as rural inevitably changes over time as the most rapidly growing rural areas cross the size and/or density threshold to become urban. Ideally, during transition years, surveys carry both new and old delineations of rural and urban. At the very least, in reporting a statistical series, years when the delineation was changed need to be noted.

National level surveys are constructed to represent the nation as a whole. Typically, sampling units are grouped or stratified on the basis of political and/or geographic characteristics and selected at random from within groups. Within these units, the sampling frame is then updated to take account of new and vacated housing units and a sample is then identified for interviewing. Survey sampling and interviewing costs are lower to the extent that the sampling is clustered, with fewer sampling units used for a given total expected sample size. However, especially where effectiveness of the stratification in capturing geographic variation is weak, reducing the number of sampling units reduces the representativeness of the survey respondents.

This sampling design may result in samples that do not well represent the populations of rural regions or areas, particularly if sampling units are not stratified along rural-urban lines. Moreover, surveys selecting different sampling units over time can yield estimates that vary over time even when no actual change is present.

VI.2.3 Labour force surveys

Labour force surveys are another primary source for rural development statistics.¹² These surveys do not only focus on employment and unemployment characteristics but also on details of hours worked, type of work, economic activity, education and training and income as well as a whole range of demographic variables. Labour force surveys may also contain various types of *ad hoc* modules related to urgent specific topics, such as computer use at home and at work. The units of measurement are both individuals and households. As for the households, it is important also to record variables that describe the composition, in particular since the definition of household varies between countries.

Labour surveys are the source that gives the most detailed information on labour market issues at an aggregate level as well as across economic sectors. However, for cost reasons the sample size is normally not sufficient to provide reliable figures on regional disaggregation or for small industrial or commercial sectors. This is of course a serious drawback when using labour force surveys for rural development statistics. By linking the surveys to censuses, this might be partly overcome by imputing territorial breakdown.

As was mentioned above, the definition of households differs between countries, which has implications on the comparison between countries. In order not to make things even worse, countries are recommended to use the available set of the most recent revisions of international standards such as:

- ◆ International Standard Industrial Classification of all Economic Activities (ISIC), United Nations;
- ◆ International Standard Classification of Occupations (ISCO), ILO;
- ◆ International Standard Classification of Education (ISCED), UNESCO.

In several countries there is a significant proportion of the population that works on their own small farm and **produce only for their own consumption**. Is such a person considered being employed? Normally the definition of employment states that a person is employed if he/she did any work for pay or profit during the reference week for the survey. Work in this context means work for pay, payment in kind or profit during the reference week, even for as little as one hour. If production for own consumption falls within the production boundaries in national accounts, it should be considered as employment and if such production of agriculture goods is

¹² A Labour Force Survey is a standard survey of work-related statistics. The following countries and territories make their Labour Force Surveys available online. Statistics and meta-data for some countries are also available from the ILO [LABORSTA](http://www.ilo.org/dyn/lfsurvey/lfsurvey.home) database (Source: <http://www.ilo.org/dyn/lfsurvey/lfsurvey.home>).

considered to be of importance in relation to the total supply of these goods in the country in question.

Another aspect of employment of importance in all countries but in particular so in developing countries and for issues of rural statistics is the **treatment of seasonal workers**. Here it is clear that during off-season, the season workers should not be considered as being employed in the reference period as they do not continue to receive wages or salaries, even though they might have a contract or an assurance to return to work at a later stage. To capture fluctuations in employment, labor force surveys are often carried out quarterly or even monthly.

Labour force surveys are often the only source of information on underemployment, where people may be working part-time but would work more hours given the opportunity to do so. These surveys may also identify people who are not in the labour force, but would work if the opportunity arose.

Finally, labour force surveys are often the only surveys that ask about people with second jobs (such as farmers who also hold off-farm jobs), employment and unemployment over the past year, or other information that helps to better understand employment conditions.

VI.2.4 Other survey sources

Agriculture censuses and **farm structure surveys** are other useful sources for rural development statistics. While the former is carried out every 5 or every 10 years, the latter is often done biannually or every 5 years. Agriculture censuses have the same advantage as the population censuses in that they provide exhaustive results with detailed territorial breakdown. The variables covered by these sources include the geographical situation of the holding, information about the owner and operator of the holding, type of tenure, land use, live stock, capital equipment and, which is particularly important for rural development statistics, farm labour force.

Some agricultural censuses also include information on off-farm work, which is potentially very useful for understanding linkages between agriculture and rural development. Although these censuses are typically not synchronous with population censuses, it would be extremely useful to use the same language in the agricultural census as in the population census in asking about labour force participation.

In developed countries where agriculture is rather marginal from an employment and sector economic point of view, **business surveys** and **structural business statistics** play a more dominant role as data sources. These surveys, which are carried out frequently, can provide detailed information about type of economic activities, their relative size in monetary terms and in employment terms as well as in their dynamics. However, the shortcomings are that they provide only rather aggregated territorial breakdown, in the best case by major regions. **Statistical business registers**, which constitute the frames for business surveys, do however contain information, which provide detailed territorial breakdown and can from this aspect be a very useful source when analysing the rural business structure. The

variables recorded for enterprises, besides information about location, type of business and owner structures are often limited to employment, wages, and gross output. This limitation could, however, be overcome by creating links between the business surveys and exhaustive processing of the data in business registers.

Health and education, other important variables with respect to rural development analysis, are partly covered in population censuses and HBS and labour force survey. However, for more detailed analysis **specialized surveys in health and education** are often required.

VI.2.5 Administrative registers

Administrative registers are another important source for rural statistics provided and have the advantage that they are readily available and are normally much less expensive to use than surveys. However, they are conditioned on that confidentially laws permit them to be used, that they have wide national coverage and that the administrative arrangements are such that extraction of relevant statistics can be made. Furthermore, the definition of key concepts such as employment, unemployment and type of occupation might not be the same as those that would be used for statistical surveys, as the registers are set up for administrative purposes. In addition, it happens frequently that the arrangements in administrative registers change, leading to discontinuity. From an international point of view, they also have the weakness that they are set up from pure national policy point of views which means that, for instance, registers for social insurance systems vary greatly between countries.

Still, in some countries, notably the Nordic countries, register data are very well developed and can produce excellent statistics as well as providing sampling frames. The type of registers mainly used for rural development statistics are population registers, tax registers (including VAT registers), vehicle registers, social security and insurance registers.

VI.2.6 Vital Statistics Records

Statistics on births and deaths are required to estimate net migration rates for a locality. If the Population Census only asks about residence in the previous year, then it provides a very partial glimpse of migration. Moreover, since the number of people who migrate in a year is relatively small, estimates for small territorial units are likely to be unreliable. With vital statistics records, one may estimate net migration relatively precisely for the entire period between censuses, under the assumption that census population counts at the ends of the time interval were equally reliable. Care must be taken, however, that the residences recorded at births and deaths are the usual place of residence—what is recorded in the Census.

Infant and age-standardized mortality rates are sometimes used as health measures. However, death rates are often so low that frequently one has to average mortality rates over several years to obtain a reliable measure for small localities

VI.2.7 Non-official statistics, e.g. from trade associations

For certain variables of interest for rural development statistics, e.g. tourism, recreation, rural amenities, local business structures and supply of cultural activities, the available sources might not be from official statistics but rather produced by trade associations, local agencies and authorities, societies and clubs (e.g. sport or cultural). If such information is used, much attention must be made to how the statistics was generated, how the definitions were used and if they are compatible with official definitions as well as to how data were collected and processed. In some cases the results of these statistics might be in the interest of the organization that collected the statistics, e.g. number of people engaged in certain cultural or educational activities and when this number might be related to some form of public or private funding. To this should be added that non-official statistics are often neither exhaustive nor based on probability sampling.

It is also a problem when statistics from these sources are combined with official statistics in order to create indicators. Here the rural statistician must be extremely careful and scrutinize compatibility and coverage of the sources. If such indicators are after all used, the statistician should clearly indicate not only sources and methods and individual meta data but also how the indicator should be interpreted and with which precision.

VI.2.8 GIS and geo-coded statistics

An increasing amount of information is being made available through Geographic Information Systems (GIS). These systems attach geographic identifiers to data, allowing them to be mapped. To visualize geographic congruence or other relationships, users can superimpose one map upon another, or create a many layered map. The standard road map, which often contains information about topography, cities and towns, and rivers and lakes as well as roads and highways, is a type of layered map. With GIS one might add a layer of soil type or agricultural products. This would provide information on how agricultural products vary with elevation and distance from a major town, for example.

All data attached to place or area, including data from censuses, satellite imaging, and geological studies, can be included as long as there are geographic identifiers (longitude and latitude) attached. For areas such as political jurisdictions, this means having "shape files," which provide coordinates for the jurisdictional boundaries. This creates an enormously powerful analytical tool, as one can combine ecological data with demographic, economic, and social data relatively easily, once the system is established. It is also possible to extract data files relating to given types of localities. Thus, one could create a table that crossed commune or township elevation with type of agricultural product or income.

Much of the information available through GIS systems is already available for standard geographic units (townships, communes, etc.). Thus, someone may already have created a data set with average elevation that could be attached to a land use file. However, the mapping of the data creates many new opportunities for the development of geographical or ecological indicators that remain largely unexplored.

First, GIS may be used to develop environmental measures for localities. Examples include the proportion of the territory that is lake or pond (or, alternatively, miles of lake shoreline), average elevation, and topographic variation. All of these measures may relate to the attractiveness of the area as a place of residence. Other measures might include miles of autoroute or major road in an area or road density.

Second, GIS may be used for generating information about distances and neighbouring areas. For instance, a given area may not have a major hospital, but with a health facilities data base in GIS and a road system layer, it would be possible to estimate driving times to a regional hospital for populations in territorial units lacking hospitals. Economic well-being of residents in territorial units may relate to job growth not only in the place or area of residence, but also in neighbouring territorial units that are within commuting range.

Third, GIS may be used to define new socioeconomic units. For instance, with data on commuting or shopping patterns, one can use GIS to define labor market or market center service areas and generate statistics for those areas.

Finally, GIS can be used to map information based on political units onto ecological units such as river basins or agricultural zones where these zones are the units of analysis.

The application of environmental measures is not always straightforward. Distances or lengths that are measured can depend on how small the basic unit of measurement is. Variegated shorelines, for instance, vary in estimated length depending on whether one selects to use units that pick up every inlet or use larger units. Much environmental data, such as climate and air or water pollution is collected at data points. Statisticians have developed various methods such as “krieking” to estimate climate and air quality between data collection points, but the estimates are not necessarily precise where the terrain is mountainous or uneven.

To analyze the importance of neighbouring units, some people define neighbouring as those abutting the territorial unit in question while others select on the basis of distance between the population unit in question and the centre of population in other units.

VI.3 Indicators – use and misuse¹³

No general definition of indicator as a special subset of statistical results has been proposed in an official document at international level that is invariably applicable to all areas of official statistics. However, this term is used with increasing frequency; one reason may be that the term “statistics” or “statistical results” is not attracting as much attention from users and media as the term “indicator”. The implicit undertone of the term indicator is that it is more than an “ordinary statistic,”

¹³ Indicators – spotlight or smokescreen? Heinrich Brünger, United Nations Economic Commission for Europe. Paper presented at the 23rd Nordic Statistical Conference, Turku, Finland, 18-21 August 2004.

such as the size of the population or tonnes of steel produced, so that indicators deserve more attention than ordinary statistics.

The **first** and most evident approach to indicators is simply the combination, through a defined algorithm, of two (or more) statistical results (“numbers”) to form a new derived measure. The simplest form is through a numerator and denominator, e.g. as per capita ratios or percentages or other forms of shares. The two elements may be taken from the same statistics (e.g. as for growth rates), or from two different statistics. The added element of the derived measure, as compared to the case when numerator and denominator are looked at separately, is the added degree of comparability; the derived measure eliminates the size effect of the denominator. This is an especially desirable property when comparing across regions or between countries, but it is also important for comparisons over time in economic statistics. With this wide characterisation of an indicator, all comparisons over time based on simple indices (with a base period value as denominator) would be included under this concept. There are more elaborate forms of algorithms to compile derived measures in official statistics, which try to eliminate effects other than size. Examples are age standardisation, constant price aggregates, or seasonal adjustments.

A **second** approach to the term indicator conveys a normative interpretation, with a possibility of establishing rankings or league tables if applied to comparisons of regions, countries, or other appropriate units, for the same period.¹⁴ Indicators are those official statistics that allow a statement of “better”/”worse” (including a quantification of the difference), both in comparisons over time and across space. To be an indicator in the first definition is a prerequisite for normative interpretation; however, not all indicators of the first definition lend themselves to a normative interpretation.

It is important that all results of official statistics, but normative indicators in particular, are accepted by all users, irrespective of their stand on certain policies and priorities, as authoritative measures of the reality. This is the true meaning of impartiality. This principle is also the reason why official statisticians should abstain from policy-prescriptive comments when disseminating normative indicators.

Normative indicators lend themselves to be transformed to rankings or league tables. Opinions differ whether statistical offices, and statistical services of international organizations, should publish rankings, as opposed to a more neutral sequence in tables. If the indicator is sufficiently robust, and if the data sources are of sufficient quality and comparability, there is nothing to be said from the point of view of the fundamental principles against presentations as ranking. Ranking is even used for other than normative indicators, with the purpose to identify countries or regions that are “outliers” in a certain way and where the ranking deserves further explanation and analysis.

¹⁴ The World Competitiveness Yearbook (WCY) ranks and analyzes the ability of nations to create and maintain an environment in which enterprises can compete (Source: <http://www02.imd.ch/wcc/methodology/>).

A **third** type of indicators measures are those indicators that cover only a part of a target concept, which remains unmeasured within the framework of official statistics for reasons that will have to be discussed. Such a type of indicators can be found in areas that lack a common measurement unit such as money in economic statistics, or joule in energy statistics. Many statistical areas lack a common unit of measurement, starting with areas of social statistics such as health, education, quality of work, crime, followed by environment, but also some new phenomena covering aspects of both economic and social statistics such as globalisation, competitiveness and information society. In this approach, the term indicator implies that the indicator stands for something broader than what it actually measures.

A **fourth** approach is the so called **synthetic indicators** that are composite measures of individual indicators (or statistics, depending on the definition of indicators), for the purpose of giving an easy overview of a complex array of indicators or statistics, through using different forms and techniques of “weighting” the individual series.¹⁵ The need for weighting arises from the fact that either the components are not expressed in the same measurement unit, or refer to different items in reality where there is no obvious method of aggregation (such as prices for different commodities). While there had been some discussion about synthetic indicators in meetings of official statistics, the issue of whether, or under what conditions, such indicators are compatible with the framework of official statistics has not been given much attention.

There are no evaluative problems with synthetic indicators that use **empirical weights**, i.e. weights based on official statistics. Price and volume indices are the prime examples of such indicators that are widely used in official statistics. If the single components are themselves produced according to the framework of official statistics, the added step of synthesizing them through empirical weighting derived from official statistics is also within this frame, and adds considerable value for some user categories.

The issue is entirely different with synthetic indicators based on **subjective weights**, be they determined by policy-makers, specific users, or based on expert opinions about importance. Many statisticians argue that such subjective weights are not compatible with the principle of impartiality. Official statisticians should therefore not publish such synthetic indicators under their responsibility as results of official statistics, even when the individual components are fully part of official statistics. Aggregating with no assignation of weights is tantamount to giving equal weights to every component, but this equal weighting remains equally subjective unless it has an empirical or conceptual foundation.

Subjective weighting is especially problematic for international comparisons, because different countries may prefer different weighting schemes, and it is hard to see ways in which an agreement between countries on subjective weights may be

¹⁵ An obvious example is found in the annual **Human Development Reports** (HDR), which was first launched in 1990. Since the first Report, four new composite indices for human development have been developed — the Human Development Index, the Gender-related Development Index, the Gender Empowerment Measure, and the Human Poverty Index (Source: <http://hdr.undp.org/aboutus/>).

reached. In practice, the “experts” from the secretariats of international organizations will impose their own subjective weights.¹⁶

Apart from its use in resource allocation, the main purpose of such synthetic indicators seems to be to allow unequivocal rankings, given that the separate rankings of each component may lead to different “league tables”. In addition, some of these composite indicators are given bombastic labels. However, official statistics and sports are two different pairs of shoes. A synthetic indicator implies the possibility of substitution, i.e. a “bad” ranking in one indicator can be offset by a “good” ranking in another component. In sports, this is indeed possible, but it is questionable to see the substitution aspect between the weakness of a country or region in one aspect, and the strength in another (unless these aspects are part of a common unit of measurement to be found in real life such as in economics). For national policy purposes, it is more important to identify where the strengths and weaknesses of a given country are; averaging them artificially hides this message. In contrast to other forms of aggregates, the simplicity of a single aggregate in this form of indicators is a smokescreen that dilutes the quality of the hard facts of the components by mixing them together with arbitrary subjective valuations.

It is recommended for official statisticians to withstand pressures for endorsing synthetic indicators based on subjective weights, and make the components visible. If they are used anyhow it has to be made absolutely clear to users about the limits of aggregation that are permitted within official statistics.

Synthetic indicators have become very popular outside official statistics, for comparing countries on corruption, competitiveness, etc., with some components of synthetic indicators being subjective indicators. Official statisticians would undermine the principle of impartiality if they felt pressured by this competition to embark on the same approach.

VI.4 Rural measurement issues

Statistics on rural development are problematic in several ways. They are largely based on national censuses and surveys, which, in most industrialized countries at least, were not designed with rural areas in mind. As a consequence, many statistics used routinely at the national level or in urban regions need to be treated with caution when applied to rural areas. And, statistics that might pertain

¹⁶ A good example of this issue is Transparency International’s annual “Global Corruption Reports,” which contains a number of Corruption Surveys and Indices: “The TI Corruption Perceptions Index (CPI), which ranks countries in terms of experts’ perception of corruption; The Global Corruption Barometer is a survey that assesses general public attitudes toward and experience of corruption in dozens of countries around the world; and The TI Bribe Payers Surveys evaluate the supply side of corruption - the propensity of firms from industrialised countries to bribe abroad. (Source: http://www.transparency.org/about_ti/annual_rep/index.html).

largely to rural areas, on land use, for instance, are often not part of national or regional data sets.

One central problem is that, because they are delineated on the basis of density and/or size of place, **rural areas do not remain constant over time and their boundaries are porous**. As noted earlier, rural areas that gain population become at some point urban areas and classified as such. This makes it difficult to track changes over time. Thus, looking at historical population trends using a current delineation of rural-urban is likely to underestimate overall rural population growth and overestimate urban growth. Without a careful presentation of the data, it is easy to conclude that areas have not grown because they are rural, forgetting that areas remain rural because they have not grown.

Rural boundaries not only change over time, they have become increasingly porous as commuting has increased. By definition, rural areas lack large centres. Many residents, particularly in regions proximate to cities, may commute outside of the rural area to work. If regions are defined in part on the basis of a commuting threshold, the problem is minimized, but, even here, until that threshold is reached, there will be some discontinuity between jobs in the region and the jobs held by residents. This affects urban as well as rural statistics, but rural statistics more substantially as the commuters are a larger proportion of the rural work force.

A third issue is that **statistics developed at the national level to reflect livelihood and well-being may be less valid in rural settings than urban settings**. For instance, while censuses and surveys typically ascertain a single occupation. Many rural people may hold more than one job. In some rural areas, for instance, a substantial proportion of people with income from farming may not consider farming as their principal occupation. The number of area farmers identified in an agricultural census may far exceed the number of farmers found in a population census of the same area.

Another, related example is unemployment, a key national and international indicator. This measure works well in urban areas, where the vast majority of people employed are wage and salary workers. In some rural settings, however, particularly agricultural settings, many people are self-employed. In this case, economic hardship is more apt to be marked by underemployment than by unemployment. Farm household members, for example, may have farm work, but have a desire to work off the farm and be able to do so without substantial loss in farm income. Whether they have off-farm work is not likely to be reflected in unemployment statistics.

A final example: rural employment is more likely to be seasonal than urban employment and estimates of labor market conditions can vary considerably depending on when a census or survey is taken. These examples point to the importance of using household budget and related surveys to supplement census data in understanding rural development conditions.

A fourth issue is a **lack of measures that pertain particularly to rural areas development, such as natural amenities**. Measures of urban assets tend to be a by-product of business and other surveys and administrative data. Thus, the presence of

transportation hubs, research universities, and substantial high-end business services sectors have all been linked to urban growth in industrialized countries. By these standards, rural regions are clearly disadvantaged. However, rural areas can have their own advantages: pleasant landscapes and climate, lakes or ocean, mountains and streams, and unique picturesque or historical settings. These assets are important for tourism and second home development, but they can also serve to attract entrepreneurs and others who prefer to reside in natural rather than urban environments. A study of rural U.S. counties found population and employment growth to be more highly correlated with natural amenities than with economic base.¹⁷

The potential importance of these natural assets is included in the concept of farm “multifunctionality,” where farms and farm land recognized to have environmental, recreational, and scenic outputs as well as agricultural product. However, the actual measures of rural amenities—what attracts people to rural areas—have not been developed. One reason is that the measures cannot be derived easily from existing statistical systems. More important is the current lack of any clear basis for ascribing amenity value. When it comes to attractive rural settings, “more” is not always “better.” Landscape preferences research has found, for instance, that people most like varied landscapes (and even then, only up to a point).

The absence of official measures of these rural amenities does not mean they are irrelevant for rural well-being. Families may be willing to “pay” considerably for these amenities, by accepting lower earnings and/or paying a higher proportion of income for housing. The result may be rural anomalies from a strictly economic perspective, such as population shifts from areas of apparently high real earnings to areas of lower real earnings.

The fifth issue is the relatively **small size of rural regions**. Table classifications of income, occupation, and other socioeconomic measures that are used nationally or in urban regions may be inappropriate for rural areas because of issues of data confidentiality. Statistical Disclosure Limitation (SDL) has become a major concern for statistical agencies as surveys and censuses are usually taken with an explicit understanding that personal or establishment information will not be revealed. Reasonable response rates depend on the belief that respondents have that their answers will be kept in confidence.

There are numerous methods of SDL. For table presentation, the most frequent is to suppress data in cells with below a certain number of responses. SDL must be considered as a strategy, however, and not dealt with piecemeal, on a table by table basis, for instance. It is sometimes possible to combine information from different tables, sometimes released in different reports, to ascertain confidential information that is suppressed any single table.

Small size also affects the reliability of statistics estimated on the basis of administrative records. National statistical agencies are increasingly called on to develop local area estimates for various measures when the costs of censuses or major

¹⁷ McGranahan, D. A. (1999) “Natural amenities Drive Rural Population Change.” Economic Research Service, USDA AER-781 (<http://www.ers.usda.gov/Publications/AER781/>)

surveys are prohibitive. For instance, in years outside of censuses, area populations may be estimated on the basis of births, deaths, school enrolments, telephone hook-ups, and other measures. Regional income or domestic product may be estimated from establishment data on employment and wages and other information. Typically, estimates are made first at the national and perhaps regional levels, where there are extensive data, and then allocated downward, based on local statistics. In rural areas, these estimates must be treated cautiously. Estimates are inherently more unreliable where territorial units are small. Moreover, the assumptions and methods that are used in making these estimates may fit urban areas better than rural areas. It is, for instance, typically much more difficult to estimate self-employment income than wage and salary income.

The small size of rural regions can also affect the reliability of estimates from surveys and, since they frequently ask complete sets of questions only for a proportion of respondents, population and other censuses. One way to overcome this problem is by over-sampling in rural regions or at least those rural regions of particular interest, such as agricultural regions or declining regions. Of course, over-sampling in one type of area means under-sampling in another. More precise regional estimates come at the expense of precision in estimates for other national subpopulations such as ethnic groups. Bayesian techniques are available to estimate the sample size necessary to obtain estimates of a given precision.

The sixth issue relates to the interpretation of rural statistics. To the extent possible, well-being outcomes (earnings, employment, health) need to be standardized for the composition of the population. This is most obvious in the case of mortality rates, which are highly related to age and sex, but it extends to other areas as well. Comparisons between rural and urban regions or localities almost inevitably find rural areas have lower incomes than urban areas. This does not necessarily mean that rural areas are somehow lagging, however, or that an explicitly regional policy is called for. As noted elsewhere, rural cost of living is typically lower than urban costs. But, adjusting for cost of living may not be enough. Rural regions also tend to have relatively low levels of educational attainment and to be older than urban populations. Since lower levels of attainment are generally associated with lower incomes at the individual level, lower rural incomes may reflect the educational characteristics of the rural population rather than any drawback to rural residence. By the same token, if income disparities increase or decrease nationally across educational attainment levels, this will be reflected in increasing or decreasing disparities between rural and urban areas. This tendency of divergence “big time” is also found across-countries between industrially advanced countries with a highly educated labour force and least developed countries characterised by a predominantly rural economy.

Unemployment is quite evenly distributed between men and women with a secondary education in most economies, but at the tertiary level of education greater country-level diversity is apparent. In all the economies shown, females with higher education levels are more often unemployed than males of the same education group. On the other hand, with the exception of Peru, men with low education levels have higher unemployment rates than females in all economies.¹⁸

¹⁸ Source: <http://www.ilo.org/public/english/employment/strat/kilm/kilm11.htm>

VI.5 Conclusions and recommendations

The rural development statistician must thus base his/her work on a **multitude of statistical sources**, some of the most important ones are mentioned above. This calls for the statistician to pay very strong attention to the:

- ◆ definitions used in the various sources;
- ◆ differences in coverage;
- ◆ difference in precision;
- ◆ problems of mixing sources in order to create indicators.

All sample surveys are subjected to both sampling errors and non-sampling errors (unwillingness of respondents to provide correct answers, misunderstanding the questions, non-response, mistakes by the interviewers, miscoding etc.). It is therefore important that rural development statisticians indicate as far as possible the reliability of each of the sources used.

Survey data should be exposed to data checking, for instance by using **consistency check** programs. Of particular importance is that the rural development statisticians are aware how the particular survey has treated **missing data**, as these will serious effect on the quality of the data sets. There are several approaches to the problem of missing data. If the household or the person included in the survey cannot be re-interviewed, then the best solution is to impute the data for which there are several methods available. However, if for a particular household or person too many variables contain errors or are missing, that household or person should probably be removed in which case the weighting has to be adjusted. In most cases survey data has to be weighted for other reasons than missing data, both by spatial and temporal weighting. If the rural development statisticians cannot control these issues, which is normally the case, they should at least be informed how the particular survey has treated these issues, so that this properly recorded in the meta data for rural development statistics.

Procedure for imputations of missing data or correction of data must often be used. The rural development statistician must then clearly indicate which procedures have been used and how they might affect the result.

As has been pointed out on several places, rural development statistics in this Handbook focus on the **well-being of the household** in various categories of territorial areas. As the household is the main statistical unit for which production, income and consumption is measured, the underlying definitions must be clearly indicated and tested for consistency as well as compared with other countries. In this Handbook it has been said that as in many cases there are large variations in definition and coverage of variables between countries, **it may not be advisable to benchmark levels but rather to changes in levels.**

Within a country there might be quite significant differences in income between regions. In these cases there are normally also significant differences in

price levels, in particular for housing, food and certain services. This calls for the use as far as it possible of **income measures expressed in purchasing power**.

Comparison of the relative purchasing power between regions can be done in several ways using:

- ◆ **national weights**, which enables comparisons for a uniform basket of goods between regions and the national average as well as between regions;
- ◆ **regional weights**, by pricing the cost of regional representative baskets. This enable comparisons to be made for a representative basket of goods for each individual region, compared with what that basket would have cost at national average prices; and
- ◆ **a geometric mean** of national and regional weights.¹⁹

Data for the United Kingdom showed that for the North East region the national weights fell from 95.3 to 91.5 between 2000 and 2003 while for London they increased from 106.8 to 107.6.²⁰ This concerns national averages (=100). For individual items there are of course even larger differences. The regional weight for housing, for instance, was 69.2 in the North East while it reached 120.0 in the South East.

Rural development statistics must be based on a multitude of statistical sources, which are not always compatible. This Handbook strongly recommends that all published rural development statistics be published together with detailed analysis of sources and methods used, detailed meta data for the individual data cells as well as a readers' guide, advising how the results should be interpreted and with what precision.

For international benchmarking the above should be recorded for each country and if incompatibilities exist this must be clearly indicated and to what extent data indicator levels can in fact be compared. If they cannot, only changes in levels of indicators should be recorded.

Because of the multitude of indicators it might be tempting to construct a rural composite indices for international benchmarking. This Handbook strongly advises against such a practise.

GROSH, M. & GLEWWE, P. (Eds.) (2000) *Designing Household Survey. Questionnaires for Developing Countries. Lessons From 15 years of the Living Standards Measurement Study.*, Washington.

¹⁹ Relative Regional Consumer Price Levels in 2003. Adrian Ball and David Fenwick, Office of National Statistics, United Kingdom.

²⁰ Ibid.