

VI DATA SOURCES

VI.1 Introduction

Rural development statistics are concerned with the measurement of a multitude of heterogeneous variables, in the areas of demography, geography and territorial descriptions, environment and, above all, the whole spectrum 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, for instance, might very well be timely and have detailed breakdown of types of employment and education but only on a national level or, in the best case, major regions. They may not be available at smaller administrative levels such as communes, which are the basic units in rural development statistics.

To be useful for rural development statistics, surveys and other sources of primary data must contain the rural-urban code that is most appropriate for analysing 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 analysed by the statistical agencies that carry out the survey, as confidentiality is not an issue. However, where rural analysts use publicly released 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 which are 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 results that vary over time even when no actual change is present.

A related problem is that rural development statistics often have to **combine data from different sources**, sources that may differ in definitions, coverage, precision, measurement period and periodicity. Statistics on employment may differ depending on the source (household surveys, establishment (business) surveys, or administrative records). The problem is compounded when data from two or more sources are combined to form an indicator. Even combining national accounts data with employment and foreign trade data is problematic despite the close complementary nature of these statistics. When constructing rural

development indicators where 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.

A good example of mixing data sources is provided by the International Labour Office's (ILO's) new agricultural labour productivity indicator (Key Indicator of Labour Market (KILM) 18e) which is largely derived from the Groningen Growth and Development Centre (GGDC) Database at the University of Groningen (the Netherlands).¹

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 quite similar 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 combination of one or several of these characteristics. There may also be big differences in the child-adult definition and how much weight is given in the household aggregate. (See Chapter IX of this Handbook for a further discussion of these issues.)

Also when it concerns the variables for which the household is characterized there is a great span in the definition among countries. This indicates that care should be taken when comparing indicator 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. While, for instance, price statisticians might be confident in presenting estimates of price changes for country comparisons, they are not that keen on comparing levels because like is not always compared to like.

The rest of this chapter reviews the main data sources used, singly or in combination, to generate rural statistics. The discussion attempts to cover both OECD members and developing countries.

VI.2 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 to 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 with respect to

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 purchasing power parity, etc., can be downloaded from the Internet site of the GGDC. Website: <http://www.eco.rug.nl/ggdc/>

expenditures, consumption and income patterns. As will be seen, however, this can be attenuated by conducting household budget surveys linked as sub-samples to censuses.

Box VI.1

An example of mixing data sources

ILO's new agricultural labour productivity indicator (Key Indicator of Labour Market (KILM) 18e

For the 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 the OECD: *National Accounts*, Volumes I and II (annual issues). Employment estimates for the aggregate economy are mostly taken from OECD: *Labour Force Statistics* (annual issues), and for the individual sectors – as far as available – from the OECD: *National Accounts*, Volume II (annual issues) and the 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, benchmark estimates of annual hours worked for a significant number of non-OECD economies have been developed.

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 purchasing power parity (PPP) taken from a 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 (KILM, 2004; van Ark and Monnikhof, 2000).

There are several international sets of recommendations which provide guidance and assistance to countries in planning the content of their census. These recommendations also facilitate and improve international comparability through the harmonization of data, definitions and classification of topics. To this end readers may wish to consult the following:

- **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, for example, 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. (See also Chapter IX of this Handbook.)

The UNECE/Eurostat Recommendations contain the following 29 core topics. Those that are 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. Some of these are also briefly mentioned.

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 the 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."²

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.³

² "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>).

³ An original ILO (LABORSTA) database was compiled from responses to a questionnaire sent to countries that were known to have coded the variable "occupation" using more than 20 occupational groups in their last Population Census or Labour Force (or Household) Survey. The 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 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 the ILO managed to increase the total coverage to 85 countries and territories of which there were: 11 in Africa, 15 in the Americas, 4 in the Arab region, 13 from other Asian countries, 18 from Transition Economies, 20 from other

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' cooperatives”;
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 whom one may prefer, for certain descriptive and analytical purposes, to group together with “employers.”⁴

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 (ILO, 2003a; 2003b).

⁴ 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>).

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 of the **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.

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.⁵

⁵ The ILO covers two relevant indicators: KILM 11 on unemployment by educational attainment and KILM 14 on Educational attainment and illiteracy.

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. Coverage for the indicator is sparse for sub-Saharan Africa, the Middle East and North Africa. Many economies of Asia and the Pacific are also missing.

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>).

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 the 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 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 multiperson household as defined below; or
- A multiperson 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 on the distinction between “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. Countries should specify in their census reports whether they used the “housekeeping unit” or the “household-dwelling” concept of a private household. (Chapter IX of this Handbook contains a detailed discussion of different household concepts in the context of measuring the incomes of agricultural households).

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:

- The head of household or one of the joint heads;
- 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;
- An adult person selected with a view to facilitating the determination of family relationships; or
- 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. (Again, Chapter IX of this Handbook also discusses this issue in the context of statistics on the incomes of agricultural households).

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**

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.

This will 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**.⁶ 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 are encouraged to obtain results which correspond as closely as possible with those obtained by countries which use the locality as the unit of classification.

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.

This Handbook recognizes as good practice that statistics on rural development should use regions as the geographical 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 centre (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 (for example, in terms of administrative boundaries, of built-up areas, of the area for which services such as shops,

⁶ This does not necessarily mean that locality is the unit for analysis but rather the unit for classification aspects. Normally, one should use localities for analysis of local issues and regions for analysis of regional issues.

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.

VI.3 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. Similar to population and housing censuses, 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 the 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 the benchmark and sampling frame) are usually conducted to update census results.

In many countries, despite the limitations indicated above, censuses of agriculture remain a major source of data on rural populations. The FAO, in consultation with other international organizations prepares the “ World Programme for Agricultural Census” every 10 years. This provides updated guidelines on methods, concepts and definitions for the organization of agricultural censuses.

The new “2010 World Programme of Agricultural Census” being prepared by the 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 basis,⁷ will provide a limited range of key structural items of importance for national policymaking, 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.

Box VI.2 contains the items recommended for core and supplementary modules.

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 the linking of data from the two sources. Also, recommendations are made for countries that wish to broaden the census to all rural households.

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

Box VI.2
FAO: World Census of Agriculture

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
0003	Sex of household members (paragraphs).
0004	Age of household members (paragraphs).
0005	Household size
0006	Main purpose of production of the holding
0007	Number of land parcels (paragraphs).
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 – 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*)

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.

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 (see Part II of this Handbook).

VI.4 Household budget surveys

The main sources for this section are: “Household Budget Survey in the EU. Methodology and recommendations for harmonization – 2003 (European Commission, Luxembourg, 2003)” and “Household Surveys in Developing Countries and Transition Countries, Implementation and Analysis (United Nations, New York).” There are two main objectives for a household budget survey (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 HBSs is the United Nations Classification of Individual Consumption According to Purpose (COICOP) (UN, 2003). 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 of HBSs, which is the main interest in the present Handbook, is to provide detailed information on 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, socioeconomic characteristics, size and composition, degree of urbanization, region, patterns of consumption between different types of household as well as information on levels of income, wealth and expenditures. Specific subpopulations 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.

HBSs are very multi-purpose in nature and cater to a large number of users and uses. For rural development statistics, they are an invaluable source of data although in some countries they do not permit sufficient breakdown by regions and socio-professional categories (for example, farmers – see Chapter XIII of this Handbook). However, it is clear that HBSs have great potential for the improvement of rural development statistics. It is here that demands for special rural statistics and the required breakdown must be formulated. If this is done it is, however, in competition with many other statistical areas of interest, and there is 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.

The frequency of HBSs varies greatly between countries, from annual surveys 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, the way households are defined is important. Normally the definition of a household is based on persons sharing an accommodation or address and/or sharing expenditures and/or income to various degrees. The family bond is another possible criterion. Even within homogenous groups of countries (for example, the European Union), different definitions might be applied. Whatever definition is reached, it has

to be made clear 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 HBSs are briefly summarized in bullet form below:

- They 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 sampling. First, a stratified sample of area units is selected by using probabilities proportional to size after stratification by geographical areas or 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.
- There are a number of different approaches to **sampling frames**.
 - One approach is to **base the HBS as a sub-sample on another survey**, such as a labour force survey, or drawn from a master sample. This means that the HBSs can be linked to other surveys and that a wider range of data 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 this 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 successfully completed previous surveys. For those countries that have sufficiently well-covered registers, a second approach is to use **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 linkages 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 HBSs is a high **non-response rate**, which might have serious implications on the reliability of the results. With a 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 objectives of HBSs are 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 better concept, is the **actual final consumption expenditure of the household** which also includes acquisitions from governments and non-profit institutions servicing households (NPISH).

- For the measurement of various types of income (disposable income, net income etc.) and their components, see part II of this Handbook.

The list of variables in a HBS usually contains the following items at 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.

HBSs are, as has been shown above, essential for many important statistical areas: national accounts, CPI, poverty statistics and rural statistics, just to mention a few. They are also, however, rather costly which puts the 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 accumulated over several years. For instance, data from three subsequent years might be aggregated and converted into annual results after having been adjusted for price changes. The advantage is that the workload is more evenly distributed and a smaller number of staff can be permanently engaged, which will increase their competence level.

VI.4.1 Living Standards Measurement Study surveys - an introduction

The overall objective of Living Standard Measurement Study (LSMS) surveys is the measurement and study of 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 surveys attempts to investigate the determinants of living standards and this 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 are nationally representative, but use 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 their complexity, LSMS surveys have rigorous quality control procedures to ensure that the data they gather are of high quality. These procedures minimize any errors caused by respondent fatigue or the use of proxy respondents. The interviewer makes multiple visits to households to find any members who were not home during the interviewer's earlier visits. This 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. The supervisor's assessment 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 characteristic of LSMS surveys is that they are multitopic and multilevel: 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 collects detailed information to measure household consumption, which is the best monetary indicator of household welfare. The household questionnaire also collects information on income and transfer income, while data on income from wage employment are collected in almost every LSMS survey. Further, many LSMS surveys also collect data on income from agriculture, household enterprises, and miscellaneous sources.

LSMS household questionnaires also record information on a variety of other dimensions of welfare and the use of social services (for example, housing and related amenities and 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 (for example, consumption, housing quality and agricultural production) is collected only at the household level, but much of it (for example, employment, education and health) is collected at the individual level (Grosh and Glewwe, 2000).

VI.4.2 International Household Survey Network

The demand for household-based economic and social data by national as well as international agencies is constantly growing. In developing countries, HBSs 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 them. 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, with surveys only conducted when donor funding is 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 a huge waste of funds and put a high burden on the 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, an initiative of the World Bank is to establish an **International Household Survey Network**. This is designed to ensure better coordination between international and national donor agencies, on the one hand, and the receiving agencies, on the other. The aim is to facilitate the collection of more and better data, make better use of available data, increase the frequency of data collection, strengthen national

capacities in receiving countries and reduce costs. This comes at the price of less freedom of survey design by a particular donor and the loss of the associated prestige. It should be acknowledged that this is a reasonable restraint on the donor.

VI.4.3 Master sampling frames and master samples

Ad hoc scheduling of surveys has now been replaced in many National Survey Offices (NSOs) with 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.⁹

Sharing of survey personnel and facilities among the surveys will make 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 and 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 MSF 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 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 MS design always represents a compromise between different requirements on the design from the surveys in the programme. The design chosen for the MS 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. This is obvious in the case where the NSO intends to use census enumeration areas as frame units but even if other (administrative) units are to be used there is usually a need for population or household data from the census for these units.

The quality of the MS deteriorates over time. The measures of size used for assigning selection probabilities become out-of-date as population changes take place. This 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 areas;

⁹ See latest report on Household Sample Surveys in Developing and Transition Countries (source: <http://unstats.un.org/unsd/hhsurveys/>).

often there is a low growth, or even decline, in some rural areas and a high growth in some suburban areas in the cities. For this reason the MSF is normally completely revised after each population census, usually every ten years. In some situations, however, it may be possible to use the MS 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. These establishments are typically many and small and widely dispersed throughout 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.

VI.4.4 Suggested integrated programme of household surveys¹⁰

The value of household surveys covering family status, access to services, economic activity, production practices, expenditure and social activity would be greatly enhanced if they were followed up by the routine use of panel surveys using sub-samples, in order to track performance. This approach would ensure comprehensiveness and consistency and avoid the use of excessive amounts of ad hoc, largely anecdotal data. It requires a long-term financial commitment by countries and international development partners. Some initiatives are under way which could help in this effort. For example, the Food Insecurity and Vulnerability Information Mapping System (FIVIMS), a UN inter-agency action, is working with countries to facilitate standard data collection for food security indicators to monitor follow-up to the World Food Summit in 1996.¹¹

One approach (see Vogel, 2004b) could be an important first step. Vogel proposes a framework and overall design of an integrated annual rural household survey programme, which would integrate a country's core statistical needs with those needed to monitor progress towards meeting the Millennium Development Goals (MDGs). The framework recognizes some of the MDG indicators and incorporates them as part of the core items and issues covered in the survey design. The survey design recognizes that there are core data items at the household level that need to be monitored every year but it also recognizes that there is no need to monitor other items every year because situations do not change that rapidly. There are two reasons for proposing an annual survey framework. One is to build a database of household data that can be used for analytical purposes for current and future policy and investment decisions, and especially to forecast the consequences of any proposed actions. A primary goal is to provide a survey framework that provides the data required to monitor progress towards meeting the MDGs, and more importantly, the inter relationships between them. Another goal is to foster statistical capacity building. When countries only do periodic household surveys, they receive training and technical assistance for that particular instance. However, when the survey is completed the sponsor or donor goes elsewhere, leaving the country without the capacity to continue.

Figure VI.1 provides a brief description of the proposed survey framework. It should be noted that the proposal suggests having a national level rural household survey to provide a cross-section of information over time. Further, the sample is divided into replicates with five being used each year. By the fifth year, each replicate will have been in the sample for five consecutive years.

¹⁰ This section includes an excerpt from the paper "Integrating Rural Household Survey" presented by F. Vogel at the 2004 MEXSAI Conference in Cancun (Vogel, 2004a).

¹¹ <http://www.fivims.net/index.jsp?lang=en>

Figure VI.1
Description of distribution of sample replicates by year and detailed questionnaire to be used

Replicate Year	Rep 1	Rep 2	Rep 3	Rep 4	Rep 5	Rep 6	Rep 7	Rep 8	Rep 9	Rep 10	Rep 11	Rep 12
1	A	A	A	A	A							
2		B	B	B	B	B						
3			C	C	C	C	C					
4				D	D	D	D	D				
5					A	A	A	A	A			
6						B	B	B	B	B		
7							C	C	C	C	C	
8								D	D	D	D	D
9									A	A	A	A
10										B	B	B
11											C	C
12												D

Detailed Questionnaires

- A. MDG 1. Poverty, Hunger, Employment, income
- B. MDG 2 & 3 Education
- C. MDG 4,5, & 6 Health
- D. MDG 7 Environment

Every replicate receives same core questionnaire every year

- A. (MDG 1) Indicators of progress towards eradicating poverty and hunger. This would also have information about employment and wage rates included in the October Inquiry. This would provide the primary data needed to enable the calculation of Purchasing Power Parities for the poor in conjunction with the International Comparison Program.
- B. (MDGs 2 and 3) Indicators of progress towards achieving universal primary education and gender equality for women.
- C. (MDGs 4, 5, and 6) Indicators of progress made to reduce child mortality, improve maternal health and combat diseases.
- D. (MDGs 7 and 8) Indicators at the household level to measure the consequences of improving the environment and improving the competitiveness of markets by removing distortions to trade.

Every sample household in every replicate will receive the same core questionnaire that will remain essentially the same over time. The core questionnaire will obtain information needed on an annual basis, have considerable year-to-year volatility, be important in monitoring food security issues, and provide an overview of progress towards meeting the MDG's. The design will provide longitudinal analysis of the core data so that short run evaluations of the effects of policy or investment decisions can be appraised.

The key to the design is that each year each household is also surveyed by a detailed questionnaire that rotates by subject matter each year. The variables to be measured on a rotating basis are those for which change would be difficult to monitor on an annual basis. During the period a household is in the sample, it will be queried by each of the detailed questionnaires at least once with one fifth of the households surveyed twice for one of the detailed questionnaires.

At the end of the fifth year a household in the sample will have been surveyed annually for the core items, once for each detailed questionnaire, and twice for one of the detailed versions. In this way the survey will provide a matching sample comparison over time.

VI.5 Labour force surveys

Labour force surveys are another important 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 specific urgent topics, such as computer use at home and at work. The units of measurement are both individuals and households. With households, it is important to record variables that describe their composition since the definition of household varies between countries.

Labour force 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 a serious drawback when using labour force surveys for rural development statistics. The problem may be partly overcome by linking the surveys to censuses and imputing territorial breakdown.

As previously mentioned, the definition of household differs between countries, which has implications for international comparisons. 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, a significant proportion of the population work on their own small farm and **produce only for their own consumption**. Is such a person considered to be employed? Normally the

¹² A Labour Force Survey is a standard survey of work-related statistics. Statistics and metadata for some countries are available from the ILO [LABORSTA](http://www.ilo.org/dyn/lfsurvey/lfsurvey.home) database (Source: <http://www.ilo.org/dyn/lfsurvey/lfsurvey.home>).

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 one's own consumption falls within the production boundaries in national accounts (i.e. it is considered important in relation to the total supply of these goods), it should be considered as employment.

The **treatment of seasonal workers** is another aspect of employment that is important in all countries, but particularly for developing countries and for rural statistics. It is clear that during the off-season the seasonal 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 seasonal fluctuations in employment, labour force surveys are often carried out quarterly or even monthly.

Labour force surveys are often the only source of information on underemployment, where people work 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, particularly in rural areas.

VI.6 Other survey sources

In developed countries where agriculture is rather marginal from an employment and economic sector perspective, **business surveys** and **structural business statistics** play a more dominant role as data sources than do farm surveys (dealt with above). These surveys, which are carried out frequently, can provide detailed information about the type of economic activities, their relative size in monetary and employment terms as well as their dynamics. However, the shortcomings are that they provide only a rather aggregated territorial breakdown, in the best case by major regions. **Statistical business registers**, which constitute the frames for business surveys, do contain information with a detailed territorial breakdown and these can be a very useful source for analysing 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, which are also important variables with respect to rural development analysis, are partly covered in population censuses, HBSs and labour force surveys. However, for more detailed analysis **specialized surveys in health and education** are often required.

VI.7 Administrative registers

Administrative registers are another important source for rural statistics and have the advantage that they are readily available and are normally much less expensive to use than surveys. Despite the fact that they are subject to confidentiality laws with regard to how they can be used, they have wide national coverage and the administrative arrangements are such that relevant statistics can be extracted. It must be remembered that 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, the arrangements in administrative registers can change frequently,

leading to discontinuity. Finally, they also have the weakness that they have a purely national policy perspective. This means that, for instance, registers for social insurance systems vary greatly between countries.

However, 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 registers and insurance registers.

VI.7.1 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 it provides only 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 for births and deaths are the usual places of residence and match 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.8 Non-official statistics, e.g. from trade associations

For certain variables of interest for rural development statistics (for example, 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, or societies and clubs (e.g. sport or cultural). If such information is used, much attention must be given to how the statistics were generated, what definitions were used, whether they are compatible with official definitions and how the data were collected and processed. In addition, the results might serve the interests of the collecting organization and may not be totally objective. An example of this would be data on the numbers engaged in certain cultural and educational activities where the results are tied to the provision of public or private funding for those activities. To this should be added that non-official statistics are often neither exhaustive nor based on probability sampling.

It is also problematical 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 used, the statistician should clearly indicate not only sources, methods and individual metadata but also how the indicator should be interpreted and with what precision.

VI.9 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 ecological data can be combined 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, for instance, 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 occupied by a 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 roads in an area or road density.

Second, GIS may be used for generating information about distances and neighbouring areas. For instance, with a health facilities database 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 socio-economic units. For instance, with data on commuting or shopping patterns, one can use GIS to define labour market or market centre 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 form the unit 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 a measurement unit that pick up every inlet or a measurement unit that will not. 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 these estimates are not necessarily precise where the terrain is mountainous or uneven.

VI.10 Conclusions and recommendations

This chapter has demonstrated that the rural development statistician must base his/her work on a **multitude of statistical sources**. 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 subject 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 rural development statisticians be aware of how the particular survey has treated **missing data**, as these will seriously affect the quality of the data sets. There are several possible 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. There are several methods available to achieve this. 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 this case the weighting has to be adjusted.

When procedures for the imputation of missing data, or correction of data, are used, the rural development statistician must clearly indicate which procedures have been used and how they might affect the results. If the rural development statisticians cannot control these issues (which is often the case), they should at least be informed how the particular survey has treated them, so that this is properly recorded in the metadata.

Though more an issue of publication than a characteristics of data sources, 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 in any single table.

As has been pointed out several times, 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 benchmark 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 to 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 (Ball and Fenwick, 2003):

- **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.¹³

Rural development statistics must be based on a multitude of statistical sources, which are not always compatible. This Handbook strongly endorses, as good practice, the publication together with the statistics of a detailed analysis of the sources and methods used, adequate metadata for the individual data cells, and 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. If incompatibilities exist, these must be clearly indicated, together with a guide 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 of indices for international benchmarking. This Handbook advises against such a practise.

¹³ 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 (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.

References

- Ball, A. and D. Fenwick (2003). “Relative Regional Consumer Price Levels in 2003”. Office of National Statistics, United Kingdom.
- Grosh, M. & Glewwe, P. (Eds.) (2000). “Designing Household Survey. Questionnaires for Developing Countries. Lessons From 15 years of the Living Standards Measurement Study.” Washington.
- International Labour Office (ILO) (2003a). “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
- International Labour Office (ILO) (2003b). “SEGREGAT database Data on employment by sex and detailed occupational groups”. Documentation May 2003.
<http://laborsta.ilo.org/>
- International Labour Office (ILO). [LABORSTA](http://www.ilo.org/dyn/lfsurvey/lfsurvey.home) database. <http://www.ilo.org/dyn/lfsurvey/lfsurvey.home>.
- 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
- United Nations (UN) and the Statistical Office of the European Communities (1998). “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 New York and Geneva. Statistical Standards and Studies – No.49.
- United Nations (UN). “Principles and Recommendations for Population and Housing Censuses”. Statistical Papers, Series M, No.67/ Rev.1.
- United Nations (UN) (2003). “Household Budget Survey in the EU. Methodology and recommendations for harmonisation - 2003”. European Commission, Luxembourg, 2.
- United Nations (UN). “Household Surveys in Developing Countries and Transition Countries, Implementation and Analysis”. New York.
- United Nations (UN) (2000). “Classification of Expenditure According to Purpose”. New York.
- van Ark, B. and E. Monnikhof (2000). “Productivity and unit labour cost comparisons: A data base”. Employment Paper 2000/5. ILO, Geneva.
<http://www.ilo.org/public/english/employment/strat/publ/ep00-5.htm>.
- Vogel, F. (2004a). “Integrating Rural Household Survey”. Paper presented at the 2004 MEXSAI Conference, Cancun.
<http://www.nass.usda.gov/mexsai/papersabstracts.htm>
- Vogel, F. (2004b). “Expert Panel on Statistical Capacity Building. Topic 3: Rural Poverty and Hunger (MDG1)”. Paper presented at the 2004 MEXSAI Conference, Cancun.
<http://www.nass.usda.gov/mexsai/papersabstracts.htm>