

**UNITED NATIONS**  
**Department of Economic and Social Affairs**  
**Statistics Division**

**ESA/STAT/AC.135/1**  
**October 2007**  
**English only**

**INTERNATIONAL LABOUR OFFICE**  
**Bureau of Statistics**

---

**United Nations Expert Group Meeting on**  
**Measuring the Economically Active Population in Censuses**  
**5-9 November 2007, New York**

# **HANDBOOK ON THE COLLECTION OF ECONOMIC CHARACTERISTICS IN POPULATION CENSUSES**

**Draft**



Statistics Division  
Department of Economic and Social Affairs  
United Nations

Bureau of Statistics  
International Labour Office

# **COLLECTION OF ECONOMIC CHARACTERISTICS IN POPULATION CENSUSES**

**Handbook**





ST/ESA/STAT/119  
Draft September 2007  
English Only

Statistics Division  
Department of Economic and Social Affairs  
United Nations

Bureau of Statistics  
International Labour Office

# **COLLECTION OF ECONOMIC CHARACTERISTICS IN POPULATION CENSUSES**

**Handbook**

## NOTE

The designations used and the presentation of material in this report do not imply the expression of any opinion whatsoever on the part of either the Secretariat of the United Nations or of the International Labour Office concerning the legal status of any country, territory, city or area, or its authorities, or concerning the delimitation of its frontiers or boundaries.

The term “country” as used in this report also refers, as appropriate, to territories or areas.

The designations “developed” and “developing” regions/countries are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process.

Symbols of the United Nations documents are composed of capital letters combined with figures. Mention of such a symbol indicates a reference to a United Nations document.

This document has not been officially edited.





## PREFACE

The United Nations, as part of its World Population and Housing Census Programme, has provided technical guidelines to its member states primarily through the preparation of (i) *Principles and Recommendations for Population and Housing Censuses*;<sup>1</sup> (ii) *Handbooks*<sup>2</sup> on general census operations, as well as on the collection of data on specific topics, such as economic activity; and (iii) supporting technical materials.

The first publication *Principles and Recommendations for Population and Housing Censuses*,<sup>3</sup> was issued for the 1980 round of censuses (1975 - 1984). It has been supplemented by further publications<sup>4</sup> for the 1990, 2000 and 2010 rounds of censuses that present changes in relevant international standard concepts and classifications, as well as accumulated experiences from national census exercises. For the 1990 round of censuses, a *Handbook on Economic Activity Status* was also issued, presenting the content of census questionnaires and national practices in the collection of information on economic characteristics of the population during the 1970 and 1980 round of censuses.<sup>5</sup> The review documented which countries had included specific items in their population censuses. In 2002, a report entitled *Collection of economic characteristics in population censuses: Technical report* was issued in draft form, and was circulated for comments and suggestions for improvement with a view to the preparation of a Handbook in 2004. This current document has been produced as a result.

Given the broad focus of the *Principles and Recommendations*, it was not possible or practical to include details that would guide countries in implementation of these concepts and standards operational in censuses. This *Handbook* is based on the *Technical Report* issued in 2002 and provides suggestions on the implementation of

---

<sup>1</sup> United Nations (1980). *Principles and Recommendations for Population and Housing Censuses*. Statistical Papers No. 67, Sales No. E.80.XVII.8; United Nations (1990). *Supplementary Principles and Recommendations for Population and Housing Censuses*, Statistical Papers N. 67/Add.1, Sales No. E.90.XVII.9. United Nations (1998): *Principles and Recommendations for Population and Housing Censuses. Revision 1*. Sales No. E.98. XVII.8. United Nations (2007): *Principles and Recommendations for Population and Housing Censuses. Revision 2*. Sales No. E.07.XVII.8.

<sup>2</sup> Examples of these Handbooks include:(a) Part I: *Planning, Organization and Administration of Population and Housing Censuses*, Sales No. E.92.XVII.8; (b) Part II: *Demographic and Social Characteristics*, Sales No. E.91.XVII.9; and (c) Part IV: *Economic Activity Status*, Sales No. E.96.XVII.13.

<sup>3</sup> Statistical Papers No. 67, United Nations publications, Sales No. E.80.XVII.8.

<sup>4</sup> Statistical Papers N. 67/Add.1, United Nations publications, Sales No. E.90.XVII.9, United Nations (1998): *Principles and Recommendations for Population and Housing Censuses. Revision 1*. Sales No. E.98. XVII.8 and United Nations (2007): *Principles and Recommendations for Population and Housing Censuses. Revision 2*. Sales No. E.07.XVII.8.

<sup>5</sup> United Nations publication, Sales No. E.96.XVII.13. The International Labour Office in 1990, 1996 and 2003 published more detailed descriptions in its publication *Sources and Methods Labour Statistics, Vol. 5: Total and Economically Active Population, Employment and Unemployment (Population Censuses)*. Geneva: ILO. The Third Edition (2003) of this publication is available in electronic form only in the on-line database <http://laborsta.ilo.org/>.

economic characteristics in population censuses, based on relevant experiences of countries, with particular focus on the questions used and the requirements for processing of responses. A separate chapter discusses the role that the census results on economic characteristics can play in the planning of statistical surveys. The *Handbook* is intended to provide census planners with a variety of approaches to assess the questions and methods of collecting economic characteristics used in their national census, as they evaluate the performance in the past decade and plan for the current decade round of censuses (2005 – 2014). Users of census results may also find this document useful when evaluating the quality of census results.

The publication of this *Handbook* is a result of collaboration between the Bureau of Statistics of the International Labour Office (ILO) and the United Nations Statistics Division (UNSD). The original *Technical Report* on which it is based was first drafted under the supervision of the ILO<sup>6</sup>, reviewed by a group of experts and finalized under the supervision of UNSD. That *Report* benefited from inputs of several writers. Special mention is made of Mr Reg Gilbert (now deceased), who prepared the concepts and national experiences presented in chapters 4 to 8 of the *Handbook*, as well as most of the suggestions for questions to be tested.

UNSD and ILO invite feedback from the users of this *Handbook*, on national experiences besides those presented here; on the relevance and practicality of the suggested (untested) questions; on how well the approaches presented have worked; and on the circumstances under which they were found to be most effective. Such feedback will be useful in preparations for subsequent census rounds.

---

<sup>6</sup> An early version of the material presented in parts 2 and 3 is available in Gilbert, R. *Asking questions on economic characteristics in a population census*. STAT Working Papers 2001-1. ILO, Geneva. 2001. The material presented in chapters 10 and 11 is available in Hoffmann, E. *Coding occupation and industry in a population census*. STAT Working Paper 2001-2. ILO, Geneva. 2001.

Comments or additional material may be sent to:

<p>The Director</p> <p>Attn: Demographic and Social Statistics Branch</p> <p>Statistics Division</p> <p>Department of Economic and Social Affairs</p> <p>United Nations</p> <p>New York, NY 10017</p> <p>USA</p> <p>e-mail: <a href="mailto:socialstat@un.org">socialstat@un.org</a></p> <p>Fax: + 1-212-963-1940</p>	<p>The Director</p> <p>Bureau of Statistics</p> <p>International Labour Office</p> <p>CH-1211 Geneva 22, Switzerland</p> <p>e-mail: <a href="mailto:stat@ilo.org">stat@ilo.org</a></p> <p>Fax: + 41 22 799 6957</p>
---	---

# Table of Contents

<i>PREFACE</i> .....	<i>iii</i>
<b>PART 1: GENERAL ISSUES</b> .....	<b>1</b>
<b>CHAPTER 1: FOCUS AND CONTENT – RATIONALE AND STRUCTURE</b> .....	<b>3</b>
A Overview .....	3
B Purpose of the Handbook.....	5
C Users and uses of the Handbook.....	8
D Organization of the Handbook.....	9
<b>CHAPTER 2: PLANNING AND DESIGN OF POPULATION CENSUSES FOR THE COLLECTION OF DATA ON ECONOMIC CHARACTERISTICS</b> .....	<b>11</b>
<i>Uses and Sources of Statistics on Economic Characteristics</i> .....	<i>11</i>
A Uses of statistics on economic characteristics .....	11
B The population census compared with other sources of data.....	12
<i>Census Administration from the Perspective of Economic Characteristics</i> .....	<i>15</i>
C General considerations in census administration.....	15
D Preparation for enumeration .....	16
E Decisions about coverage and format of the enumeration .....	19
1. Scale of the census and the use of sampling to expand content .....	19
2. Issues relating to counts of population present and usual resident population .....	20
3. Method of census enumeration.....	22
4. Forms used to conduct enumeration.....	23
F Selection, training and supervision of field staff.....	25
G Field tests and the pilot census.....	27
H Issues relating to questionnaire design for data processing.....	28
I Data processing and dissemination of census results.....	30
<i>Decisions related to topics for Measuring Economic Characteristics</i> .....	<i>32</i>
J Selecting topics to be included.....	32
K Determining the location and sequence of the economic items on the census form .....	33
L Formulating questions on selected topics.....	34
<b>PART 2: MEASUREMENT OF ECONOMIC CHARACTERISTICS IN A POPULATION CENSUS</b> .....	<b>39</b>
<b>CHAPTER 3: MEASUREMENT FRAMEWORKS</b> .....	<b>41</b>
A Understanding the concept of economic activity.....	41
B The scope of production in the system of national accounts .....	42
C Correspondence between economic activity and production within the SNA production boundary .....	46
D Decisions related to measuring economic characteristics .....	46
1 Setting the reference period.....	47
2 Setting age limits.....	50
E Organization of Chapters: Rest of Part 2 & Part 3.....	53
1 Relevance of past experiences .....	53
2 Format of the presentation.....	54
<b>CHAPTER 4: THE CURRENTLY ECONOMICALLY ACTIVE POPULATION (THE LABOUR FORCE)</b> .....	<b>57</b>
A The Labour Force Framework .....	57
B The Employed Population.....	58
1 The operational definition .....	58
.....	59
2 Requisites for applying the definition .....	59
3. Special cases: identification of the employed .....	60

4.	Employed: Review of questions used in national censuses .....	63
5	Employed: Suggestions for enhancing effectiveness of questions.....	72
6	Employed: Issues of measurement .....	76
C	The Unemployed Population .....	76
1.	Unemployed: The operational definition .....	77
2.	Unemployed: Requisites for applying the definition .....	77
3.	Unemployed: Review of questions used in national censuses .....	79
4.	Unemployed: Suggestions for enhancing effectiveness of questions .....	87
5.	Unemployed: Issues of measurement.....	88
D	Population not currently active.....	88
<b>CHAPTER 5.</b>	<b>USUAL ECONOMIC ACTIVITY STATUS.....</b>	<b>89</b>
A	Introduction.....	89
B	The operational definition.....	90
C	Requisites for applying the definition.....	90
D	Review of questions used in national censuses .....	93
E	Suggestions for enhancing effectiveness of questions.....	99
F	Issues of measurement .....	102
<b>PART 3:</b>	<b>MEASUREMENT OF CHARACTERISTICS OF JOBS, ESTABLISHMENTS AND</b>	
	<b>PERSONS .....</b>	<b>105</b>
<b>CHAPTER 6.</b>	<b>DESCRIPTIVE CHARACTERISTICS FOR THE MAIN JOB.....</b>	<b>107</b>
A	Status in employment .....	108
1.	The operational definition .....	109
2.	Requisites for applying the definition .....	109
3.	Review of questions used in national censuses .....	110
B	Suggestions for enhancing effectiveness of questions.....	114
C	Issues of measurement .....	114
D	Occupation.....	117
1.	The operational definition .....	117
2.	Requisites for applying the definition .....	118
3.	Review of questions used in national censuses .....	118
4.	Suggestions for enhancing effectiveness of questions.....	122
5.	Issues of measurement.....	122
E	Place of work.....	124
1.	The operational definition .....	125
2.	Requisites for applying the definition .....	125
3.	Review of questions used in national censuses .....	126
4.	Suggestions for enhancing effectiveness of questions.....	130
5.	Issues of measurement.....	130
<b>CHAPTER 7.</b>	<b>DESCRIPTIVE CHARACTERISTICS OF THE ESTABLISHMENT.....</b>	<b>131</b>
A	Industry .....	131
1.	The operational definition .....	131
2.	Requisites for applying the definition .....	131
3.	Review of questions used in national censuses .....	132
4.	Suggestions for enhancing effectiveness of questions.....	135
5.	Issues of measurement.....	135
B	Institutional sector.....	137
1.	The operational definition .....	137
2.	Requisites for applying the definition .....	138
3.	Review of questions used in national censuses .....	140
4.	Suggestions for enhancing effectiveness of questions.....	142
5.	Issues of measurement.....	142
C	Informal sector and informal employment.....	143

1.	The operational definition .....	145
2.	Requisites for applying the definition .....	146
3.	Review of questions used in national censuses .....	147
4.	Suggestions for enhancing effectiveness of questions .....	150
5.	Issues of measurement .....	155
<b>CHAPTER 8.</b>	<b>PERSON TOTALS .....</b>	<b>159</b>
A	Working Time .....	159
1.	The operational definition .....	159
2.	Requisites for applying the definition .....	160
3.	Review of questions used in national censuses .....	161
4.	Suggestions for enhancing effectiveness of questions .....	163
5.	Issues of measurement .....	163
B	Income .....	164
1.	The operational definition .....	165
2.	Requisites for applying the definition .....	165
3.	Review of questions used in national censuses .....	166
4.	Suggestions for enhancing effectiveness of questions .....	170
5.	Issues of measurement .....	170
<b>PART 4:</b>	<b>DATA PROCESSING, TABULATION, ANALYSIS AND DISSEMINATION .....</b>	<b>173</b>
<b>CHAPTER 9.</b>	<b>DATA PROCESSING, TABULATION AND OTHER CONSIDERATIONS .....</b>	<b>175</b>
A	Introduction .....	175
B	Data processing aspects for particular topics .....	176
1.	Current and usual economic activity status .....	176
2.	Employment .....	177
3.	Unemployment .....	177
4.	Status in employment .....	178
5.	Occupation .....	178
6.	Place of work .....	178
7.	Industry .....	179
8.	Institutional sector .....	180
9.	Informal sector .....	180
10.	Informal employment .....	181
11.	Working time .....	181
12.	Income .....	181
C	Dissemination .....	182
1.	Printed publications .....	183
2.	Electronic dissemination .....	183
3.	Micro data .....	183
4.	Confidentiality and data security .....	183
5.	Graphical and related output .....	184
D	Meta Data .....	184
E	Emerging technologies for data processing and output .....	185
F	Data capture and coding methods .....	185
G	Output .....	186
<b>PART 5</b>	<b>CODING OF OCCUPATION AND INDUSTRY .....</b>	<b>189</b>
<b>CHAPTER 10:</b>	<b>PREPARATIONS FOR CODING OCCUPATION AND INDUSTRY .....</b>	<b>190</b>
A	Objectives .....	190
B	Strategic coding and processing options .....	191
1.	Process all cases or a sample only .....	191
2.	Field or office coding? .....	192
C	Planning and organizing coding operations .....	199

1.	Finance and resources .....	199
2.	Expertise, experience and rehearsal .....	200
3.	Estimating coding rates .....	200
4.	Coding staff .....	200
5.	Coding teams and supervisors.....	201
6.	Coding tools .....	202
7.	Coding problems and queries .....	204
8.	Quality assessment and quality control.....	204
9.	Premises, infrastructure and equipment .....	205
10.	Process in one location or in several .....	205
11.	Handling of documents .....	206
12.	Use of automatic or computer assisted coding.....	206
13.	The problem of different languages .....	208
<b>CHAPTER 11. THE DEVELOPMENT AND USE OF CODING INDEXES .....</b>		<b>210</b>
A	What is a coding index? .....	210
B	Developing and updating the occupation coding index.....	212
1.	Sources of information for construction of the index.....	212
2.	Organization and structure of the index .....	214
C	Using the occupation coding index.....	217
1.	Using the occupational response.....	217
2.	Using ancillary information on industry or name and type of employer.....	218
3.	Use of other ancillary information for occupation coding .....	219
4.	Inadequate occupation responses and queries .....	220
D	Developing and updating the industry coding index.....	221
E	Using the industry coding indexes .....	225
1.	Using the industry response.....	225
2.	Using ancillary information on occupation.....	226
3.	Use of other ancillary information for industry coding.....	227
4.	Inadequate industry responses and queries .....	228
<b>PART 6: USING POPULATION CENSUSES TO IMPROVE LABOUR FORCE AND RELATED STATISTICS .....</b>		<b>230</b>
<b>CHAPTER 12. TYPES OF DATA COLLECTION ON THE LABOUR FORCE AND OTHER ECONOMIC CHARACTERISTICS.....</b>		<b>232</b>
A	Organization of the Part.....	232
B	Diversity of sources .....	232
C	Household-based inquiries .....	234
D.....	Establishment-based inquiries from information on the population's economic characteristics and activities .....	235
E	Administrative sources .....	237
<b>CHAPTER 13. LABOUR FORCE SURVEY STRUCTURE AND ARRANGEMENTS .....</b>		<b>240</b>
A	Purpose of the Labour Force Survey .....	240
B	Common characteristics of surveys on the economically active population.....	240
C	Diversity of labour force survey structure and designs.....	242
D	Survey frequency .....	243
1.	Continuing surveys for current data.....	243
E	Linkages of the labour force survey with other surveys.....	245
1.	Separate labour force surveys.....	247
2.	Labour force survey as vehicle for other data.....	247
3.	Labour force "module" attached to other surveys.....	247
4.	Omnibus multi-purpose and specialised surveys.....	248
5.	LFS as periodic round of a survey system with varying focus .....	249

<b>CHAPTER 14</b>	<b>SAMPLING AND RELATED USES OF CENSUS INFORMATION.....</b>	<b>252</b>
A	Sampling in conjunction with the population census.....	252
B	Census as source of sampling frames .....	253
1.	The survey population.....	253
2.	The sampling frame.....	255
3.	Common problems with sampling frames.....	259
4.	Physical representation and maintenance of the frame .....	262
5.	Master samples.....	264
C	Census-based area frames for labour force and other household surveys.....	267
1.	The frame of census EAs.....	267
2.	Defining primary sampling units.....	268
3.	Construction of strata.....	271
D	Selecting a master sample of EA's for household surveys.....	272
1.	Size of the master sample for household surveys.....	273
2.	Allocation across domains.....	273
3.	Selection of master sample areas.....	273
4.	Replications.....	274
<b>CHAPTER 15</b>	<b>THE POPULATION CENSUS AS A FRAME FOR ECONOMIC AND AGRICULTURAL SURVEYS AND CENSUSES {ILO COMMENT: THIS CHAPTER TO BE REFERRED TO FAO FOR COMMENT} .....</b>	<b>276</b>
A	Introduction.....	276
B	Censuses and sample surveys of establishments: Similarities with household survey design .....	277
C	Special features of sampling for economic surveys.....	278
1.	Heterogeneity.....	278
2.	Uneven distribution.....	279
3.	Sampling versus survey units.....	279
4.	Sampling different types of units.....	280
D	Data required for sample design and selection .....	280
1.	Information on economic characteristics of area units and corresponding measures of size.....	281
2.	Collecting information on economic characteristics of population census enumeration areas (EAs) .....	282
E	Primary sampling units and stratification.....	284
<b>CHAPTER 16</b>	<b>USE OF CENSUS DATA FOR THE PRODUCTION OF SURVEY ESTIMATES .....</b>	<b>286</b>
A	Introduction.....	286
B	Weighting of sample data .....	287
1.	Source of information.....	288
2.	A step-by step procedure.....	289
C	Design weights.....	289
D	Non-response weights.....	290
1.	Defining weighting classes.....	290
2.	Data sources .....	291
E	Correcting the sample distribution with more reliable external information .....	292
F	Estimating totals .....	294
<b>CHAPTER 17</b>	<b>THE CENSUS AS A BASIS FOR EVALUATING SURVEY DATA .....</b>	<b>296</b>
A	Survey content .....	296
1.	Limitations of census data on labour force characteristics.....	296
2.	Potential for comparison with and evaluation of LFS data.....	297
B	Survey coverage and 'representativeness' .....	298
C	The importance of preparing simple unbiased estimates .....	298



<b>CHAPTER 18</b>	<b>COMBINED USES OF CENSUS AND SURVEY DATA: CURRENT ESTIMATES FOR SMALL DOMAINS .....</b>	<b>300</b>
A	The need for small domain estimates .....	300
B	What are ‘small domains’? .....	300
C	Diversity of methods .....	301
1.	Illustration of a procedure .....	304
ANNEXES	.....	306

List of Tables:

TABLE 1: EXAMPLES OF NON-MARKET HOUSEHOLD PRODUCTION ACTIVITIES.....	43
TABLE 2: ILLUSTRATIVE SET OF ECONOMIC AND NON-ECONOMIC ACTIVITIES .....	45

List of Diagrams

DIAGRAM 1: ECONOMIC PRODUCTION AS DEFINED IN CURRENT SNA.....	42
DIAGRAM 2: LABOUR FORCE FRAMEWORK.....	58
DIAGRAM 3: CURRENT ECONOMIC ACTIVITY PROMPT CARD.....	74
DIAGRAM 4: USUALLY ACTIVE POPULATION.....	91
DIAGRAM 5: CONCEPTUAL FRAMEWORK: INFORMAL EMPLOYMENT .....	144
DIAGRAM 6: ILLUSTRATION OF A SMALL AREA ESTIMATION P ROCEDURE.....	305



**PART 1:           GENERAL ISSUES**



## **Chapter 1: Focus and Content – Rationale and Structure**

### **A Overview**

1. Countries have been collecting, for decades, through population censuses, data on economic characteristics of the population. For many countries the population census remains the only source for nationally representative statistics on economic characteristics, while for others the population census complements other sources such as household and labour force surveys, establishment surveys, and administrative registers. As the census is conducted only once every ten years (in a few countries once in every five years), the economic characteristics are often competing with demands for coverage of a wide range of other topics.
2. The decisions on content and topics to be covered are generally guided not only by the resource availability (which determines the size of the questionnaire and number of questions that can be accommodated), but also on the availability of other sources of similar statistics, their reliability and scope, as well as the potential usefulness of the statistics that the population census would generate. This is the context in which the question of whether to include economic characteristics in a census is considered.
3. The population census potentially covers the entire population, and therefore in the absence of a comprehensive population and administrative register, it is the only source that can provide information on each individual, and thus finely detailed statistics on occupations etc. for the country as a whole as well as for small geographical areas. However, some of the basic concepts on economic characteristics are complex and will in principle require a depth in interviewing and probing that is not easily achieved in the population census, except where these considerations are taken into account in the planning and execution of the census operation.
4. For more complete coverage of economic characteristics, inquiries based on a sample of the population that are designed for in-depth interview are generally more suitable for the collection of economic characteristics than a population census. Such inquiries may be attached to the population census or be household-based sample surveys such as labour force surveys. The labour force survey can provide more extensive coverage of this topic because of its specific focus, but will not be able to provide statistics with satisfactory precision for many of the groups that may be of interest for policy formulation and implementation purposes, or for analysis of the economic and social structure of the country and its regions.
5. Thus the population census (or for some countries, population and administrative registers) is an indispensable source for obtaining broad-based estimates on selected economic characteristics (such as the employed and unemployed; and some characteristics of the job or of the employing establishment). It also provides benchmark statistics, such as labour force participation and unemployment rates, occupational and industrial distributions, for labour policies and programmes, particularly at regional,

district and other levels of administration. Population censuses are also useful for evaluating and also producing national estimates from sample-based inquiries of these characteristics. It is therefore important to produce accurate and reliable estimates through population censuses and provide better results, not only for those countries for whom this is the only source of data on economic characteristics but also for those that carry out labour force surveys. An important underlying issue is how the population census design and procedures may be chosen so as to maximize its usefulness for these purposes.

6. Despite these many benefits, population censuses have some disadvantages, including the high cost to the collection agency, the heavy reporting burden on the community and a certain vagueness in the measurement of topics that would require many questions to correctly classify particular groups of the population. One must also reconcile the population census results with those from other data collections. Such reconciliation is facilitated if comparable concepts, definitions, units and classifications are used.

7. This *Handbook* complements other resource materials that are already available to help measure economic characteristics; namely: -

- (a) International standards on concepts and definitions, in:
  - (i) “Resolution concerning statistics of the economically active population, employment, unemployment and underemployment”, of the Thirteenth International Conference of Labour Statisticians, 1983;
  - (ii) Definition of SNA production boundary, and related concepts, such the institutional sector, in the *System of National Accounts 1993*;
  - (iii) “Resolution concerning Statistics of Employment in the Informal Sector”, in the *Report of the Conference of the International Conference of Labour Statisticians, 19-28 January 1993*;
  - (iv) “Guidelines concerning treatment in employment and unemployment statistics of persons on extended absences from work, of the Sixteenth International Conference of Labour Statisticians, 1998;
  - (v) “Guidelines concerning a statistical measurement of informal employment”, of the Seventeenth International Conference of Labour Statisticians, 2003;
- (b) International standard classifications of industry, occupations and status in employment:
  - (i) *International Standard Classification of Occupations (ISCO)*,<sup>7</sup>

---

<sup>7</sup> Geneva, International Labour Office, 1990a.

- (ii) *International Standard Industrial Classification of All Economic Activities (ISIC)*,<sup>8</sup>
- (iii) *International Classification of Status in Employment (ICSE-93)*.<sup>9</sup>

(c) Methods on economic characteristics:

- (i) ILO Manual on concepts and methods, *Surveys of Economically Active Population, Employment, Unemployment and Underemployment*,<sup>10</sup>
- (ii) *Principles and Recommendations for Population and Housing Censuses, Revision 2, 2007*.<sup>11</sup>

8. This *Handbook* focuses specifically on population censuses. It discusses the strengths and weaknesses of population censuses as a source of data on economic characteristics and emphasizes the uses of these data and their role in the national statistical system. It reviews the general census operations, highlighting aspects that have direct relevance to the collection of economic characteristics data, and how these operations could take greater cognizance of the requirements of economic characteristics data. It also reviews some of the critical considerations for including economic characteristics items in the census, such as issues related to ordering of items and formulating questions. The *Handbook* discusses how to operationalize the conceptual underpinnings of the items on economic characteristics that are presented in the *Principles and Recommendations, Revision 2*, and highlights, with illustrative examples from national census questionnaires, aspects of question formulation that could provide more accurate representation of the relevant concepts. In some cases the *Handbook* proposes samples of questions (untested) that might be considered for trial and testing. Procedures for coding industry and occupations and the necessary tools to accomplish these tasks are also presented. The *Handbook* also deals with the comparability and complementarities of population censuses and household-based sample surveys that measure economic characteristics and discusses specific ways in which sample estimates could be enhanced with population census data.

## **B Purpose of the Handbook**

9. The primary aim of the *Handbook* is to draw on the experiences of countries in measuring economic characteristics and provide guidance on the implementation of the recommendations on the economic characteristics outlined in *Principles and Recommendations for Population and Housing Censuses, Revision 2*. This is a revision of the 1998 edition of the same publication, which the United Nations undertook in order

---

<sup>8</sup> Statistical Papers, No. 4, Rev. 3 (United Nations publication, Sales No. E.90.XVII.11).

<sup>9</sup> Report of the International Conference of Labour Statisticians, 19-28 January 1993.

<sup>10</sup> Hussmanns and others (1990). Geneva: International Labour Office.

<sup>11</sup> Statistical Papers, No. 67/Rev.2 (United Nations publication, Sales No. E.07.XVII.8.).

to take account of changes and new developments in related international standards on topics that are commonly included in national population and housing censuses.

10. The main changes specific to economic characteristics that have taken place since the 1970 and 1998 editions of the *Principles and Recommendations* that were published, and which are reflected in the current revised edition, are: (a) the revised definition of the concept of the economically active,<sup>12</sup> (b) a revised delineation of the boundary of production in *United Nations System of National Accounts 1993 (SNA)*,<sup>13</sup> (c) revised versions of the standard international classifications of industry (i.e., ISIC Revision 3), occupation, (i.e., ISCO-88) and status in employment, (i.e., ICSE-1993), and (d) the adoption, in 1993, of a framework for defining the informal sector, by the International Conference of Labour Statisticians (ICLS).<sup>14</sup> The first two revisions (that is of the concept of economic activity and of the SNA production boundary) have profound implications for the definition as well as measurement of economic activity, and necessitate a review of questions and methods used by many national census offices to collect data on economic characteristics.

11. A majority of countries, at least 180, {UNSD COMMENT: THIS WE CAN CHECK BEFORE (OR AFTER) THE EGM. PLACE OF WORK, HOWEVER, IS NOT AS COMMONLY ASKED AS THE OTHER CHARACTERISTICS.} have in the past included items on economic characteristics (in particular, employment and unemployment, status in employment, industry, occupation and, less frequently, place of work) in their census taking.<sup>15</sup> It is expected that while many users of census results will want statistics on employment in the informal sector and on informal employment for analysis of local labour markets situations, only a few census operations will consider including these topics, as the identification of those employed in such activities is quite complex and will require more probing than is possible in many census operations. However, by outlining how this could be done in a census this *Handbook* provides countries with the option of including the informal sector and informal employment and also gives guidance on the extent to which other characteristics included in the census may be used to derive estimates of employment in the informal sector and informal employment.

---

<sup>12</sup> 1983 Thirteenth International Conference of Labour Statisticians: Resolution concerning statistics of the economically active population, employment, unemployment and underemployment. Presented in *Current International Recommendations on Labour Statistics*. Second edition. Geneva: International Labour Office, 2000.

<sup>13</sup> Commission of the European Communities, and others (1993). *System of National Accounts 1993* (United Nations publications, Sales No. E.94.XVII.4), Chapters VI and XVII.

<sup>14</sup> Presented in *Current International Recommendations on Labour Statistics*. Second edition. International Labour Office, Geneva: International Labour Office, 2000.

<sup>15</sup> *Handbook of Population and Housing Censuses: Part IV Economic Activity Status, 1996*. United Nations publication, Sales No. E.96.XVII.13, Table 4. See also *Sources and Methods: Labour Statistics*. Vol. 5, second edition: Total and economically active population, employment and unemployment. ILO, Geneva, 1996.



12. As a supporting document to the *Principles and Recommendations, Revision 2*, the *Handbook* provides more detailed and technical operational guidance concerning the recommended economic characteristics than is contained in the *Principles and Recommendations, Revision 2* in three respects: (i) elaborating on the internationally recommended concepts in light of the practical possibilities, (ii) examining illustrative examples of questions used in national population census and their limitations; and (iii) suggesting suitable procedures for collecting and processing the relevant information, drawing on national and international experiences.

13. The *Handbook* does not discuss the theoretical justifications for the underlying concepts and their definitions, but concentrates instead on the practical issues of identifying the economically active population in a population census. It aims to translate the current SNA-93 boundary into questions that can be used in population census enumerations in which the economic characteristics of the population are being collected. While the *Handbook* focuses primarily on techniques and methods to be used in population censuses, the suggestions made are also applicable to household sample surveys which include economic characteristics as background variables, as well as for specialized labour force surveys and when samples are included in censuses to investigate the economic characteristics. There is, however, a wider latitude in terms of the type and number of questions that can be included in a sample survey whose focus is to measure employment characteristics. Other manuals on the topic of surveys<sup>16</sup> need therefore to be consulted for additional and more specific guidance when making preparations for such specialised surveys.

14. In addition to the *Principles and Recommendations for Population and Housing Censuses*,<sup>17</sup> mentioned earlier, the *Surveys of economically active population, employment, unemployment and underemployment: An ILO manual on concepts and methods*, published by the International Labour Office<sup>18</sup> provides extensive discussions on the concepts as well as conceptual basis for many of the decisions related to parameters that are needed to make the concepts operational. This ILO Manual was written before the SNA-93 was finalised and uses a concept of economic activity that was updated in SNA-93. An article published in the ILO's *Bulletin of Labour Statistics 2007-1* (Husmanns: *Measurement of employment, unemployment and underemployment – current international standards and issues in the application*, see [www.ilo.org/public/english/burea/stat/papers/listart.htm](http://www.ilo.org/public/english/burea/stat/papers/listart.htm)) updates the definition of economic activity to be consistent with SNA-93. In addition to the *Principles and Recommendations* and the ILO Manual, several other United Nations and International Labour Office publications provide some relevant information and guidance on some of the topics and issues addressed here. These include: *“What Kind of Work Do You Do?”*

---

<sup>16</sup> See in particular, Husmanns and others, 1990.

<sup>17</sup> The most recent version was adopted by the United Nations Statistical Commission at its thirty-eighth session in 2007 and published in 2007 as *Principles and Recommendations for Population and Housing Censuses, Revision 2* (United Nations publication, Sales No. E.07.XVII.8). For ease of reference, throughout this guide it will be cited simply as *Principles and Recommendations, Revision 2*.

<sup>18</sup> By Husmanns and others (1990), Geneva: International Labour Office.

*Data collection and processing strategies when measuring occupation for statistical surveys and administrative records*”,<sup>19</sup> a Working Paper issued by the International Labour Office, and *Handbook of Population and Housing Censuses, Part I: Planning, Organisation and Administration of Population and Housing Censuses*,<sup>20</sup> and *Handbook of Population and Housing Censuses, Part IV, Economic Activity Status*<sup>21</sup> both published by the United Nations.

## **C Users and uses of the Handbook**

15. The *Handbook* is mainly addressed to census planners, and managers responsible for designing and carrying out the census; as well as other producers of statistics, including survey experts, data processing staff and trainers. For these producers, the *Handbook* is intended as a basic reference for (i) identifying potential problems in measuring specific characteristics and in the formulation of certain types of questions, (ii) considering a range of possible questions or improving upon previously used questions, and (iii) deciding on the extent to which economic characteristics are to be covered in the population census. It is a guide not only for the application of international standards for data collection but also for coding and adapting these standards to national classification requirements. Many of the issues raised are also important considerations in training of field staff and in developing manuals for enumerators and their supervisors.

16. This *Handbook* is addressed also to users of census statistics on economic characteristics. The main users of these statistics have generally been various national government ministries, committees, and agencies concerned with the formulation, implementation and evaluation of economic, social and labour policies. Provincial and local government authorities with similar responsibilities and concerns are also a core set of users. Development programme officers also have need of such data, e.g., when planning and implementing—at the national, regional and local levels—projects for building of roads, water supply, electricity and telephones as well as waste disposal; or when deciding on the location of special zones for industrial development; etc. There are in addition, at least potentially, a wide range of different users, among whom are:

- (i) those who are interested in understanding the economic and social structure of a country and of its small as well as large regions and population groups;

---

<sup>19</sup> International Labour Office, Bureau of Statistics, STAT Working Papers, No. 95-1(February 1995). Geneva. Prepared by Eivind Hoffmann from contributions by Peter Elias, Brian Embury and Roger Thomas.

<sup>20</sup> Series F, No 54 (Part I). United Nations publication, Sales No. E.92.XVII.8. New York: United Nations, 1992.

<sup>21</sup> Series F, No 54 (Part IV), United Nations publication, Sales 96.XVII.13. New York: United Nations, 1996.

- (ii.) employers' organizations and managers and planners in private sector enterprises who need to understand their markets and the potentials for recruitment of their most important resource: people to work in their establishments;
- (iii.) members of the research community (at universities and in economic, labour and social research institutes), political leaders and other policy makers, social activists and representatives of special interest groups, journalists and international agencies; and
- (iv.) those concerned with the welfare of workers and their households, such as the labour unions, as well as those providing financial, marketing and social services to either management or labour.

17. Increasingly, advocates of gender issues and equal opportunity questions are becoming more involved as a growing user group with very specific data availability and quality concerns, particularly as relates to the different work statuses of women and men, as well as girls and boys.

18. Users of statistics need a broad knowledge of the issues of data collection in order to engage in and maintain a fruitful dialogue with producers at all stages of the census operation, and with them arrive at a mutual understanding of the kinds of statistics for which the population census is well suited, as well as the limitations of such statistics. The *Handbook* will acquaint users of these statistics with the concepts, definitions, and methods and the constraints that producers of statistics face while responding to users' needs. With this understanding, census planners and users of census results may enter into a mutually beneficial dialogue about the content of and procedures for the census in the planning stages. Equally important, informed users can meaningfully communicate to the census authorities about the form that tabulations and other census outputs should take to best meet the needs of users and provide feedback on the adequacy of the data relative to their data requirements.

## **D Organization of the Handbook**

19. There are four main parts in this Handbook in addition to this introductory chapter. Part 1 which also covers the planning and design of population censuses. The second Part reviews the measurement frameworks and general issues related to the inclusion of economic characteristics in the population and housing census. Part 2 also presents conceptual definitions associated with activity status and discusses some of the constraints to measuring economic characteristics from population censuses.

20. It also addresses in more detail practical considerations in collecting information for each of the characteristics relating to economic activity that are included in the Principles and Recommendations, Revision 2. It illustrates these details with examples of questions used in national censuses, indicating some aspects of the questions that might be modified to enhance their effectiveness. The chapter also emphasizes the role of training and discusses relevant issues to be investigated.

21. The third Part continues the more detailed presentation of definitions, examples and suggestions for enhancing effectiveness but focuses on the measurement of the characteristics of jobs, establishments and persons.

22. The fourth Part deals with data processing, analysis and tabulation. It describes the procedures for checking for errors and assuring quality control. The discussion on tabulation outlines those tables that are considered essential, separately from those that are recommended and those that are only optimal.

23. Part 5 focuses on two main characteristics for which there are international standard classifications, i.e., occupation and industry. It discusses the possible strategies that the census planners can adopt for the inclusion of these variables and their classifications, as well as the implications for enumerators or office coders to code responses to the questions on the census questionnaire. It describes the tools to be used for coding operations, such as coding indexes and instructions, and explains how to develop and test these tools.

24. The sixth Part covers issues of comparability between census and labour force/household survey results and discusses some specific uses of population census results, such as in preparing a sampling frame for labour force surveys, evaluating labour force data, and for small area and small group estimation. It reviews sources of differences between labour force data collected through the population census and the labour force survey, with respect to the scope, coverage, timing, survey instruments, extent of application of the concepts and definitions, methods, sampling and non-sampling errors, etc., and proposes ways to maximise the effective use of census results, in improving sample designs and estimates from labour force surveys.

## **Chapter 2: Planning and Design of Population Censuses for the Collection of Data on Economic Characteristics**

### **Uses and Sources of Statistics on Economic Characteristics**

#### **A Uses of statistics on economic characteristics**

25. Social and demographic information on the population engaged in the production of goods and services is vital to the analysis of the economic performance of a country as well as for the regions of which it is composed. One of the main strengths of the population census as a source of data is the provision of locality-specific statistics. As a first step to including economic characteristics questions in a census questionnaire, census authorities and potential users of census results need to work together to develop lists of the actual and anticipated uses of census statistics in their own country. Such lists will be helpful in planning the census content and types of outputs, as well as in mobilizing support for the census.

26. Statistics on the scope and structure of the economic activities of the population provide a useful basis for the:

- (a) formulation and planning of economic and social policies and campaigns, for example, planning for employment promotion in public and private sectors, special remedial programmes to reduce unemployment, underemployment and other decent work deficits, and making decisions about the location of new economic activities and markets.
- (b) administration and management of programmes, such as unemployment benefits and welfare schemes;
- (c) setting of quantitative targets and thresholds to be achieved in the implementation of or as a result of policies, studying labour market behaviour, and local labour market situations as a basis for making investment, zoning, and community development decisions, and identifying education, and technical and vocational training needs as well as understanding linkages with poverty alleviation;
- (d) evaluation of national or local employment or social welfare programmes and policies, including assessing the current state of the economy and providing indications of differences in the economic opportunities among different population subgroups;
- (e) conduct of scientific research and hypothesis testing, such as construction of working life tables and related estimations used to design pension schemes and other social insurance programmes; and

- (f) teaching of and general descriptions about the economic and social situation, analysis of the labour market situation for major population groups or subgroups in different localities, regions or the whole country.<sup>22</sup>

27. In addition to the examples of direct use of census results, the census is also used as a basis for preparing more intensive sample surveys to study special issues. The census can provide the basis for:

- (a) the development of a national statistical capability and particularly of sample surveys and providing a convenient and efficient basis for launching continuing survey programmes;
- (b) constructing the sampling frame and for sample design and selection, not only for population surveys but also for other types of surveys, such as agricultural surveys of farming households and holdings, economic surveys of establishments and of economic activities in the informal sector;
- (c) stratifying the sample for a survey, e.g., to ensure that the sample will include sufficient number of relatively rare circumstances, which may be of interest to planners; and
- (d) evaluating sample survey results;
- (e) producing small-area estimates.

More detailed discussion of this subject is presented in Part 6.

28. In spite of their potential, the usefulness of census statistics on economic characteristics to particular users may be limited by factors such as: timeliness of the census results; the form in which they are disseminated to the users; and the number of topics on which data are available. In addition users' need for specialized cross-tabulations (with multiple variables) may not be fully taken into account, or the statistics may not be made available in formats that lend themselves to further analysis. This means that it is important to consider carefully not only which characteristics to include in the census, but also who should be the main targets for the dissemination of the results. It also means that it is necessary to plan carefully the effective and timely processing of the data that have been collected and the dissemination of the results.

## **B The population census compared with other sources of data**

29. Sound planning and design of a population census should take into account other available sources of statistics on the economic characteristics of the population. A fully developed national statistical system will also involve (a) one or more national survey programmes based on samples of households or persons, including periodic labour force and income and expenditure surveys, (b) industrial or establishment-based censuses and surveys, and (c) statistics based on a range of administrative reporting systems, such as population registers, social security and pension records, income tax records, and

---

<sup>22</sup> See United Nations (1998, 1996a) and Hussmanns and others (1990), for discussion on uses of economic characteristics data.

unemployment compensation reports.<sup>23</sup> Census designers can then focus their attention on those types of data needs for which the census is best suited or represents the only alternative.

30. In many countries the population census is the only form of measurement for obtaining complete information on the economic activity rates for different population groups, as well as their occupational and industrial distributions. Even though the population census is generally not the most effective device for collecting information on economic characteristics, in particular when the statistics needed are mainly for large groups at the national level, in many countries it is still the most used vehicle for collecting such information. Questions on economic characteristics have been included in almost all population census enumeration exercises, according to an analysis of 1995-2004 round of censuses and census-type count<sup>24</sup> {UNSD COMMENT: NEED TO CHECK LATER WHETHER THIS IS STILL CORRECT FOR 1995-2004}. The major advantages of the census compared to alternative sources are: (a) its national scope and consistent geography, (b) availability of other demographic and socio-economic variables for the same individuals, (c) that it includes all persons, permitting extensive and detailed cross-tabulations and the testing of complex models, and (d) the possibility of examining long-term socio-economic and demographic trends through comparisons of statistics from two or more censuses (assuming a certain minimum degree of conceptual and operational consistency between the censuses).

31. The population census covers the entire population in all locations in the country, and therefore potentially covers all types of economic activity, including the most rare occupation or highly localized branch of activity, for even the smallest population group as determined by other variables. A census can produce, for example, summary measures of labour force participation for all civil or administrative divisions of the country, consistent with those for the country as a whole. Tabulations can be generated not only to the lowest geographic detail but also to the most detailed occupational and industrial distributions without having to be concerned about the precision of the estimates, as expressed by the sampling error<sup>25</sup>, that would be inherent in similar statistics derived from a sample survey. Moreover, multi-variate classifications and tabulations are feasible, within the limits of the census budget. {UNSD COMMENT: FROM JEAN MICHEL DURR: I DISAGREE WITH THIS PARA. IT IS THE USUAL NAIVETY TO BELIEVE THAT A CENSUS CAN DESCRIBE CORRECTLY VERY SMALL POP GROUPS. IT IS ON THE OPPOSITE MORE LIKELY THAT THE CENSUS PROCESS (QUESTIONNAIRE, ENUMERATORS, DATA CAPTURE, CLASSIFICATION, CODIFICATION....) WILL NOT TREAT CORRECTLY RARE ITEMS. BUT THIS IS LEARNED BY EXPERIENCE AND NOT IN THE UN HANDBOOKS! SO, I WOULD LIKE TO EXAMINE MORE DEEPLY THE TR

---

<sup>23</sup> See Hussmanns and others, 1990; p. 4, for a discussion of the comparative strengths and weaknesses of these various sources.

<sup>24</sup> This is to be checked and updated with new reference after Expert Meeting. xxxx.

<sup>25</sup> Concerns may be expressed by some census experts as to the quality of measuring rare items, but one of the purposes of this document to improve this.

ITSELF AND THE DRAFT REVISION. PEMBER COMMENT: HENCE I HAVE ADDED FOOTNOTE 25}

32. The main disadvantage of the population census as a source of statistics on economic characteristics is a function of its strength: because of its unparalleled geographical coverage it is also the major source of baseline data on the characteristics of the population in general. Economic characteristics thus represent just a few of the numerous topics that compete for coverage in a population census. Moreover, because the population census aims to cover the entire population in a country, the scale of operation is great and this presents a major constraint on the number of questions that can be included. These considerations substantially influence the decision of;

- (i.) whether or not economic characteristics should be covered,
- (ii.) how many questions can be included on economic characteristics,
- (iii.) who is to be the respondent, and
- (iv.) what depth of interviewing and probing can realistically be used to elicit the correct information on each individual?

33. In order to ensure that the recorded information and the resulting statistics from the census are accurate and give meaningful representations of economic activity in the country a lot of effort is required when designing the instruments for their collection and training enumerators. Typically, however, most population censuses have to abstain from the use of detailed questioning and probing required for identifying whether some individuals are economically active. Moreover, because the census interview relies on proxy respondents it does not always capture reasonably accurate information on certain characteristics which only the individual might know, such as occupation, industry, hours worked, income, sector of employment, etc., (but to a large extent it shares these problems with the other statistical sources mentioned).

34. As a consequence, the reliability of the measurement of each respondent's characteristics as obtained in a well-run sample survey is likely to be higher than in a census, with consequent implications for comparability of results. By the same token, administrative reporting systems are more likely than the census to closely track those concepts and data items that are important for the agency and staff responsible for the administrative reporting system.

35. The censuses in some countries are administering long forms on a sample basis in order to investigate certain topics in greater depth. In such cases, the basic census questionnaire is often the short form that is applied to all individuals, while the designated sample receives the long form, which covers the questions on the short form as well as several others that may be investigated in greater depth. Data collected for a sample in the census exercise, however, have similar limitations as any sample survey data. In addition to the non-sampling errors (such as under coverage of the target population, non-response, recall lapses, inaccuracies in reporting, etc.), common to both the population census and sample surveys, the estimates based on a sample rather than the whole population will suffer from imprecision due to the sampling. Moreover, the



potential for generating estimates at the smallest level of geographical detail and for coverage of rare occurrences of certain characteristics such as certain occupations and types of economic activity, etc., are greatly diminished with the use of sampling. However, the strengths of household survey-based statistics include the flexibility of the data collection instrument to accommodate a larger number of questions on a variety of topics; and also the possibility of estimating parameters comparable to those measured in the population census, within the margins of error associated with the sample design.<sup>26</sup>

36. Specialized surveys such as time-use surveys, which are increasingly being used in many countries, is another source that can be designed to provide comprehensive and accurate measures of the extent of economic activity in a country, for large groups of the population. Time use methodology may be used as a tool for improving statistics on unremunerated work of women, men, girls and boys. As at this stage, such studies have not been extensively used for this purpose. Many of them, particularly those few in developing countries, have been small studies restricted to a single or a few localities—and therefore do not allow the use of their results for national estimates. Nevertheless the experiences with these studies can serve to sharpen the instruments and questions used to measure economic activity in national censuses and household surveys. A guide on collecting time-use statistics, with a proposed international classification for time-use statistics has been issued by the United Nations to support this area of work.<sup>27</sup>

## **Census Administration from the Perspective of Economic Characteristics**

### **C General considerations in census administration**

37. The search, in census design and planning, for the proper balance between substantive and operational considerations, is not an easy task in the face of conflicting demands, resource constraints, the inter-related nature of various aspects of the entire census process and uncertainties about how full-scale operations will proceed. Part I of the *Handbook of Population and Housing Censuses*<sup>28</sup> and the *Handbook on Census Management*<sup>29</sup> {PEMBER COMMENT: 1992 HANDBOOK IS NOT SHOWN ON UNSD WEBSITE AND PERHAPS COULD BE DELETED. OTHER MORE RECENT PUBLICATIONS MIGHT BE CITED. UNSD TO ADVISE PLEASE} present detailed discussions on issues related to the planning, organization and administration of population and housing censuses, in general. In this section, issues that have a direct bearing on obtaining data on economic characteristics are highlighted and decisions on items that have significant implications for over-all census planning are discussed in the context of measuring economic activity.

---

<sup>26</sup> See Hussmanns and others (1992), chapters 1 and 10 for discussions on sample surveys.

<sup>27</sup> United Nations (2004) “Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid work”, Sales No. E.04.XVII.7.

<sup>28</sup> United Nations (1992), Sales No. E.92.XVII.8.

<sup>29</sup> United Nations (2000). Handbook on Census Management for Population and Housing Censuses, Sales No. E.00.XVII.15.

38. Census administration involves establishing the organizational structure, which has typically included organizing the central, regional and field offices; designating personnel and recruiting additional staff, as necessary; constituting a planning group; and convening technical advisory committees. It is important that the census-planning group has at its disposal expertise in the collection, compilation, processing and analysis of statistics on economic characteristics. Ideally, especially given the intricacies of the concepts associated with economic characteristics and the difficulties they present in their application (detailed discussions are presented in the next section) it would be necessary to have a labour statistics expert on the planning team when any of the economic characteristics are among the topics to be covered in the census enumeration. This may be ensured, at the minimum, by having experts in this field on one or more of the advisory committees (depending on how these committees are organized). Two major forms of organisation are proposed in Part I of the *Handbook* series: one is to form committees by subject-matter areas, in which case one of the committees would need to be on economic characteristics. The other is to set up a committee with subject-matter subgroups, again with one such sub-group devoted to economic characteristics. If the former is chosen, it is important that there are opportunities for all the committees to meet together, to ensure consistency and coherence in approaches and better coordination. The advisory committees provide professional input and represent the user perspective. It also strengthens the census planning by providing a forum for certain decisions to be debated among a wider expert audience.

## **D Preparation for enumeration**

39. Some of the pre-enumeration activities, ranging from the determination of enumeration areas to carrying out organisational pre-tests, apply to the census exercise as a whole. Other activities, such as deciding on the content of the census and the design and testing of the questionnaire are specific to the characteristics that are being considered for inclusion in the scope of the census. The latter activities are the focus of this section.

40. *Mapping:* One aspect of the mapping exercise that needs particular attention, is mapping non-typical areas, such as large tracts with sparse population, rapidly growing areas, congested urban areas and areas with ill-defined boundaries.<sup>30</sup> This is intended to ensure the full coverage of the population in a census; but it is also of particular importance for measuring economic activity. In many agricultural societies, some farmers or farm workers reside on part of their farmland. These tracts of land may be difficult to reach or to identify clearly defined boundaries. Such problems could lead to the under-enumeration of persons employed and residing on these types of farms. Areas with very dense population such as the slums in some major cities may also present a problem for the measurement of economic characteristics: Many of the residents of these areas are involved in marginal activities that tend not to be reported with the traditional

---

<sup>30</sup> See paras. 189 to 193 of Handbook of Population and Housing Censuses: Part I Planning, Organization and Administration of Population and Housing Censuses, (United Nations publication Sales No. E.92.XVII.8).

census questions on economic activities. In addition, the crowding in these areas could present difficulties in the listing of households, leading to the undercount of the population and therefore under enumeration of these marginal activities; as well as unequal division of the enumeration areas and workload.

41. *Pre-tests:* The use of pre-tests to determine the efficacy of the data collection instrument, field operations and even data processing (as noted later) is a critical issue to consider for every census round since circumstances and the characteristics of the population change in a way that could affect the way respondents receive and respond to questions. Both new formulations of questions and those that were previously used need to be subjected to pre-tests. It is particularly indispensable, in light of the changes in the scope of the production concept and consequently in the intended understanding of “economic activity”, for pre-tests to be conducted to ensure that persons participating in any of the relevant activities can be separated from those that have not participated in any of them during the defined reference period.

42. *Producer-user consultations:* The importance of communicating information about census plans, operations and outputs to user groups, opinion leaders and the general public through out the entire census process cannot be over emphasized.<sup>31</sup> The primary initial responsibility for these communication and publicity efforts lies with the census office itself. At a certain point, however, major public and private sector users can make critically important contributions to these efforts. The users of statistics generated by the census are particularly well situated to provide reinforcement, through their own formal and informal communications channels, on the value of the census and the importance of cooperation by the public.

43. Users of the statistics have valuable inputs to make with respect to the relevance of the concepts and methods, the adequacy of the statistics that will be produced and the forms of dissemination, and can state how the statistics could best meet their needs. The users are not only those from the government side, such as the ministries and government departments, but also researchers and academicians as well as representatives of employers and workers’ organizations, as well as other non-governmental institutions and special interest groups (see para. 14). Two groups of users who should be represented in advisory committees are researchers who have analysed statistics on economic characteristics, and gender statisticians and analysts. In recent decades, many of these users have drawn attention to the limitations in available statistics on women’s economic activity and they will provide invaluable assessment and feedback on how to avoid some of these shortcomings in the statistics that will be available in the census.

44. Producer-user dialogue should also serve as a forum for educating the users about other data sources and their availability. Users need also to be informed of the outputs of the major alternative sources of statistics on the economic characteristics of the population, such as, household and establishment surveys and administrative reporting systems. From users’ perspective the sharing of census and other experiences will most

---

<sup>31</sup> See United Nations (2007); see paras. 1.112-1.116.

likely result in improvements in terms of content, detailing, validity, reliability, and timeliness, and, for some users, in addressing the concern that consistency be preserved between different data sources and from census to census. In many individual circumstances, improvements in terms of subject-matter detail, validity, reliability, and timeliness cannot, within the resources available, be obtained simultaneously so that trade-offs between different kinds of improvements will be involved. Moreover, improvements are often accomplished at a price, in terms of consistency over time and with statistics from other sources. Here again, trade-offs are involved. Informed users can assist in determining whether and how these trade-offs will be resolved in ways that are acceptable to the most important users of the resulting statistics.

45. *The role of users:* Forward-looking leaders of the financial and business community, who use census data to help with investment planning and marketing, may often be able to convince others in the financial and business community of the value of the census and census results. This support can be useful in securing political support for the census in its preparatory phases as well as more general public support during the actual enumeration, and in some instances mobilising resources from the private sector for the census activities. Labour organisations are also in a strong position to provide political and public support for the census both in the planning and enumeration stages. In both cases, it is useful that support for the census also continue during the post-enumeration stage of the census, such as processing, analysis, and dissemination. Periodic specialized reports about the census plan and progress, as well as outputs tailored to different sets of users are important tools for mobilizing continued support both in the business community and among labour representatives. The likelihood that users will take an active role in supporting the census is directly related to their involvement in consultations on census content and outputs.

46. There is considerable variation among countries as to who makes the final decision on census content. This responsibility may be given to the census commissioner, to a minister, a parliamentary body or provided for in a presidential decree. However, regardless of who has ultimate authority in the matter, the process of providing input into that decision should be as broad as possible. Users of census data on economic characteristics should be invited to play an active role in this consultation process and, once invited, should be encouraged to participate actively. It is important for all users and potential users to understand from the start that the process is subject to considerable give and take, with a considerable amount of compromising involved. As mentioned elsewhere, many trade-offs are involved not only between subject-matter interests, but also between subject-matter concerns and operational considerations. Moreover, because of the high degree of inter-relatedness among the various aspects of census design, it is important that the number of economic items included in the census, is balanced with other topics to be covered, to assure the production of reliable data or of the cost-effectiveness of the census operation.

47. *Building on past experiences:* In planning any new population census the logical starting point is the previous population census. Accordingly users should endeavour to obtain relevant documentation relating to the previous census. At a minimum this would

involve reviewing the questionnaire(s) or schedule(s) used in the last census, reading about the basic census procedures, and reviewing the main census outputs related to the economic characteristics of the population. In addition, evaluation studies of the previous census and any census monographs focussing on the economic characteristics that may have been produced can provide helpful information. In this regard, it is of great importance that census organisers prepare full documentation of the implementation of the census programme, not only of successful experiences but also of failures, so that both the users and current and future census planners benefit from them. Where documentation is limited, the production of such documentation needs to be considered and planned for early in the census-planning period.

## **E Decisions about coverage and format of the enumeration**

### ***1. Scale of the census and the use of sampling to expand content***

48. Census authorities often wrestle with pressure to expand the list of topics covered in the census beyond what they consider to be consistent with preserving public good will and a reasonable level of data quality in the over-all census results. These pressures are particularly acute in countries that lack alternate sources of data, for example, continuing programmes of sample surveys, to provide statistics on many important topics. It is important to note, however, that in general it is the statistically most developed countries, i.e., those with a range of different sources for national statistics that tend to also have the most comprehensive and complex population censuses. This suggests that considerations of the costs of the census and the capacities of the census organisation will normally be decisive in determining the scope of the census.

49. One strategy used in some countries, developed and developing, to ensure that the basic census form is relatively brief, is to use sampling as part of the census operation itself. This approach, which is described in *Principles and Recommendations, Revision 2*,<sup>32</sup> involves obtaining information on some topics for only a sample of the population enumerated in the census (see para. 31).

50. When sampling has been used in national population censuses (combined with the use of long form), economic characteristics have often been among the topics included in the long form. Due to the complexities inherent in the concepts related to economic characteristics, more than one question is usually required to measure an economic characteristic. Moreover, for some topics, such as, employment, income, informal sector, further probing may be required to ensure that all relevant situations have been covered. The decision to include economic characteristics in the census on a sample basis involves deciding (i) which, if any, of the characteristics should be in the short form, thus addressed to the whole population; and (ii) which of the characteristics should be included in long form, which is administered to only a sample of the population.

---

<sup>32</sup> United Nations publications, Sales No. E.07.XVII.8. Paras. 1.422-1.427.

51. The use of samples provides greater flexibility for investigating topics on economic activity. It can significantly reduce the response burden for respondents who are canvassed through the short form, as well as certain aspects of the census processing costs. The possible impact of the use of sampling on the cost of census fieldwork will vary with factors such as the method of sampling used, population density, and whether self-enumeration is involved. Gains in terms of reduced total operational costs, could be offset by the cost in terms of complicating the interviewing instructions and the procedures used in processing the census results and of the introduction of sampling errors into the census results. Nevertheless, some countries have found that the gains from the use of sampling to expand the subject-matter scope and detail of the census outweighed the costs in complications, which need to be addressed if the application of sampling is to be successful.

52. Census users and producers in some countries argue that, as the provision of small area (such as for localities, or polling district level), and small group data on a nationally consistent basis is one of the main functions of a census, restricting data collection (or processing) to only a sample of units is inconsistent with the basic objective of a census. It may also be relevant that in most small countries the relative cost reductions from the use of sampling in the collection or processing of census data are much more limited than in larger countries.

53. Sampling, if used, is usually carried out at the data collection phase and more rarely at the processing stage of census results. Sampling at data collection stage is carried out in three, broadly, different ways: (a) a sample of persons (that is, only one in  $n$  persons in the census), (b) a sample of households (that is, all persons in one in  $m$  households), or (c) an area sample of some sort (that is, all households and unrelated individuals in one in  $k$  enumeration areas or census blocks). Although the latter type of sample is the simplest to administer in the field and may result in the greatest cost saving, it will usually yield estimates with the large standard errors because of the intensity of clustering. Whatever design is used, an expert in sampling should be associated with the choice, the actual sample design, its implementation, and in the preparations of the estimates based on the sample, including estimates of sampling errors.

54. Sampling in data processing has been used in the past to produce advance tabulations or to control the costs of the processing operation. (Because the coding and processing of detailed occupation and industry information can be a time-consuming and costly part of the census processing, occupation and industry items are often among the first topics to be proposed for the sample, if sampling is to be used.) With modern processing technology, there is rarely a strong rationale for the use of sampling in processing census data.

## **2. *Issues relating to counts of population present and usual resident population***

55. The issue of whether people are generally enumerated at the place of their usual residence (that is, a *population present* approach) or at the place they are found at the time of the census (that is, a *usual resident population* approach) gives rise to much

discussion in many census offices around the world.<sup>33</sup> However, it should be stressed that for many countries, the issue is actually of minor importance in the context of the population census. If most of the population will be found at their usual places of residence on the census enumeration day, there is little practical difference between the two concepts. For this reason, countries often choose as the census date a day that maximizes, to the extent possible, the likelihood that persons will be found at their usual places of residence. Furthermore, most censuses actually use a combination of both approaches. For example, homeless persons are always enumerated on a *population present* basis even in a *usual resident population* census, while people away from their usual place of residence at midnight on the census reference night are usually enumerated on a *usual resident population* basis even in a strictly *population present* census if they are working or in travel status on enumeration night. {UNSD COMMENT: THIS SEEMS TO DOWNPLAY THE IMPORTANCE OF THE DIFFERENCE. JEAN MICHEL DURR'S COMMENT: THIS PARAG IS TO BE REWRITTEN IN ACCORDANCE WITH THE P&R. REGARDING ECO CHARAC, THERE IS ALSO AN ISSUE IN MEASURING DAILY MIGRATIONS (COMMUTERS), AND THE DEFINITION OF THE PLACE OF RESIDENCE AND PLACE OF WORK ARE IMPORTANT. (SEE MARGARET). PEMBER COMMENT: AWAITING MARGARET'S INPUT ON THIS POINT}

56. From the perspective of users of census statistics on economic characteristics the two main concerns should be: (a) which approach will minimize under-enumeration of the economically active population; and (b) which approach will yield the most reliable responses on the economic items. In both cases, there will be a trade-off between the coverage and response errors for visitors present compared to those for absent usual residents. Proxy-respondents, which are still used in many censuses, could present special problems with either approach, if someone other than the individual is reporting on characteristics such as occupation, industry, hours worked, income, and usual activity status.

57. Objective assessment of these trade-offs can only be derived from empirical evidence from earlier census, survey experience or census tests. Similarly, although in theory the census coverage concept adopted should be in accord with national accounting practices for the country,<sup>34</sup> the practical trade-off considerations relating to coverage and response errors should be determinative in deciding upon the concept to use in the census. Many national accounting concepts and distinctions are not possible to implement in the context of the census. In some countries where a substantial number of persons are absent for work-related reasons, information about absent household members can distort the results from the economic items under a *usual resident population* enumeration, unless those absent can be easily identified separately. For this reason, a time limit for absence is sometimes introduced. For example, to be included in the census the absent person has to have been at the household for a specified duration (e.g., at least six

---

<sup>33</sup> United Nations publications, Sales No. E.07.XVII.8. Paras. 1.461-1.468 and 2.71-2.77.

<sup>34</sup> *System of National Accounts 1993*, United Nations publication, Sales No. E.94.XVII.4, particularly paras. 14.7 - 14.21.

months) in the last year and/or have no other private dwelling in the country that could be called the person's usual residence.

### 3. *Method of census enumeration*

{ILO COMMENT: UNSD MAY WISH TO EXTRACT TEXT RELATING TO REGISTER-BASED CENSUSES FROM P&R, EG, PARA 1.64 OF REV 2}

58. A major design issue in planning any population census is whether the information will be obtained largely by interviews with each household or by self-administered questionnaires (which each household completes on their own without the use of census enumerators). The former approach is usually termed the "interview" method and the latter the "self-enumeration" method. The choice of which method will be the most cost-effective in terms of the level of over-all enumeration and the quality of data obtained depends on a variety of factors, including among others, the level of literacy in the population, the pool of persons available to serve as census enumerators, the size of the country and population settlement patterns.<sup>35</sup> Linked to any decision on using the self-enumeration approach, is how the census form is to be delivered to each household and, once completed, returned to the census organization. Countries with reliable postal systems and good address registers may rely primarily on a mail-out and mail-back approach; while other countries may use census enumerators to drop-off and/or pick-up the census forms. In other words, the operational and quality issues that go into the choice of methodology do not generally turn on issues related to any single subject-matter field.

59. The interview methodology is particularly suitable for the economic questions as it allows enumerators to probe, in a uniform manner, into the wide range of possible economic activities and their characteristics. However, the interview methodology places a heavy weight on the enumerator's understanding of the questions and how to ask them and record the answers, and thus on their selection and training, as well as on the questionnaire. By contrast, from the perspective of the economic questions, the self-enumeration methodology puts an even stronger dependence on a clear simple questionnaire and the few easily understandable notes, which a respondent with average literacy can read and follow. The questionnaire design and testing is important for both approaches, but is particularly critical when the self-enumeration method is used.

60. The approach to be adopted should be reconsidered very early before each census. With increasing literacy levels, some countries that formerly used only the enumerator approach switch to using a mixture of both approaches: for example, self-enumeration in large urban centres and the interview approach elsewhere. Such a combination of enumeration methods may lead to a smaller field staff and hence lower enumeration costs. However, it also requires procedures to obtain responses from those households in

---

<sup>35</sup> This is discussed more fully in Part I of the *Handbook of Population and Housing Censuses*, paragraphs 307-315.



self-enumeration areas who are unable or unwilling to complete the census form on their own. Any new procedure will require thorough testing before it can be accepted.

#### 4. *Forms used to conduct enumeration*

61. Traditionally, a number of developing countries designed their census enumeration form so that it would fit onto one page for most households (a second page being used for large households).<sup>36</sup> This arrangement had the advantage of being simple to use in the field and easy to print. The single-page format also permitted the individual sheets to be bound together so that all the forms for one Enumeration Area (EA) could generally be preserved together in a single packet with an EA cover sheet and a stiff back. This approach is operationally simple and has many practical advantages.

62. However, the one-page format also has a number of undesirable features. For example, the decision to include or exclude specific topics depended, in part, on whether there was sufficient room to include them on a single page. As a result, in some countries, the decision has been to exclude items or topics from the census while in others additional items have been accommodated on a single-page form by enlarging the size of the page to such an extent that the census forms became expensive to produce and difficult for the enumerators to carry and work with. In other countries additional items were accommodated on a single-page questionnaire by so reducing either the length of some questions or the print size that the form was difficult to read or the space for recording responses was insufficient. It is particularly important that sufficient space be provided for recording responses to items, such as industry and occupation, which should not be pre-coded; and a too simplified or shortened question on the activity status, does not elicit an accurate response that captures the intended definitions of the concepts.

63. With economic development and change, countries that in the past had gathered little or no economic information in the census, often give serious consideration to expanding the number of topics to be included in the next census. The increasing education levels experienced by most countries also means that the public and the census staff may have a better appreciation of the need for more information and are better able to cope with more questions. Several options for expanding the census form are possible. If a one-page form was used in the previous census, using the back of this form provides a simple way of making a minimal expansion. Some countries use a booklet for each household with a standard paper size (e.g., A4). This allows the use of larger and clearer fonts and provides more space in general for recording responses as well as facilitating data processing operations.

---

<sup>36</sup> In a number of English-speaking countries the term “questionnaire” has traditionally been used to refer to enumeration forms that contained the verbatim wording of the actual questions to be asked, while the term “schedule” was used to refer to enumeration forms that simply listed the topics being investigated at the tops of the columns. In this *Handbook*, the terms questionnaire, enumeration form, and schedule are used interchangeably.

64. Many countries separate the census enumeration form into two parts: (1) the household roster with basic information on one page and (2) a separate page for each person above a specified minimum age containing the detailed items on economic and other topics. These two parts are usually printed as “household booklets”, with provision for a single household roster and 8 to 15 individual (adult) person forms all stapled together. This is an approach that is similar to that used in many household surveys.

65. A variety of other design issues also arise. For example, a decision will have to be made as to whether the census form should be laid out in “portrait” or “landscape” format. (The terms “portrait” and “landscape”, as used here, have the same meaning that they do in common word processing software.) The “landscape” layout generally has the questions across the top of the page (in columns) and the individuals listed along the side (in rows). This format is best suited for a form containing relatively short questions that need simple pre-coded or numeric answers, e.g., age or number of children ever born. The “portrait” layout commonly has the questions listed along the rows, covering more than one page, and individuals listed across the columns on top of the pages. The portrait layout definitely is better for more complex questions, such as those related to economic topics. For countries that use a booklet for each household or individual, the choice between these two layouts may not be as critical, because space limitations are not so severe. In some cases a mixed methodology has been employed, e.g., “landscape” for the basic household listing and then “portrait” for each adult person.

66. The collection of data on economic characteristics often involves the use of complex concepts that in turn sometimes call for complex skip patterns to ensure that each person is exposed only to those questions that fit their circumstances. These concepts, and the related skip patterns, have often evolved on the basis of experience with labour force surveys. In the context of the census, however, certain simplifications must be introduced to allow the census enumeration to proceed smoothly. Some of the strategies for reducing the complexity of the economic items for the census are discussed in Parts 2 and 3. It is essential that any simplifications introduced to accord with the constraints imposed by the scale of the census be: (a) carefully reviewed by experts in labour force statistics and analysis and (b) thoroughly tested.

67. Countries with two or more languages commonly spoken raise particular problems for census taking. In some cases questionnaires have been printed in different languages. In others text in two or more languages are put on the questionnaires. As previously noted, space is usually a major problem for census questionnaires and the inclusion of two language versions of each question can lead to a very cramped questionnaire.<sup>37</sup> The alternative is, however, to rely on the abilities of the enumerators to translate accurately on-the-spot both the census questions and the replies. This procedure obviously places extreme demands on the enumerators and their training.

---

<sup>37</sup> See also the discussion in *Principles and Recommendations for Population and Housing Censuses*, Revision 2, para. 1.187.

68. Decisions related to the design of the census enumeration forms depend on a wide range of substantive and operational factors so that it is important for each country to review and adequately test its design and layout well before the actual census. In addition to consideration of the census form in light of issues at the data collection phase, data processing considerations also will need to be taken into account. Moreover, census questionnaires need to be attractive looking documents. The population in almost all countries now are accustomed to attractive layout of most things they see. Self-enumeration questionnaires particularly must be attractive, but it also applies to interview schedules. If the enumerator finds the questionnaire to be an attractive well laid out document he/she is much more likely to fill it out neatly and correctly.<sup>38</sup> Coders and data entry staff will function better if the form they are using is well designed. Questions on economic items (or those on any other topic) that are cramped into a small column heading in a truncated form, usually give poor results since it is difficult to fully reflect the intricacies of the relevant concept. In some cases, when the responses are not pre-coded, the space provided for responses is not adequate for recording the complete response.

69. Accordingly census planners should, if at all possible, use or consult professional layout designers. At a minimum, advice could be obtained from professional printers on type styles, sizes and the use of colour, if they are not already doing so. Excessive use of bright colours should be avoided as it can distract from the main objectives. It can also be wearing on the eyes of staff working on the forms particularly data processing staff. However, the use of limited pastel soft colours and shading can make questionnaires attractive and will most likely be filled in better. Similarly, employing different fonts and type sizes for different purposes can make the form easier to use. For example, by consistently assigning differing formats to the various elements (that is, questions, prompts, instructions and pre-coded answers) of the census enumeration form one can greatly facilitate its readability. The cost of professional assistance in layout design is usually well repaid in better quality interviewing/responses. Often also such services are available free or at little cost from other government agencies.

## **F Selection, training and supervision of field staff**

70. The field staff (enumerators, and first and second level field supervisors) are an essential element of a successful census, regardless of whether the interview or self-enumeration method is employed, although the field staff plays a more critical role when the interview method is used. Given the importance of this subject, both the *Principles and Recommendations for Population and Housing Censuses, Revision 2* (see paras. 1.210 - 1.215) and the *Handbook of Population Census Methods, Part I*, (see paras. 140-142 and 379-390) give it considerable attention. [PEMBER COMMENT: UNSD TO ADVISE ON MORE RECENT PUBLICATIONS]

---

<sup>38</sup> Respondents may also be influenced by the appearance of the questionnaire, if they view it as professional looking.

71. The relatively short duration of enumerator training for the census cannot make up for major deficiencies in their basic education. On the other hand, enumerators with relatively advanced levels of formal education often do not make the best census enumerators. The repetitious asking of relatively simple questions can bore persons with very high level of education and it has been known for such persons to start asking their own version of the questions in the, normally mistaken, belief that they are gaining a better insight into people's problems than the designers of the census. Pretests based on the use of staff more qualified than the actual enumerators may give misleading results. Middle-level officers, in close contact with the local community have been generally found to be most effective in asking the economic questions in censuses. Schoolteachers working in their own areas where they know community activities have often been found to be particularly effective.

72. Training of field workers will need to include training for both trainers and for supervisors as well as that for enumerators. Trainers should, if at all possible, have some earlier training or teaching qualifications and experience. The training emphasis for the economic questions must be on asking the questions, understanding the major skip patterns, applying any prompts and probes as necessary and on how to record the answers properly. The content of training on the economic items should leave the census field staff comfortable as to how they are to complete the census questionnaire for the major categories of persons who will be encountered during enumeration: those currently active (wage and salary workers, the self-employed and contributing family workers), those not now economically active but who were so formerly, and those who have never been economically active. It is important that field workers understand the full range of possible replies that can be given to the different questions and the type of information that they are designed to obtain. They do not need to understand the underlying concept and scope of economic activity. Thus enumerator training should give little or no attention to the economic theory or conceptual frameworks that underlie the economic items in the census. With the limited amount of training possible for the census enumerators complex notes or long lectures need to be avoided and the training focus on practical and operation issues and the understanding of the questions that have been formulated.

73. Training materials and examples must be simple and aimed at bringing all enumerators up to a minimum level of competence. Avoid complex examples and problem cases that may entertain the brighter enumerators but will almost certainly confuse most. Unusual cases will always occur and it is important that they are written on blank spaces on the questionnaire with full details for later resolution by supervisors or office staff. As with any other type of statistical sources, uniform treatment of similar situations is a key to good economic statistics, so that good enumerators' manuals and sets of examples covering all major common economic situations are essential. Training examples need to be selected with care so as not to foster gender-based or age-based biases, or other stereotypes related to economic activity. In some cases, somewhat different examples may be prepared for enumerators working in distinctly different areas, for example, urban examples for urban enumerators and rural examples for those in rural areas. However, this approach should be used cautiously since enumerators in rural areas

may encounter persons with urban work experience and urban enumerators, rural work experience. Accordingly, census enumerators, wherever located, will need at least some preparation to deal with the typical situations in both urban and rural areas.

74. Practical mock interviews in groups covering the main situations have proved an effective technique of training, but by far the most effective training is field practice in a real situation with a detailed review on the spot or immediately afterwards. At least one and preferably two practical field sessions should be included in any census-training programme. It is often useful to include in the enumerator training sessions some audio-visual (television, movie, film-strip, or slides) or audio (tape recording) materials that have been centrally prepared. Such material, that illustrate the basic structure of major variants of common interview situations can help to offset the damage arising from a poor trainer. This approach is particularly valuable for helping enumerators obtain a correct grasp of complex items with skips such as those that frequently occur in the economic questions.

## **G Field tests and the pilot census**

75. For the purposes of this *Handbook*, pre-tests are taken to mean any test of census questionnaires or procedures. Many of the tests used may be quite narrow in scope, particularly in the early testing of alternate wording of census questions. Census pre-tests can be based on purposive samples designed to maximize the number of relevant cases in terms of the content or procedures being tested. As census planning proceeds and options are narrowed, tests tend to become larger and involve more diverse settings. However, purposive samples are still adequate.

76. It is essential that all items proposed for use in the census, including the economic questions, be adequately tested.<sup>39</sup> The results of these tests should play an important role in decisions about census content and procedures at every stage of census preparations. The fact that the same questions have been successfully used in other countries, in labour force or other sample surveys in the same country, or even in the previous national census, does not obviate the need for testing in light of current national conditions and using members of the census staff and adopting planned census procedures. Ideally, small separate tests should be held for different segments of the main census questionnaire with at least one test focusing on the economic questions. Pre-testing is generally not expensive if test areas are carefully selected to include many different situations for the population and the test is well designed.

77. It is important that a clear objective be defined for each census test. Too often census tests lack such well-defined objectives so that the results are difficult to assess and the process of developing a reliable and cost-effective census enumeration form is delayed or led astray. Census staff who plan and conduct tests need to be able to set out one or more clear test objective(s) prior to the test and then give a critical evaluation of

---

<sup>39</sup> *Handbook of Population Census Methods, Part I*, United Nations publication, Sales No. E.92.XVII.8. (paras. 238-240).

the results in light of these objectives. Very rarely do economic questions, particularly those for use in a census, work perfectly. There are always problems with persons or circumstances that are difficult to enumerate or classify. The best design will be tuned to function well in the common situations and minimize the errors arising in the rare situations. This sort of fine-tuning of the questions, forms and procedures can only be achieved through field tests.

78. A pilot test is generally taken to mean a major test of both questionnaires and procedures. Often there is only one pilot test that serves as a complete test of all the questionnaires, forms and procedures. The final pilot test should be, in effect, a dress rehearsal of the complete census operations. As such, the final pilot census should involve one or more relatively large administrative units so that census managers have some sense of the logistical problems that arise in a large-scale operation. In theory, no changes in census content or operations should occur between the final pilot and the actual census, although in practice some minor refinements are often introduced. Since the use of untested questions or procedures in a census involves considerable risk, any changes introduced after the final pilot require strong justification and, even then, great restraint should be the guiding principle.

79. Having carried out census tests, it is important that the results of these tests be preserved in written form. This will require a certain amount of discipline by census managers. With numerous demands on the time of those responsible for census preparations, there is a temptation to avoid or delay writing up the results of pre-tests and pilot tests. However, such documentation is a building block for future improvements so that it is always to be prepared and included in the handbook on the census that can be passed onto the census staff who will be working on the next census. Such documentation is an important element of the knowledge base needed to build up experience and expertise in the country. Moreover, the results of census pre-tests and pilots, properly analysed, can also usefully contribute to the census evaluation programme, as will be discussed in Part 6.

## **H Issues relating to questionnaire design for data processing**

80. It is important that data processing staff be involved from the start of census preparations. Their participation in the development of census questionnaires from the earliest stages is critical. Such participation will help to ensure that data processing considerations are adequately taken into account at the design stage. In addition, by allowing the data processing staff to better understand the substantive and logistical priorities and concerns of others involved in census operations and planning, it will minimize the likelihood that subsequent decisions at the processing stage will introduce major problems.

81. Where data processing staff have not been involved at the early stages of census planning, it has some times been necessary to completely redesign the proposed questionnaire on the basis of data processing considerations. (Even first drafts of questionnaires should have some allowance for coding or Optical Mark Reader boxes,

etc.). As with all aspects, the final layout should be a co-operative effort. The data processing requirements should not however overshadow other important census design considerations, such as substantive content, or the clarity and attractiveness of the forms for use in the field. For example, an unattractive form with a confusing skip-pattern and full of boxes with little space for clear questions or recording responses is to be avoided regardless of how easy it might be for data entry. Compromises are usually necessary to produce a questionnaire that functions well for all purposes.

82. *Use of pre-coded response categories:* One approach that facilitates data capture is the provision of pre-coded answer categories in the census schedules and questionnaires, wherever possible. Pre-coded answer categories have many advantages. As well as making processing easier they function very well as a guide to enumerators (or respondents) as to the type of responses needed. However, pre-coded answers do have limitations. They function best when there are only a relatively few relevant answers to a clear, precise question, or for numeric answers. The pre-coded list must cover all possible situations and it is usually best to allow for an “other, please specify...” category which can help capture misunderstandings of the pre-coded alternatives as well as situations that are rare or otherwise not anticipated. Testing should be used to help establish the pre-coded list, as well as identify problems with some of the categories, so that notes or examples can be supplied on the questionnaire or in the enumerators’ manual.

83. Enumerators often use such lists of pre-coded answers as a prompt list to respondents and to do this can be useful but should be planned and not left to the enumerator, e.g., “Did you do any of the following... (list of activities) ... during the (reference period)?” Experience shows that most lists used as prompts have some order bias: i.e., the first items tend to have more answers recorded than corresponding to reality. This bias is most pronounced in attitudinal surveys but can still be a factor for factual census economic questions. It is therefore best to try to get enumerators to read through all categories on the list before asking the respondent to identify a specific activity (or response category). There is no predetermined order in which questions or response categories are to be presented. However, many survey techniques, such as changing the order of the prompt list with different respondents or placing commonly forgotten items early in a list, might be tried in testing before the order of the items in a final pre-coded list is decided upon. It is often surprising what difference a change in order can make to the frequency of responses. Note that the order of categories for publication or in conceptual manuals definitely does not mean that they have to occur in that order on the questionnaire.

84. Pre-coded answers should only be used when the list of possible answers can be reduced to a short list from which an enumerator can easily choose on the basis of the respondent’s reply, and for which the categories are all clearly distinguishable. For this reason, pre-coding should be avoided for “occupation” and “industry.” Not only is it almost impossible to make such lists for these two questions simple and clearly understandable in any language, even with good training for enumerators, but limiting the number of categories for these two variables (as will be required for pre-coding) results in

statistics of very limited usefulness, especially if the pre-defined categories have not been chosen to reflect the economic structure of the country but only the structures of the classifications. Nevertheless, because of the expense involved in coding these two items, some countries have chosen this solution to the coding of occupation and industry, and the response alternatives have sometimes been a list based on the first digit of the respective classifications. However, the unique opportunity the census provides to obtain detailed occupation and industry data should not be missed, as this limits the kinds of basic economic analysis that could be carried out with detailed occupation and industry data.

## **I Data processing and dissemination of census results**

{UNSD COMMENT: THIS SECTION IS VERY GENERAL, MORE SO THAN THE OTHERS. NOTHING IS SAID ABOUT ECONOMIC CHARACTERISTICS EXCEPT IN PARA 85.}

85. During data capture, the information should be edited, including checks for valid ranges for codes and checks for consistency (for example, no answers to the questions on economic activity if the person's age is less than the level shown on the questionnaire).

86. In many census operations there is little coordination between discussions on census content and those related to the processing and dissemination of census results. Although initially both census producers and users are often open to discussions about outputs, they are usually soon pushed aside by discussions of census content under the pressure of deadlines, first for testing questions and procedures and then for printing the census forms. Furthermore, as the census date draws near, senior census staff tend to be increasingly involved with the management of field operations. As a result of these circumstances, many decisions on processing with critical implications for the quality, cost and timing of census outputs may be made solely by the census data processing staff without the benefit of user or subject matter specialist inputs. However, such inputs to decisions on census outputs and certain aspects of the data processing time-table can be invaluable both for the user and the census office. {UNSD COMMENT: FROM JEAN MICHEL DURR: GENERALLY SPEAKING I AM NOT VERY HAPPY WITH THIS REDUNDANCE WITH P&R OR OTHER HB, AND THERE IS A RISK OF CONTRADICTION. I DON'T SEE THE INTEREST OF WRITING SOME BANALITIES ON CENSUS ENUMERATION, TEST PROCESS AND DISSEMINATION. THE HB SHOULD FOCUS ON WHAT IS REALLY LINKED TO ECO CHAR ONLY. PEMBER COMMENT: I APPRECIATE THIS POINT OF VIEW BUT SEE THE NEED TO INCLUDE SOME GENERAL POINTS TO SET A CONTEXT. CH 9 ALSO INCLUDES GENERAL COMMENTS AS WELL AS SPECIFIC POINTS ON ECON ACTIVITY PROCESSING AND TABULATION.}

87. To ensure that such user and subject matter specialist inputs are obtained, it is essential that they be planned for. In view of the tight timetable of operations in the pre-census period, it is important that users and subject matter specialists make time and staff available for such consultations as well as those on census content. The agenda for these



discussions will cover processing issues, such as how to deal with inadequate responses, as well as issues related to such traditional census outputs as:

- (a) the extent and timing of the basic tabulations, including those on economic characteristics, for the nation as a whole, for each province, and for other geo-political aggregates;
- (b) the extent and timing of the basic cross- tabulations of data across subject matter fields and detailed tabulations of topics with very extensive response categories such as detailed occupation and industry data, income distributions or place of birth data; and
- (c) the content and timing of various published census reports and monographs.

88. In addition to the traditional census outputs, there are a range of newer intermediate and final census outputs that should also be covered in the consultation process. Chief among them are:

- (a) various types of data bases developed to store census information to facilitate the production of additional tabulations once the initial round of census tabulations have been completed;
- (b) public use samples of anonymised census micro-data that enable users on their own to carry out special analytical studies, including multivariate analyses; and
- (c) other types of census outputs in machine readable formats and Internet dissemination of census results and/or public use micro-data files.

89. The focus of these consultations should be the specification and clarification of user needs and the provision of advice from a user perspective on trade-offs that may arise in processing and dissemination. It is best if users avoid becoming enmeshed in providing advice on the actual processing work. In a country where, in some highly specialized area, such as occupational coding, the centre of technical expertise lies outside the census office, arrangements may be made to draw on this expertise on a regular basis outside the usual user consultation process. Users need also to be made aware that, at times, their interests in certain kinds of census outputs may be frustrated by lack of resources for census processing or by the need to avoid compromising laws or other restrictions designed to protect the confidentiality of individual responses. The former set of problems can sometimes be resolved by the user, particularly if it is a public agency, helping to underwrite the processing costs involved. Issues related to census confidentiality, which are closely related to the public confidence in the census and indeed most statistical work in the entire country, are not always solved so simply. However, several approaches have been used by different countries to help meet user needs for detailed data while still providing the required protection for individual respondents.<sup>40</sup>

---

<sup>40</sup> United Nations publication, Sales No. E.07.XVII.8, paras. 1.339 and 1.376.

## Decisions related to topics for Measuring Economic Characteristics

### J Selecting topics to be included

90. The *Principles and Recommendations, Revision 2*, presents a list of economic characteristics topics that may be included in a census undertaking. These are: (a) activity status (employment, unemployment, inactive); (b) status in employment; (c) occupation; (d) industry; (e) institutional sector; (f) informal sector; (g) informal employment; (h) type and geographic place of work; (i) working time; (j) income.

91. Most population censuses will only cover some of these topics. For still too many countries the census is the major, if not the only source of information on several important population characteristics. Thus users of statistics and special interest groups often compete to have their topics of interest included in the census. The magnitude of the census exercise requires that the number of questions be limited in order for the enumeration to be completed within a reasonable period of time, even though the costs of finding and contacting all the relevant respondents are usually the main cost element for the census operation.

92. The economic characteristics topics are therefore assessed in competition with other important topics. In many countries, the space allocated for the economic characteristics would permit four or five questions, and the most common topics investigated have tended to be the activity status (employment and unemployment), status in employment, occupation and industry.<sup>41</sup> Given that these topics have been included more frequently than most others, there is more experience and knowledge about how well questions on them have worked. For most of the other characteristics, and in particular institutional sector and informal sector, the limited experience with them in censuses and the requirements with respect to the number, format and content of the relevant questions to ensure a reasonable degree of accuracy and quality of responses, more testing may be needed before it is decided whether it will be possible and worthwhile to include them in the basic census questionnaire.

93. A general principle for selecting among the topics for census exercises is therefore to include only those for which experience and testing have demonstrated that a reasonable degree of accuracy can be assured with the resources available for the census operations. The issue of whether other sources of statistics on the topic exist, and the reliability of their estimates for other than national aggregates will also be relevant. For example, countries with a regular household survey programme may do well to limit the topics covered in the population census and consider collecting data on the other topics through a the labour force survey or other specialised survey such as an informal sector survey; when the users' main concern is to get reliable statistics for national aggregates.

---

<sup>41</sup> For more detailed discussion and review of national practices with respect to inclusion of topics in census, see (United Nations publications, Sales No. E.96.XVII.13) and *Sources and Methods: Labour Statistics. Vol. 5. Total and economically active population, employment and unemployment (population censuses)* Second edition. Geneva: ILO, 1996.

In the context of the population census, another option when such alternative sources are not available could be to leave some of these questions to be investigated on a sample basis; if the short and long forms approach is adopted in the census (see Chapter 4, C.1 above and Part 6, for discussion on the use of sampling in population censuses).

## **K Determining the location and sequence of the economic items on the census form**

94. The position of questions within a questionnaire and on a page should reflect their importance. Other things being equal, the later an item appears in a form or the lower it is on a page, the less attention it will be given and, consequently, the completeness and quality of responses will suffer. In many cases unfavourable positions for individual items cannot be avoided. Some items must appear later in the form than others. However, a variety of actions ranging from the design of the census form (for example, the use of arrows, different type faces, bolding, colour) to training of enumerators and supervisors, and field supervision can be used to help overcome the potential drawback of an unfavourable location on the census form.

95. The usual sequence of census questions generally does not favour the economic questions. There is very often a progressive exclusion or filtering out of the younger population from successive sets of items: whether the census is carried out with a self-enumeration form or by a census enumerator. For example, sex, relationship, and age, are obtained for everyone; information related to education for all but young children; and information on training, fertility, and economic activity and related topics for the population above an even more restricted age cut-off. This sequence is to a certain extent logical but does tend to de-emphasise the economic questions. Following this approach, economic characteristics could be one of the last set of topics in a census questionnaire, unless there are questions on fertility and mortality which are sometimes placed at the end of the questionnaire due to their potentially sensitive nature, and also because fertility questions are often for women only. Questions placed at the end present special challenges especially if respondents find the questionnaire or interview to be too long or the nature of prior questions to be sensitive or personal.

96. The first question or questions of the economic block of questions are almost always designed to identify those who are or (sometimes) have been 'economically active'. The subsequent economic questions on the census questionnaire mostly apply to this population and thus the question(s) to identify them are the key to the whole block of questions on 'economic' topics. For example those who are not currently economically active do not have a current occupation, or industry or employment status, and those who are not employed can not respond to questions related to hours worked or place of work. Thus the positioning of questions or topics should facilitate the use of skip patterns and thereby produce efficiencies, e.g., by reducing the length of the interviews. While different ordering of questions is possible, it is therefore useful to group questions relating to descriptive characteristics of the job separately from questions relating to descriptive characteristics of the establishment. One possible way of ordering the questions is to start with the employment status of the individual, followed with the

selected topics on descriptive characteristics of the job for those that are employed; the descriptive characteristics of the establishment, and other characteristics on the individual (depending on which set of items are to be investigated). It is important that the ordering of questions are pre-tested to ascertain their effectiveness.

97. Following the above principle, it is recommended that the 'occupation' question be asked before the 'place of work' and 'industry' questions and after the 'status in employment' question. The focus in the 'status' question is on the individual's work contract situation, and it is logical to follow it with 'occupation' where the focus is on what the individual does. For 'place of work' and 'industry' the focus then shifts to where the person works and the activity at that establishment. This order assists to prevent confusion between 'occupation' and 'industry'. 'Occupation' may also be considered to be the most important topic, and may therefore be given priority by being placed first. (That the sequence is different in the well-tested questionnaires of the United States, Australia and Canada may, however, indicate that question formulation may be as important as sequencing for good results.)

98. When an income question is asked it is usually put as the last question of the economic block of questions. This is partly because it is logical to ask all details of employment and then to ask about the income from that employment. It is also often done to sequence all adults to this point, not just the 'employed' persons, and ask about total income including investment income, pensions, etc. It would be illogical to sequence those not employed to a question part of the way through the economic block and then out again. It is also often thought that it is best to have a sensitive question last in the block of questions, as some persons may refuse not only this question but also all questions from that point onwards. However, as the last question (often on the questionnaire) the question does face the danger of being forgotten, missed or ignored. For example, in Australia the question on total cash income was placed before the economic block of questions and at the top of a page, presumably, to emphasise the importance given to this question. Countries should test the position and sensitivity of the question in their situation, but it does generally seem logical to have the question at the end of the economic block of questions.

## **L Formulating questions on selected topics**

99. The act of measurement in a census in a face-to-face interview involves two persons, the enumerator and the respondent, with the questionnaire serving as a tool to facilitate the communication between them and to record the outcome. The level of understanding of the questions by the respondent will depend upon the individual's level of literacy, if not education, as well as on how questions are formulated on the questionnaire and communicated by the enumerator. In a census operation, enumerators' level of understanding of concepts and definitions underlying the employment variable cannot be expected to be high enough primarily because of the limited training enumerators undergo, a function of the large number of them that is required to be employed in a census, and the cost of training. Thus they have to be trained to communicate the meaning of the questions rather than the underlying concepts, which

have lead to their formulation. The following are some principles for formulating questions.

100. *Avoid technical terms*: Simplicity is always the key to good questions. Do not use technical, long or complex words. Certainly do not confuse the terms for concepts with the words to be used in the questions. Such words as ‘economically active’, ‘own account’, ‘self employed’, ‘occupation’, ‘industry’ or ‘contributing family worker’ should definitely be avoided in questions or response categories given to respondents as they are very unlikely to know what these are intended to mean in this context. As described more fully in Part 2 with respect to individual items, the questions should be put in language that operationalizes the technical concepts, using words and examples that make sense to a broad range of respondents. Such an approach, while always important, takes on an added significance where the questions will need to be translated to other languages in the field by enumerators. In practice, the translations of some apparently simple economic questions or pre-coded categories have yielded surprising results.<sup>42</sup> All translations should therefore be carefully checked during pre-testing and during the training of enumerators.

101. *Always keep questions short*: If necessary use a second part to the question or a second question. Definitely avoid the use of ‘or’ in questions. Such questions usually lead to the assumption that the enumerator will ask another question to find out which part of the ‘or’ a ‘yes’ answer applies. If follow-up questions are implied for enumerators it is best to put them on the questionnaire to ensure they are asked and asked uniformly. Particularly also avoid negative or vague formulations, such as “You did not do any of these activities in the last seven days.” “What did you do?” or “What did you do when not working?” They can be confusing. Also, qualifying clauses at the beginning or end of the question, and questions with negative formulations are all easily misunderstood, particularly in an oral interview.

102. *Pre-coded response categories should be exhaustive*: Pre-coded responses are commonly used for most economic characteristics; the usual exceptions are occupation and industry. Pre-coded response categories reduce the time and cost of data entry and processing, but require much testing and experimentation to generate the appropriate categories. If only one response category is to be checked as response to a question the categories should also be mutually exclusive and exhaustive. Categories of economic activity status (i.e., worked, had a job but did not work, seeking work, and available for work, etc.) have sometimes been combined with reasons for non-economic activity (such as, schooling, home duties, retirement, etc.). These situations are not all mutually exclusive, and therefore present special problems for the wording of questions (such as, whether more than one category should be checked or respondents are to be directed to focus on the “main” activity). If this style of questioning is adopted special instructions and training will need to be given to enumerators to ensure that the priority rule is correctly applied.

---

<sup>42</sup> See, for example, Studies in Methods, Series F, No. 54, Part I (United Nations publication, Sales No. E.92.XVII.8), para. 227.

103. *Minimize the use of explanatory notes:* Some explanatory notes are almost always necessary in connection with the economic items but they should be kept to a minimum. Only very important points should be put on questionnaires (or self-enumeration forms), otherwise they will clutter the forms and the enumerator (or respondent, in the case of self enumeration) will either ignore them or become confused. Notes on questionnaires that are instructions to the enumerator and those that are prompts to the respondent should be clearly identified as either one or the other. Both can be quite helpful. Differing type styles or sizes or colours are often effectively used to distinguish instructive notes from prompts.

104. *Limit the use of skips in the questionnaire:* A related issue to the sequence of items on the census form is the issue of the nature and number of “skips” in the form. Skips, (that is, instructions to jump from one census item,  $n$ , to another one, not  $n+1$ , depending on the response given to item  $n$  or an earlier item) are used to direct the flow of the census interview whether the census is based on the interview or self-enumeration approach. Although quite complex skip patterns, along with repeated probes for economic activity, are sometimes used effectively in labour force surveys, as a general rule skips and multiple probes have to be held to a minimum in a census, particularly when self-enumeration is used. To the extent that skips are used in the census, they should be as simple as possible, with clear guidance given in the form of arrows, instructions, or both on the correct path to follow. With respect to the economic items, it is difficult to completely avoid the use of skips but every effort should be made to keep them as simple as possible. Common skips associated with the economic items include age-related filters and the exclusion of those from most or all of the subsequent items in the economic block, those for whom no current (or past) economic activity is reported in the census.

105. *Build on past experience:* Changes in the definition of concepts in the census, and the ways in which a census gathers information related to these concepts may lead to major problems of comparability between one census and the next, and with other sources of data on economic characteristics. This does not mean that changes in definitions and methods of data gathering should not be introduced. It does mean, however, that changes and differences, whether introduced to improve data quality or to save costs, should be the result of a careful decision-making process, and steps taken, to the extent possible, in the design of the questionnaire and the tabulations to maintain some degree of comparability, e.g. by making it possible to identify separately those elements that represent the modifications to previous concepts or scopes.

106. In most instances, this approach may well lead to more questions on the Census questionnaire, but the counter balancing reward is that it is possible both to measure change over time since the last data collection and to try out improved methodologies. When using such an approach, however, it is important to recognize that it is likely to be most effective where only a limited number of core issues are treated in this way.<sup>43</sup> For

---

<sup>43</sup> United Nations (1988), *Improving Statistics and Indicators on Women Using Household Surveys*, Sales No. E.88.XVII.II; pages 24-25 and 40-41.

example, given the changes in the production boundary and the implied changes in the definition of economic activity, the building-block approach would suggest keeping any past question(s) on activity status and introducing a question that focuses on those other activities in the household sector which tend to be overlooked by respondents or enumerators. Tables on changes between the censuses should then be designed to show the differences in the measures, by presenting separately statistics based on the old set of questions and those with the extended set. The additional question can however be costly and should be examined in the pre-test, and assessed in relation to other questions on economic characteristics that need to be included.

107. *Be realistic and keep the limitations of the census in perspective:* Decisions to include or exclude a particular (economic) topic in a census, and the degree of refinement that can be attempted in its measurement, should be based on experience and careful testing. First, do not include in a census what has been difficult to measure in a sample household survey with well-trained enumerators and a sound sample design. Second, attempt to obtain only “rounded” or approximate measures of those variables that are measured more precisely by such surveys. That is, do not try to assure all marginal exclusions or inclusions to a particular category if doing so requires complex (sequences of) questions. Conversely, within the limitations of a census, it should be made to yield what a sample survey will not provide, that is, statistics at the lowest levels of administrative, planning, or political unit and on relevant social groups of a country’s population, for major economic topics (e.g., number of carpenters or nurses in a town or a local government area, by age group and sex and type of training). The census must not fail in this unique function for major economic characteristics. Attempts to provide statistics with high levels of conceptual precision for some of the more difficult to measure topics (e.g., underemployment, employment in the informal sector, or many of the national accounts concepts and distinctions) can have negative effects on the overall quality of the basic census results. Such conceptually precise estimates should be obtained through specialized surveys. When required also for small areas and groups the ‘cruder’ census results may provide a good starting point for targeting such surveys, as discussed in Part 6.





**PART 2: MEASUREMENT OF ECONOMIC  
CHARACTERISTICS IN A POPULATION  
CENSUS**



## Chapter 3: Measurement Frameworks

### A Understanding the concept of economic activity

108. The single most important issue for obtaining reliable information on economic topics in a population census<sup>44</sup> is accurately determining the activity status of the population—who is economically active and who is not? Statistics on the size and composition of these two groups are important for formulating almost all economic and social policies and related planning and research. Moreover, most economic characteristics in the census are obtained only for the economically active, and therefore the item(s) used to identify the activity status are the key to the whole block of economic items. Thus omissions and misclassifications of persons can have a major impact on a broad range of census data on the socio-economic situation of the population.

109. The conceptual basis for distinguishing the economically active in the population is currently derived from the 1982 Thirteenth International Conference of Labour Statisticians: Resolution concerning statistics of the economically active population, employment, unemployment and underemployment.<sup>45</sup> This basis is tied to and fully compatible with the United Nations System of National Accounts (SNA),<sup>46</sup> —the standards define the “economically active population,” as comprising “all persons of either sex who furnish the supply of labour for the production of economic goods and services as defined by the United Nations System of National Accounts and Balances”.<sup>47</sup>

110. Over the years, the definition of the economically active population has evolved from “Persons who perform work for pay or profit”,<sup>48</sup> in 1954, to “All persons of either sex who furnish the supply of labour for the production of economic goods and services”<sup>49</sup> in 1966. The 1983 revision enlarged the concept even further and thereby inextricably linked it to the SNA. As the scope of production within the SNA has changed, so has the scope of the concept of economic activity. The current (2007) situation is reflected in an article published in the ILO’s *Bulletin of Labour Statistics 2007-1*.<sup>50</sup> A basic theme running through these revisions has been to obtain a more precise representation of the economic activity of the population, both in their contribution to production by way of their labour as well as the creation of the product itself. However, changes in concepts have not automatically been translated into

---

<sup>44</sup> See *Principles and Recommendations, Revision 2*, paras. 2.231 - 2.349, for the presentation on economic characteristics items for the census.

<sup>45</sup> See in particular, Hussmanns and others (1990).

<sup>46</sup> See in particular, Commission of the European Communities, and others (1993), *System of National Accounts 1993* (United Nations publication, Sales No. E.94.XVII.4), Chapters VI and XVII.

<sup>47</sup> International Labour Office (1983).  
<http://www.ilo.org/public/english/bureau/stat/download/res/ecacpop.pdf>

<sup>48</sup> International Labour Office (1955).

<sup>49</sup> United Nations (1967).

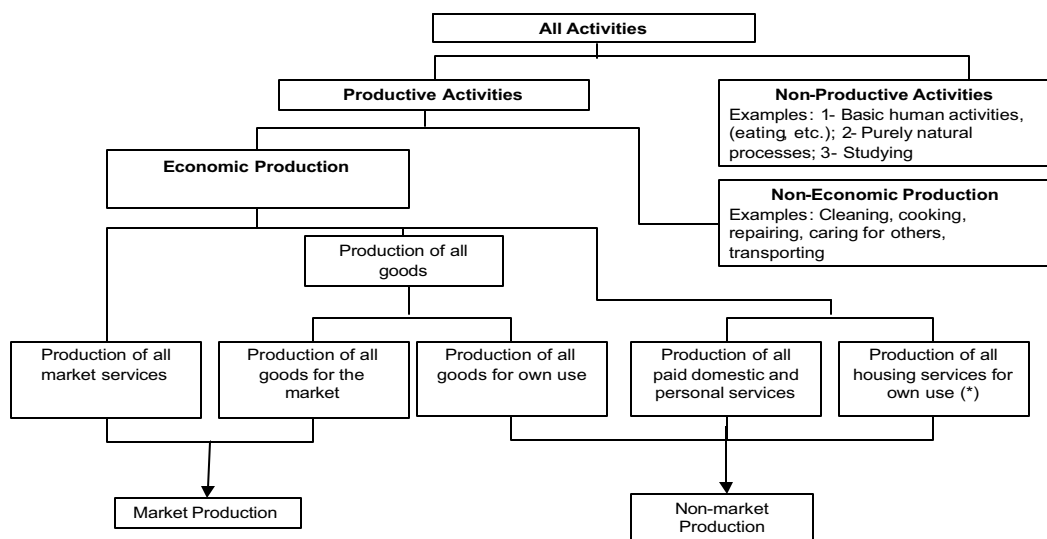
<sup>50</sup> International Labour Office (2007)

procedures that could improve the measurement or coverage of the economically active population correspondingly.

## B The scope of production in the system of national accounts

111. The 1993 SNA includes all production of goods for own use within its production boundary, as goods can be switched between market and non-market use even after they have been produced, but it excludes all<sup>51</sup> production of services for own final consumption within households, such as production of personal and domestic services by members of households for their own final consumption.<sup>52</sup> (See Diagram 1) For consistency between the 1983 revision of economic activity concepts and classification and the SNA, the types of production that constitute production boundary of the SNA, should be clearly delineated. According to the 1993 SNA, the types of production such as are presented in Table 1, which are produced for households' own consumption, are considered within the production boundary.

**DIAGRAM 1: ECONOMIC PRODUCTION AS DEFINED IN CURRENT SNA**



(\*) No labour input

112. Fetching water and collecting firewood, are examples of two specific activities, which some experts tend to object to being regarded as economic activities. The contention is that these activities, like many others routinely carried out by women homemakers in particular, such as processing of agricultural product for household consumption, and dress making for members of the household, would inflate the numbers

<sup>51</sup> All, that is, except the provision of housing services (which does not involve any labour input).

<sup>52</sup> Commission of the European Communities, and others (1993), Pages 4-5.

of economically active in a manner which make the resulting statistics less useful for certain types of analyses or economic policies. There is also concern about maintaining comparability between measures of economic activity that are consistent with the extended production boundary of the 1993 SNA and earlier data sets. A building block approach is recommended, whereby consistency with the past is protected by maintaining the possibility of providing statistics according to the older formulation of the concept through the use of questions that make this possible while also introducing new ones that would make it possible to cover the more comprehensive set of activities included with the newer concept.

**TABLE 1: EXAMPLES OF NON-MARKET HOUSEHOLD PRODUCTION ACTIVITIES**

<b>Primary production</b>	<b>Processing primary products</b>	<b>Other production of goods</b>	<b>Fixed capital formation</b>
Growing or gathering field crops, fruit and vegetables	Threshing and milling grain	Storing crops	Construction of dwellings
Producing eggs, milk and food	Making butter, ghee and cheese	Dressmaking and tailoring	Construction of farm buildings
Hunting animals and birds	Slaughtering livestock	Handicrafts made from non-primary products	Building boats and canoes
Catching fish, crabs and shellfish	Curing hides and skins		Clearing land for cultivation
Cutting firewood and building poles	Preserving meat and skins		Major repair and maintenance of dwelling and farm buildings
Collecting thatching and weaving materials	Making beer, wine and spirits		
Burning charcoal	Crushing oil seeds		
Mining salt	Weaving baskets and mats		
Cutting peat	Making clay pots and plates		
Fetching water	Weaving textiles		
Collecting firewood	Making furniture		

Adapted from Hussmanns and others, 1990; Table 1, p.18.

113. The treatment of volunteers (those working without pay or profit outside the household) has also been the subject of discussion. (See for example, United Nations (2003): *Handbook on Non-Profit Institutions in the System of National Accounts*, Series: F, No.91, Sales 03.XVII.9.) Under the SNA-93, the production of goods by volunteers, and the production of services by volunteers for market enterprises or for non-profit enterprises operating in the market (such as some hospitals and schools) is within the production boundary and therefore treated as economic activity. The production of services by volunteers for households or for non-profit enterprises not operating in the market is not treated as economic activity. Therefore volunteer work caring for neighbours' children or providing services for a religious group or a sporting body would not be considered as economic activity, whereas voluntary work to construct a bus shelter for the community or to provide services to a private (fee-paying) school would be

treated as economic activity.

114. For easy reference Table 2 lays out a summary of the major activities that are regarded as economic and non-economic. The emphasis in the table is on activities commonly found in many developing countries, but obviously not every situation would be considered to be important in all of them, nor can a general table address every situation relevant for any individual country. The recommended treatments for some specific activities shown in the table may be inappropriate in a given census because of operational considerations. On the other hand, rather than exclude some activities because this would suit certain user interests, it would be preferable to cover those activities and identify them separately, and exclude them from tabulations tailored specifically to those interests. The census planners should try to draw up a similar table specifically for common activities in their country, to guide questionnaire development for the economic questions. Discussions on the final list of economic activities for the census should include National Accounts experts as well as other major users of census results on the population's economic characteristics. In these discussions it is important to remember (i) that inclusion of these activities as "economic" only will lead to an increase in the number of 'employed' persons if the persons concerned do not carry out any other economic activities; and (ii) that the census questionnaire can be designed to ensure that these persons can be identified separately and therefore be excluded for certain tabulations if this corresponds better to the needs of certain users.

115. The proper identification of activities that are economic is vital for all subsequent data that are collected with respect to economic activities. It is therefore important to pay particular attention in the design of the census form, the training of interviewers or the instructions given on the form, to ensure that the questions translating the above notions are correctly understood by the respondents. This is particularly important in economies in which a substantial proportion of the economically active population are not engaged in full-time, full-year employment but in part-time employment, casual work, home-based work, work paid for largely in kind, unpaid family work and production of goods for own consumption. In these instances, the use of an activity list could be very instrumental.

**TABLE 2: ILLUSTRATIVE SET OF ECONOMIC AND NON-ECONOMIC ACTIVITIES**

Economic Activities	Not Economic Activities
<p><input type="checkbox"/> Working in wage jobs</p> <ul style="list-style-type: none"> <li>- Full time or part time</li> <li>- Permanent or temporary</li> <li>- Casual or piecework</li> </ul> <p>Including <u>paid</u> child minding and other <u>paid</u> domestic work</p> <p>Can be paid in cash or kind (e.g., food or accommodation)</p> <p><input type="checkbox"/> Having businesses activities</p> <ul style="list-style-type: none"> <li>- Large or small, agricultural or non-agricultural</li> <li>- Small shop/Kiosk/Street Stall</li> <li>- Preparation/Selling Juice, Soft Drinks</li> <li>- Taxi Operator</li> <li>- Shoe cleaning/Sewing Business</li> </ul> <p><input type="checkbox"/> Any activities on own or family farms for the purpose of production for sale or for home consumption including:</p> <ul style="list-style-type: none"> <li>- Planting crops</li> <li>- Harvesting crops</li> <li>- Keeping birds and other pests off crops</li> <li>- Weeding</li> </ul> <p>Transport of goods from the fields for storage or for sale</p> <p>Fetching water and firewood for domestic use</p> <p><input type="checkbox"/> Fishing, collecting shells or seaweed for sale or home consumption</p> <p><input type="checkbox"/> Processing of agricultural or natural products for sale <u>or</u> home consumption</p> <ul style="list-style-type: none"> <li>- Mats, hats from natural or grown fibres</li> <li>- Furniture from natural timber</li> <li>- Butter/cheese, etc., from milk</li> <li>- Oil from Oil seeds/fruit</li> <li>- Preparation of charcoal</li> </ul> <p><input type="checkbox"/> House or farm building/construction</p> <ul style="list-style-type: none"> <li>- Fence/enclosure/storage construction</li> <li>- Road/irrigation construction</li> <li>- House construction/additions</li> </ul>	<p><input type="checkbox"/> Studying full time</p> <p><input type="checkbox"/> All types of housework including:</p> <ul style="list-style-type: none"> <li>- Unpaid child minding-own/other children</li> <li>- Education/training of own children at home</li> </ul> <p>- House cleaning and decorating exclusively for own household</p> <p><input type="checkbox"/> Other types:</p> <ul style="list-style-type: none"> <li>- Persons doing nothing due to <ul style="list-style-type: none"> <li>- Retirement</li> <li>- Sickness</li> <li>- Disability</li> </ul> </li> <li>- Living off investment, rental or pension income (no current activity to earn it)</li> </ul>
<p>Note: A similar list of activities for each country should be made early in census preparations to assist in questionnaire design and during testing, and to be used later in manuals and training.</p>	

## **C Correspondence between economic activity and production within the SNA production boundary**

116. The scope of economic activity and the production boundary as delineated by the SNA, though closely related, are not necessarily equal, in the way they are being measured. A component of production that is difficult to fully capture, is that of the household sector comprising, among others, subsistence agriculture, informal sector<sup>53</sup> and processing of primary products (see also list in Table 1). Although the ILO economic activity concepts seek in principle to cover all kinds of economic activity in the household sector, including part-time paid and unpaid work of children in small family businesses or farms, as well as part-time paid work of students after school or during school holidays, there are practical limits to this ideal. For example, in providing guidance on household production, the *SNA* indicates “When the amount of a good produced within households is believed to be quantitatively important in relation to the total supply of that good in a country, its production should be recorded. Otherwise, it is not worthwhile to estimate it in practice.”<sup>54</sup> Even more strictly, the Resolution of the 13<sup>th</sup> ICLS states that “Persons engaged in the production of economic goods and services for own and household consumption should be considered as in self-employment if such production comprises an important contribution to the total consumption of the household.”<sup>55</sup>

117. Similar considerations apply in deciding upon the value of attempting to measure the economic activity of children and early adolescents. Concerns with child labour, may however mean that their activities would need to be recorded even if the criterion of their contribution being “quantitatively important in relation to the supply of that good” in the country is not satisfied.

118. Difficulties encountered in translating the operational definitions into questions that would elicit the desired responses, and in applying interviewing and other data collection techniques that would reach hard-to-measure categories of people, are also reasons why the estimates of economic activity may deviate from the production boundary of the SNA.

## **D Decisions related to measuring economic characteristics**

119. The operational definition of the economically active population limits the coverage of production to those performed:

- (i.) within a specified time frame or reference period,
- (ii.) by individuals meeting an age-specific criterion, and

---

<sup>53</sup> See paras. 368-371 for the definition of informal sector.

<sup>54</sup> Commission of the European Communities, 1993; para. 6.25. See also, *Principles and Recommendations, Revision 2*, para. 2.238.

<sup>55</sup> Sub-paragraph 9(b)(6) of International Labour Office (1983).

See also <http://www.ilo.org/public/english/bureau/stat/download/res/ecacpop.pdf>



120. *The reference period*: As explained in the *Principles and Recommendations, Revision 2*, there are two types of reference periods recommended. The short reference period of a week or a day, (though one month has been used in some cases), is referred to as representing the ‘current economic activity’, and corresponds to the labour force concept; and the long reference period of one year, is referred to as the ‘usual economic activity’. The former is the focus of the discussion in Chapter 4 and issues related to measuring “usual activity status” are presented in Chapter 5.

121. *Age cut-offs*: For practical considerations, such as to speed-up operations, to limit costs, and to build rapport with respondents, age limits are frequently used to exclude segments of the population that are either too young or too old to be engaged in any economic activity or are at ages at which the level of economic activity is deemed to be very low. For example, in many countries it makes little sense to ask about the economic activity of infants or very young children in any population census. It is thus usual to specify at least a lower age limit, below which answers are not required even though some level of economic activity may take place among this group. An *upper age limit*, that is the maximum age beyond which economic characteristics should not be collected, is also used in some countries. The *Principles and Recommendations, Revision 2* however cautions against the use of a maximum age,<sup>56</sup> since valuable information especially about the activity of the elderly or the ageing population, will then be lost. Indeed in many countries, the previous practice of a recognized retirement age for employees is now being abandoned.

## **1      *Setting the reference period***

### *1.1.    Period corresponding to current activity (labour force)*

122. The currently economically active population, referred to as the labour force concept, has been the basis of employment and unemployment measures in both censuses and household surveys for many years. The reference period for the current activity concept has in most national censuses ranged from a time frame as short as one day to as long as one month.

123. Some countries have been experimenting with one day (e.g., the day before the enumeration day) as a reference period in surveys, but it is not known if any country has used this in a census. Such a measure would be the most “current” measure possible, and would minimise problems of recall by the respondent. An Indian study<sup>57</sup> has shown that as expected employment tends to be lower and unemployment higher with a one-day reference period compared to the one-week period (using the priority rule discussed later, see Chapter 4, para 2). In a population with high incidence of intermittent work the chances of some individuals having one day without work is higher than their having a complete week without work. Temporary absences from work for the one day can cause

---

<sup>56</sup> *Principles and Recommendations, Revision 2*, para. 2.241.

<sup>57</sup> Paper by P. Visaria for ESCAP/ILO Seminar on Employment and Unemployment Statistics, January 1989, Bangkok.

coverage problems. Most countries have thought one day as a reference period is too short and there is always a danger that something unusual may happen on that day if not nationally at least in some regions of the country. There are also problems of avoiding Sundays or other traditional rest days for the country concerned, when the one-day reference period is adopted.

124. A one-month reference period has been tried in some countries. However recall lapse is likely to be present for some respondents and there will be more problems when formulating questions about different activities during the month than during a shorter period. The usual approach is to determine a main activity if there are more than one employment activity. If both employment and unemployment occur during the month, one also has to decide whether to use the approach of “any work” during the month, or the approach of the “main activity” during the month that is also used in the usual activity approach.

125. On balance, the seven-day reference period, which has long been used in most countries for both censuses and surveys, seems the most useful one. The recommended reference period for “employment” is the seven days prior to the Census enumeration day. This is obviously the closest short period. However, some countries prefer to use the calendar week prior to enumeration day as they find respondents relate better to a fixed calendar period. When using the calendar week approach, the exact starting and ending day should be given, e.g., as from Monday to Sunday last week. The use of either the previous seven days or the calendar week should be included in any pre-testing for the census. It is particularly useful to apply the same period for the census as for the national labour force surveys, for comparability purposes (within reasonable limits). However, if it is deemed necessary for whatever reason, to use another reference period this should definitely be subjected to pre-testing and an evaluation made of the effect of the change on the comparability with the results from previous censuses, and the results of the evaluation published.

126. The measures are termed ‘current’ measures because of the short reference period close to the census date. Yet it is unlikely that they or any other census measures will give very “current” indicators of the employment situation at the time when census results are released, because of the usual long period for census processing. As noted earlier, household surveys are used to obtain more “current”, i.e. up-to-date, estimates for national aggregates and large groups. However, given the recent developments in the use of technology in census operations, such as in the use of scanning, in place of manual data entry, the census processing time has been drastically reduced in many instances, in some cases to within a year of the completion of census enumeration. It is also true that there are usually not major changes in the basic structure of the economically active population while waiting for census results, and that they consequently are very valuable indicators of the economic structure, particularly for small geographic areas.

*1.2. Period corresponding to usual activity*

127. The usual activity concept defines participation in economic activity and relevant

characteristics, i.e., of the job, and of the establishment or employer, over a long reference period. The usual activity approach is in principle to represent the typical or usual employment patterns over the long period. The commonly used period is a year although in some instances a period such as the agricultural season is used. The latter is particularly useful when associating employment statistics with production statistics. When a year is used, it is necessary to decide if this should be the previous twelve months or a calendar year. If census day is not early in the year, the latter period has a greater risk of recall and memory losses. It has therefore been found best to use the twelve months ending at the month prior to census day.

### *1.3. The choice between current and usual activity reference periods*

128. A relatively short reference period, close to the time of the interview, minimises memory lapses for the respondents. It is also easier to measure the characteristics of economic activity for the short reference period, as there are usually fewer changes during that period than during a longer one. The disadvantage is that it can only represent the activity situation during that current period, and this may not be representative or typical of the annual or average activity patterns if the country or region has striking seasonal patterns, of employment and unemployment, and economic activities. This could be particularly significant in developing countries in which, due to serious logistical challenges during the rainy season, census day is chosen during the dry season. Many users might prefer to get a reflection of these “typical” situations and differences between regions and groups, especially when such statistics are only available at long intervals. However, since the data collected are measured retrospectively, the measures are subject to more recall errors than the data on currently active population.<sup>58</sup>

129. Also, in order to truly represent the typical employment situations over the entire reference period, all economic activities including the main, secondary and so on activities have to be covered. Over a long period as one year, the labour market and therefore the population could be subjected to changes – individuals’ activity statuses may change between active and not active; employed and not employed. Thus, in order to obtain accurate information about usual activity, the approach would require several questions about these activities, way beyond the minimum set that would be required for the current activity approach. For example, at the minimum the amount of time spent (measured either in weeks or months) in the employment, unemployment and non-economic activity states should be asked of the respondents, otherwise respondents would have to be counted on to accurately determine what their main situation or activity was during the reference period. Collecting these data would prove a major challenge for the design of the census instrument and increase the cost of the census.

---

<sup>58</sup> Hussmanns and others (1990); p. 47.

130. There could also be the possibility of changes in occupations and other such characteristics of the job or jobs over a long period such as one year. These would invoke special problems, such as the greater effort in collecting information on time spent in the possibly greater number of different jobs in order to objectively determine the main job for the period.

131. The short reference period, usually seven days, has been used in most countries for censuses and surveys for a considerable period and so there is accumulated experience in handling the limited difficulties that occur. The relative precision of the current measure and general ease of interpretation for enumerators and respondents makes it strongly preferable to any usual measure for economic characteristics questions. Census planners therefore need to weigh these advantages of a current measure against the more difficult-to-measure item of usual activity, which however, can reflect the activities over the different seasons of a twelve-month period. It is strongly recommended that any country starting to use usual activity for all their economic characteristic questions conduct a very extensive pre-testing programme to identify the difficulties and to build experience in applying the concepts.

132. As comparability with previous results is also of major value for census data, those considering adopting the usual activity approach for the first time, should keep the current approach and carefully assess the extent to which questions applied to the usual approach should be included, and *vice versa*.

133. Using both approaches has obvious advantages but also some drawbacks: An obvious disadvantage is that every additional question adds to the costs of the census and that the two approaches will generate different numbers. There are in addition a number of difficult technical problems in using the longer reference period of the usual measures that are discussed below.

134. Statistics on the economically active population should always be published with the reference period prominently noted so that users are fully aware of this: e.g., "Employed population in the week of..., 2008". Although this has often not been done in the past, it is important if statistics in countries with pronounced seasonal variations in certain types of economic activity in certain regions, e.g., agriculture or tourism, are to be understood in their proper contexts.

## **2. *Setting age limits***

135. Items under the different topics covered by the census will usually have different age cut-offs so that one of the first matters to be decided is the appropriate age limit(s) to apply for each population-specific item in the census. The rationale for setting an age cut-off is to define a more realistic target population to whom the economic activity questions are to be addressed. It is usual to specify at least a lower age limit, below which answers are not required for these questions. With regard to the economic characteristics, the same age cut-off should be applied to all items dealing with these topics. For many countries, the cut-off used in asking questions on economic activity in a population

census has been age 14 or 15 years. In countries where a large proportion of the population enter and complete secondary school, there is a tendency to adopt age 16 years as the lower limit. There are also those countries where lower cut-offs, below age 10 years, sometimes as low as age 6 years, are used<sup>59</sup>, for a variety of reasons including inadequate education funding and consequent downward trend in school attendance, or because of an increased interest in measuring the extent and types of child labour. This is especially so in developing countries.

136. The main factors to be considered in determining an appropriate lower age-limit in asking economic questions in a census are:

- (a) the age at which significant numbers of the population begin to engage in or seek to be engaged in economic activity, often a function of income levels and school enrolment rates, and the distribution of school leaving ages; and
- (b) the objectives and intended uses of the data.

137. In some countries the consideration of this lower age limit is confused with the minimum legal age for paid employment. Such considerations should be avoided, as in principle there should be no connection between the minimum age for these census questions and the legal age for paid work. 'Economic activity' questions should ideally record any type of economic activity including part-time work and unpaid work in small family businesses or farms, and there has been considerable interest in the extent and type of work activities of children in recent years. In most countries some school-age children engage in part-time, paid or unpaid employment, for example part-time work after school or during school holidays. In rural areas children often help in agriculture at all ages whether going to school or not, and these activities are not usually covered by any minimum age legislation. In a population census, the aim should be to record the whole range of economic activities and to set lower age limit that is too high, could lead to the loss of valuable information.

138. In some countries an upper age limit is also used, e.g., 65 years. This procedure is even more likely to lead to the loss of valuable information. In almost every country some people work either full or part-time, well beyond any legal or conventional retirement age. This is particularly so in rural societies but is not uncommon in urban areas, even if older persons often work fewer hours than when they were younger. Many Governments are now considering the need to encourage some form of work at higher ages, as active older people are often healthier than before. In almost all countries the population is living longer, which suggests that a greater number and range of economic activities may be missed if an upper age limit is imposed for such questions.

139. The choice of a lower age limit to use for the economic activity items should be explicitly reviewed before each census, although usually the same cut-off will be used as in the previous census. The trade-off that arises when deciding whether to raise or lower

---

<sup>59</sup> Studies in Methods, Series F, No. 54, Part IV (United Nations publication, Sales No. E.96.XVII.13), Table 4.

the age cut-off used previously, or to continue using the same cut-off, involves balancing the benefits gained by asking a broader group of people the economic items against the associated field and processing costs and added proxy respondent and enumerator burden. The lower the age cut-off the larger the population base to be interviewed (or of whom the information is to be sought). This increases the interview time and therefore the average cost per household. The lower age cut-off will generally give more realistic coverage of the economically active population in countries where there is either low enrolment in or high dropout from primary school. However, if there are legal repercussions for children's non-attendance in primary school the quality of responses may be poor, and therefore the benefits in coverage from increasing the population base will be diminished and will be outweighed by the cost. Another factor that could affect the outcome of the interview is the difficulty of eliciting accurate responses with the questions that have typically been used in censuses, e.g., when the activities involved are informal in nature (without contracts, irregular hours of work, and so on). Therefore, if there is an objective to measure child labour, either the questions should be modified for the whole census population or it should be done on a sample basis in a census long form or left for another more specialized survey specific to this objective, such as time-use survey. Such modifications may be warranted also for a number of other reasons, some of which are discussed further below.

140. In countries where child labour has become a sensitive issue, the overall effect on the quality of the census results obtained when using a low age limit should be taken into account and studied as part of the pre-census testing programme. In general, a low age limit for the economic questions will impose some extra burden on respondents, usually proxy respondents who have to report on several other members of the household, and on the data processing while providing valuable, otherwise unknown information. More relevant information about the form and circumstances of child labour and other "difficult-to-measure" production activities may however be obtained from specially designed surveys, including special purpose child labour surveys, labour force surveys and perhaps time-use surveys. However, a discussion of the merits and modalities of such methods is beyond the scope of this *Handbook*.<sup>60</sup> In general, it would be an advantage if the use of a new age cut-off could be introduced first in ongoing sample surveys and then be introduced in the census subsequently on the basis of the experiences gained. The cost-effectiveness of reducing the age limit should in any case be assessed at the pre-testing stage.

141. It is generally preferable that age-based cut-offs, such as raising the lower age limit or setting an upper age limit, are introduced during tabulation and analysis, if there is a need, for particular policy, research, or administrative needs; for example, to produce tables comparable with those from previous censuses or to estimate the impact on certain totals of the use of a higher minimum age for asking the economic items.

---

<sup>60</sup> United Nations (2004) *Guide to Producing Statistics on Time Use: Measuring Paid and Unpaid work*, Sales No. E.04.XVII.7.

## **E Organization of Chapters: Rest of Part 2 & Part 3**

### ***1 Relevance of past experiences***

142. There is often much to benefit from building on past experience, whether of our own or of others, as opposed to adopting new questionnaire formulation or design from scratch. The starting point for revising old questions or introducing new ones is therefore the lessons from previous censuses (and sometimes even surveys). The next six chapters present examples of a range of questions that have been used in national censuses, for each of the items on economic characteristics, as proposed in the *Principles and Recommendations, Revision 2*. Strengths and limitations of the questions are identified, with some suggestions on improving their effectiveness.

143. The examples have largely been drawn from both developed and developing regions around the world, except when no national experience was available to draw from, in which case untested examples captioned as “hypothetical questions”, are proposed. For each item, the discussions also address issues related to measurement, field-testing, and data processing. The starting point for the material presented in these chapters are the broad topics related to the economic characteristics of the population proposed in the United Nations *Principles and Recommendations for Population and Housing Censuses, Revision 2* and are discussed under the following chapters - current activity status, usual activity status, descriptive characteristics of the job, descriptive characteristics of the establishment and person totals.

144. In general, topics/items for which national experience abounds include: employment and unemployment; status in employment; occupation, industry; place of work; and working time. National experience on the measurement of income and usual activity status in population censuses<sup>61</sup> {PEMBER COMMENT: UNSD PLEASE UPDATE FIRST SENTENCE OF THIS FOOTNOTE} is somewhat limited, and even more so on measurement of institutional sector, informal sector and informal employment.

145. The aim in the presentation of examples has been to select questions that illustrate particular points. For all topics on which limited national experience is available, as well as for any introduction of new or modified questions, extensive programme of pre-tests and pilots needs to be considered; even if there is some household survey experience in gathering corresponding information.

146. Several census questionnaires illustrating a range of national practices in formulating questions about economic characteristics in a population census are

---

<sup>61</sup> See Studies in Methods, Series F, No. 54, Part IV (United Nations publication, Sales No. E.96.XVII.13), tables 3 and 4 for a summary of national census experience covering 1965-1984. See also *Sources and Methods: Labour Statistics. Vol. 5. Total and economically active population, employment and unemployment (population censuses)* Third edition. Geneva: International Labour Office, undated (about 2005). See “Sources and Methods” under <http://laborsta.ilo.org>.

presented as an annex. Elements of these examples are referred to in the text wherever possible and relevant.

## **2      *Format of the presentation***

147. The chapters in these Parts focus on preparing questions for inclusion on the census form. The principles for developing questions are generally the same regardless of the characteristic of focus. However, each characteristic has some very specific considerations that influence the level of difficulty in obtaining accurate measures. Therefore some issues are given more extensive discussions than others, depending on the characteristic.

148. The characteristics are discussed in five chapters corresponding to the unit to which the information sought pertains. These are as follows:

### Rest of Part 2

Chapter 4: Current economic activity status

- Employment
- Unemployment

Chapter 5: Usual Economic Activity

### Part 3

Chapter 6: Descriptive characteristics for the main job

- Status in employment
- Occupation
- Type and geographic location of place of work
- Whether in informal employment

Chapter 7: Descriptive characteristics of the establishment of the main job

- Industry
- Institutional sector
- Informal sector

Chapter 8: Person totals

- Working time
- Income

149. At the beginning of the discussions of each broad topic, an introduction is presented on the general issues pertaining to the topic, following which the issues specific



to each of the characteristic are discussed in turn. The presentation follows the structure shown below and is explained in the remainder of this section:

- The operational definition
- Requisites for applying the definition
- Review of questions used in national censuses
- Suggestions for enhancing effectiveness of questions
- Issues of measurement

(a) *The operational definition*

150. The internationally recommended definition for the characteristic is presented. This is generally a restatement of the definition presented in the *Principles and Recommendations, Revision 2*. Where necessary additional information is drawn from other sources, such as the ILO manual on surveys of the economically active population.<sup>62</sup>

(b) *Requisites for applying the definition*

151. The specific elements of the definition and their implications for translating the concept into questions are discussed. These may include setting of criteria, identifying key words that require special attention, and peculiarities that could affect the measurement of the characteristic.

(c) *Review of questions used in national censuses*

152. This section presents examples of questions that have been used in previous national population censuses and reviews them with respect to how well they could be expected to measure the characteristic. The names of the countries have however been suppressed since the examples are used purely for illustrative purposes. Wherever possible several solutions are presented and discussed. The discussion explains aspects of the question that either enhance or limit the potential for achieving the intended result. (Note that the discussion of national examples will assume that the English version of the question(s) represents an accurate translation from the original. However, that may not necessarily be correct, and the translation may have introduced some limitations that the original formulation did not have.)

(d) *Suggestions for enhancing effectiveness of questions*

153. In light of the review in the previous sub-section, suggestions may be made for improving the formulation. When reasonably good and well-tested questions are not available in some cases because of limited experience on the measurement of the

---

<sup>62</sup> Husmanns, Ralf and others, 1990. *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*, Geneva: International Labour Office.

characteristic in a census the suggestion is in the form of a proposed question, often untested in a population census exercise, but based on experiences from a sample survey. It has been necessary to include such “hypothetical” questions in the manual to address either topics that were included in the *Principles and Recommendations, Revision 2*, for the first time, such as informal employment, or emerging issues on which there is no prior census experience. This approach is used to make more options available to users of the *Handbook* and to encourage experimentation and pre-testing in census pilots.

(e) *Issues of measurement*

154. To complement the suggestions given for the formulation of relevant questions, methods for ensuring greater success in fielding these questions are also discussed. Complementary tools such as prompt cards, areas to be emphasized in accompanying interviewers’ and supervisors’ manuals, training and supervision, etc., are also discussed, where they are deemed particularly critical.

## Chapter 4: The Currently Economically Active Population (The Labour Force)

### A The Labour Force Framework

155. The labour force comprises all persons who during the reference period were either employed or unemployed as described below. Diagram 2 below illustrates the relationship between the total census population (that is, the total population net of any groups excluded from the scope of the census, such as foreign diplomats), the labour force, the employed population, the unemployed population and the population not in the labour force. Thus the questions in the census questionnaire for determining the labour force should be directed at identifying these categories in a way that is clear, unambiguous and mutually exclusive.

156. In order to achieve the last of these some priority rules of assignment are required. The *basic* rules for the economic activity questions are that, within the reference period:-

- (a) Economic activities take precedence over non-economic activities; and
- (b) Within economic activities the status of being employed takes precedence over the status of being unemployed.

Thus priority is assigned to participation over non-participation and to employment over unemployment.

157. This means that in the measurement of economic characteristics a great deal of effort needs to be put into making the definition operational and adopting methods that will correctly classify those that are economically active. Provided the set criteria are met, a person should be classified as having been economically active during the reference period, whether or not they were also engaged in non-economic activities such as studying or home duties. It should be emphasized that school attendance is to be recorded in a census independently of other characteristics and the same principle is to be applied to the enumeration of economically active population provided the age of the individual falls within the applicable age limits. Furthermore, if a person either worked for the set minimum number of hours or had a job (even though s/he did not work) s/he should be classified as employed even if they were seeking another type of employment, for whatever reason.

158. Students and all other persons, who are within the ages for reporting economic activity, should be subject to exactly the same economic questions as others. School attendance or attendance at other training institutions, should be recorded in the education and training block of questions on the census questionnaire. It is important to remember that "students" can be employed either part time or full time, particularly during vacations. In some countries, they form a significant proportion of the employed population. Similarly students can be actively seeking work and be currently available for work that is either full time or part time. Sometimes they are studying only because

there is no work available, or they are seeking part-time work to help pay for their studies. Some users may not want full-time students in, e.g., tabulations of the unemployed, in which case the responses to the education questions can be used to screen out full time students during tabulations.

159. It should be noted that as the labour force is comprised of the employed and the unemployed, the approach to its measurement is to identify each of these two components. The population not currently economically active is then determined as the residual of the total population after separating out the population that is identified as neither 'employed' nor 'unemployed' in the reference period. The rest of the chapter follows this ordering in its presentation by first discussing the measurement of the employed population, then that of the unemployed population and finally the population not currently economically active.

## **B The Employed Population**

### ***1 The operational definition***

160. An employed member of the economically active population is broadly speaking a person who during the reference period does any activity to produce goods or services of the type that falls within the economic production boundary defined by the SNA, or who is temporarily absent from an activity of this type. The "employed population" comprise all persons above a specified age who, during a short reference period of either one week or one day (*Principles and Recommendations, Revision 2*, para. 2.227):

- (a) performed some work for pay, profit or family gain, in cash or in kind; or
- (b) were temporarily absent from a job in which they had already worked and to which they maintained a formal attachment or from a self-employment activity such as a business enterprise, a farm or a service undertaking.

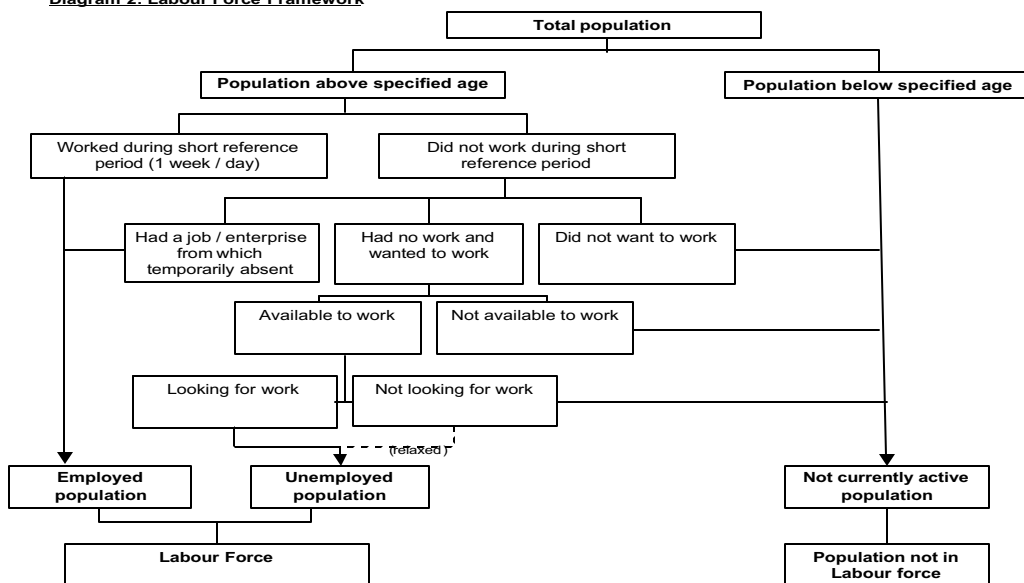
161. The operational definition requires the minimum number of hours that an individual should have worked during the reference period, to be considered employed. A one-hour criterion is often used to ensure greater correspondence between total employment and measured production, and that all types of employment, in particular irregular employment, such as stand-by work, casual labour, short-term work, etc., are covered.<sup>63</sup> Other criteria that have been used are (a) the majority criterion—working most of the time during the reference period; or (b) the one day criterion—working at least one day during the reference period; and (c) a specified number of hours higher than one.

## **DIAGRAM 2: LABOUR FORCE FRAMEWORK**

---

<sup>63</sup> Hussmanns and others (1990). page 71.

Diagram 2: Labour Force Framework



## 2 Requisites for applying the definition

162. A person is taken as employed if s/he does any work at all in the reference period of one week. Any work is commonly taken as lasting for a minimum of one hour during the reference period. Although some are concerned about this use of one hour, it is essential for (i) giving employment priority over all other activities; (ii) consistency between employment statistics and national accounts production statistics; and (iii) capturing part-time work as well as marginal forms of employment. Moreover, it has generally been found that only a small number of persons are working very few hours (e.g., less than five).

163. The aim is always to include as ‘employed’ those persons who are temporarily absent from their jobs. The following persons who were temporarily not at work during the reference period should be included as employed: paid employees who continue to have a “formal job attachment” and persons with an enterprise which continues to exist. However these distinctions are not very clear-cut. The details have been the subject of much discussion<sup>64</sup> and fine distinction has been made, particularly in developed countries. In developing countries and particularly for censuses, trying to draw these fine distinctions could be confusing and it is suggested that only the following rules be applied when defining those who are “temporarily absent” from an employment:

- (a) Paid employees who are:

<sup>64</sup> See <http://www.ilo.org/public/english/bureau/stat/download/guidelines/exleave.pdf> for the relevant international guidelines.

- (i.) Absent on paid leave (recreation/annual/sick/maternity leave) or paid training
  - (ii.) Absent from work (with or without pay) because of industrial disputes, reduction in economic activities or temporary disruption of work due to machinery breakdowns, shortage of materials etc., but continue to have a formal relationship to their job or feel that they definitely have a job to go back to.
- (b) Persons with an enterprise who respond that they are temporarily absent from work but that their enterprise continues in their absence, i.e., because other members of the family are continuing to work on the family farm or business, orders are still being received or in some other way which indicate that the activity is still operational.

164. For persons engaged in seasonal activities and contributing family workers, the correct treatment of the concept of ‘temporary absence’ is as follows:

- (a) during the off-season, seasonal employees, especially in agriculture, should not be included as employed unless the employer continues to pay most of their wages or salary;
- (b) self-employed persons engaged in seasonal activities should not be included as employed during the off-season unless their enterprises continue to exist and their owners continue to work in them;
- (c) contributing family workers do not have an enterprise and so cannot be included as employed when temporarily absent from work.

165. In general a common sense understanding of temporary absence has usually managed to cover most cases successfully. Detailed discussions of unusual cases have often confused interviewers and produced excessive numbers of persons being ‘temporarily absent’. This situation should be carefully examined in pre-tests and re-interviews to ensure their proper recording.

### 3. *Special cases: identification of the employed*

166. There are numerous categories of persons in all countries who present special measurement issues (for example, persons engaged in several different types of activities during the reference period, persons who have a job but are temporarily “laid-off” or persons seeking employment for the first time). Some categories of employment present special problems of identification. People in these groups frequently may not, without prompting of one kind or another, report themselves or be reported as having undertaken an activity which will lead them to be classified as being “employed” or “having a job”. The following are some of the major groups that without probing or special care in formulating the questions will tend to be under- enumerated.

- (a) Persons in Part Time or Casual Work – Part-time work, casual work, piece work or similar activities are often under-reported in many censuses for several reasons, including misconception that the government is only interested in recording full-time jobs for the census, suspicions about the purpose of the question particularly when no taxes are being paid on income. When such work are performed by persons who are not the principal “breadwinner”, they are not considered to be “real” jobs.
- (b) Persons with small businesses, particularly informal businesses – Small-scale informal activities such as street selling, shoe cleaning, beer brewing and the wide range of informal businesses common in some countries, are often not regarded as real jobs or real work and are not systematically recorded.
- (c) Unpaid work in a family business (non agricultural) – Persons, particularly women, working without pay in family shops, small family manufacturing activities, or similar family-based enterprises, are one of the most difficult groups to record as employed in censuses (and even in sample surveys), since these activities are seen to be part of family responsibilities or obligations rather than as ‘economic’.
- (d) All types of agricultural work, not only small scale farms – there is usually little problem in obtaining information about employment of those working on commercial or formal registered farms either as owners or as regular employees. However, work on a small common village farm, on a peasant farm, or as a casual farm labourer is often more difficult to identify in a census; a deliberate effort would need to be made to improve the coverage of these activities in a census, such as including a module for screening farming households. (See Chapter 15, for discussion on using population census to develop agricultural frames.) Moreover, some farm work, particularly subsistence farming activities carried out by women, is not always regarded as a real job or real work by respondents.
- (e) Economic activities not yet widely acknowledged: The following are examples of activities that are considered as economic because they contribute to production as conceptualized by the SNA. They are however yet to be accepted by many of the planners of censuses and are recognized least of all by the respondents and even enumerators.
  - (i.) Agricultural work for home consumption – The biggest problem, besides unpaid work in family businesses in many countries is agricultural work mainly for own consumption. In most of these countries, this work is very significant and usually done by women who do not do any other type of economic activity, and who generally will report themselves as homemakers unless the questions or prompts used probe specifically for these types of activities.
  - (ii.) Processing primary products for home consumption – These activities include making mats, pots, furniture, processing milk for butter or oil seeds for oil, weaving textiles, mainly for home use. The SNA states that these activities should be included within the

- production boundary if they represent a significant proportion of the total production of the relevant item for that country. When households sell part of the products, the activities are sometimes recorded in special surveys, but they will usually be missed.
- (iii.) House and farm building work – Construction work, not simple repairs or maintenance, which adds to the household’s capital is regarded as economic work by the SNA and the persons engaged are to be recorded as “employed” when doing this work. These activities include fence making, construction of animal enclosures (e.g., for cattle, pigs or poultry), which are sometimes agricultural off-season activities, road building or irrigation ditch making, as well as home building or major additions to the housing structure.
  - (iv.) Voluntary production of goods and voluntary production of services for market-oriented enterprises – As explained in paragraph 6 of chapter 3, such activities should be classified as economic but may be easily overlooked. In developing countries, voluntary “self-help” community work to construct schools, bus shelters, bridges etc are common but may not be reported since there is no payment for the work undertaken. Even in countries with more advanced economies, the efforts by volunteers in fee-paying hospital and school may be overlooked. (Recall however, that volunteers producing services for households and non-profit institutions should not be included in this group.)<sup>65</sup>
  - (v.) Collecting firewood and fetching water – These two activities have only recently (i.e., since 1993) been included within the boundary of the SNA and are therefore economic activities, even though there is some reluctance to accept them as such. In many villages and some urban areas fetching water is a daily activity of women and often children. Apart from the difficulty in perceiving these activities as economic, the effect of including them would in some countries make almost all women economically active and employed; with corresponding reduction in estimates of the not economically active population and in many cases unemployment estimates. This can be worrying for those wanting to use the census results as a basis for formulating employment policies, if those who are classified as ‘employed’ solely on this basis cannot be separately identified. However, to do so will be very important for a better understanding of the economic and social circumstances of particular groups and localities.

---

<sup>65</sup> See chapter 3, para 111.



#### 4. *Employed: Review of questions used in national censuses*

167. Questions aiming to establish activity status, i.e., whether a person should be classified as employed, unemployed or not economically active, have been formulated in a variety of ways. In some cases the questions have addressed each status separately, in others the statuses have been combined as alternative responses to one question. Focusing on the employed, the examples presented here are classified under four headings:

- (a) One simple question used to determine if an individual was employed during the reference period (examples A.1, A.2 and A.3);
- (b) Two questions used, one to determine if an individual did any work during the reference period; and the second to determine if those who were not at work had a job or business to which they would return (examples A.3 and A.5);
- (c) One composite question that includes elements to determine two or more economic characteristics of the individual (example A.6);
- (d) Two or more questions, one of which is a basic question on activity status (or employed status), and any additional ones to cover specific categories of workers who are not easy to enumerate (examples A.6 to A.10).

##### *Example A.1*

A.1.1	<input type="checkbox"/>	A.1.1 Do you have a job for which you worked 4 hours or more in the past week (or would have worked if you had not been away due to vacation, illness, pregnancy, or a labor dispute, etc.)? (Fill in space <small>A.1.1</small> )
		1. yes    → GO ON TO QUESTION XX
		2. no     → GO ON TO QUESTION Y.Y

168. The single question in Example A.1 is posed to determine whether the individual was employed during the reference period. The question incorporates the main element of the concept – worked for a minimum time or attached to a job during the reference period. The key word “job” can be understood in a narrower sense than is intended (i.e., as a formal, standard, or regular wage employment), and it may not be understood that it as well should include informal-type employment, unpaid work in a family business, or in some cases own-account work. The two parts of the question, the main and the parenthetical, would be better separated to simplify and make it easier for respondents to know to what they have answered “yes” or “no”. There is also some ambiguity introduced with “had not been away”, which should be qualified.

169. It has been noted that there is a potential problem with questions such as “Did you have any job or business of any kind in the last.”, because they are likely to produce an immediate answer ‘No’ from respondents who are farmers, petty traders or similar. They seem to have heard the word ‘job’ and assumed the question was about wage jobs. Such

questions alone without probes or follow-up questions will almost always produce an underestimate of the employed population. The word "job" should definitely be avoided in the basic question on this topic. Some of the major problem groups for identification of the 'employed' were presented in chapter two. Persons engaged only in these kinds of activities may not regard themselves as 'employed' or as 'having a job' without some prompting by the questionnaire and/or the enumerator.

**Example A.2**

*35 - 45 questions to be asked  
for person 15 years old and  
over about work in the last  
week before census  
(March 26 - April 1)*

**A.2.1 Are you employed?**

**yes** → Y.Y. question

**no** → A.2.2 question

**A.2.2 Are you unemployed?**

**yes** → *Do you have  
a status of  
unemployed?*

**yes**

**no**

**no** → X.X question

170. The question in Example A.2 is simple and can easily be responded to with a "yes" or "no" by those familiar with the concept of employed. This group may exclude people who judge their economic activity to be insignificant or of little value. Pertinent information is sacrificed when the question is this short. It would be more effective for the reference period to be stated in the question. This would ensure more consistent use or mention of it. The way the questionnaire has been designed, the reference period is captioned with instructions to the enumerator and may not be systematically used when the question is introduced.

**Example A.3**

<p><b>A.3.1 Last week, did the person have a full-time or part-time job of any kind?</b></p> <ul style="list-style-type: none"> <li>• Mark one box only.</li> <li>• A 'job' means any type of work including casual, temporary, or part-time work, if it was for one hour or more.</li> <li>• See page 12 of the Census Guide for more information.</li> <li>• Remember to mark box like this: <input checked="" type="checkbox"/></li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Yes, worked for payment or profit</li> <li><input type="checkbox"/> Yes, but absent on holidays, on paid leave, on strike, or temporarily stood down</li> <li><input type="checkbox"/> Yes, unpaid work in a family business ► <b>Go to XX</b></li> <li><input type="checkbox"/> Yes, other unpaid work ► <b>Go to YY</b></li> <li><input type="checkbox"/> No, did not have a job ► <b>Go to YY</b></li> </ul>	<ul style="list-style-type: none"> <li><input type="checkbox"/> Yes, worked for payment or profit</li> <li><input type="checkbox"/> Yes, but absent on holidays, on paid leave, on strike, or temporarily stood down</li> <li><input type="checkbox"/> Yes, unpaid work in a family business ► <b>Go to XX</b></li> <li><input type="checkbox"/> Yes, other unpaid work ► <b>Go to YY</b></li> <li><input type="checkbox"/> No, did not have a job ► <b>Go to YY</b></li> </ul>
---	---	---

171. In Example A.3, respondents are given clear instructions and guidance in completing the questionnaire. The inclusion of the fourth response category (“other unpaid work”) will ensure that volunteers are given an obvious choice, but the question sequencing will treat these a not employed even if they are working as volunteers for profit-making enterprises.<sup>66</sup>

**Example A.4**

<p><b>A.4.1 a) Did ... work at any time LAST WEEK, either full time or part time?</b>          Work includes part-time work such as delivering papers, or helping without pay in a family business or farm. It also includes active duty in the Armed Forces. Work does NOT include own housework, school work or volunteer work. Subsistence activity includes fishing, growing crops, etc. NOT primarily for commercial purposes.</p> <p><i>Read each category and mark (X) the ONE box that applies.</i></p> <p>1 <input type="checkbox"/> Yes, worked full time or part time at a job or business AND did NO subsistence activity</p> <p>2 <input type="checkbox"/> Yes, worked full time or part time at a job or business AND did subsistence activity</p> <p>3 <input type="checkbox"/> Yes, did subsistence activity only )</p> <p>4 <input type="checkbox"/> No (did not work OR did only own ) <b>SKIP TO xx</b>          housework, school work, or )          volunteer work) )</p> <p><b>A.4.2 Was ... on layoff from a job or business LAST WEEK?</b>  <i>If “No”, ask “Was ... temporarily absent or on vacation from a job or business last week?”</i></p> <p>1 <input type="checkbox"/> Yes, on layoff</p> <p>2 <input type="checkbox"/> Yes, on vacation, temporary illness, labour dispute, etc.</p> <p>3 <input type="checkbox"/> No</p>
--

172. The basic economic activity question in A.4.1 (a), is to determine if any work was done during the reference period.

- (a) When taken together with the clarifications given on what are included or excluded from work, the question appears to be too long.
- (b) For an interview-based approach, it should be made clear to whom the explanatory notes are directed (i.e., to the interviewer or to the respondent).

173. The notes serve as prompts on the major problem categories, and the response categories can also be useful prompts. However the question tends to become rather complex with these features, and in particular with the combining of both basic economic activity, subsistence activity and other activities being asked in one question.

- (a) The use of the term “subsistence” further complicates the question, as the scope of subsistence activities may not be fully grasped.

174. One way to improve upon this question is to give separate categories for

---

<sup>66</sup> A number of other countries have specific separate questions to identify different categories of volunteer work, but no country sees to treat them as employed, even if working for profit-making enterprises or producing goods.

subsistence and non-subsistence work (either in the response categories to be checked in addition to the basic ones, or in a separate question).

175. The second question, A.4.2, ascertains if there is an attachment to a job, and complements the identification of the “employed”. Without this question the count based solely on A.4.2 (that is, only those did any work in the last week) would incompletely measure the employed, by excluding (most of) those temporarily absent. Both questions have the same reference period. This question is followed with one on unemployment, so that those responding “no” to both questions are asked questions to establish their unemployment status.

**Example A.5**

<p>A.5.1 Last week, how many hours did you work (not including volunteer work, housework, maintenance or repairs for your own family)?</p> <p><i>Include as work:</i></p> <ul style="list-style-type: none"><li>• <i>working without pay in a family farm or business (e.g., assisting in seeding, doing accounts);</i></li><li>• <i>working in your own business, farm or professional practice, alone or in partnership</i></li><li>• <i>working for wages, salary, tips or commissions.</i></li></ul> <p>Number of hours (to the nearest hour) OR</p> <p><input type="radio"/> None</p> <p>A.5.2 Last week were you on temporary lay-off or absent from your job or business</p> <p><i>Mark one circle only</i></p> <p><input type="radio"/> No</p> <p><input type="radio"/> Yes, on temporary lay-off from a job to which expect to return</p> <p><input type="radio"/> Yes, on vacation, ill, on strike, or locked out or absent for other reasons</p>
---

176. This pair of questions from the interview schedule is slightly different from the pair in example A.4. The basic question used to determine if an individual did any work during the reference period is directed at measuring “number of hours worked”. (Issues specific to this characteristic will be discussed below in the appropriate section.) In both examples, the need for clarifying what comprises work is demonstrated. The exclusions are inserted as part of the question, and examples of what should be considered “work” is presented as prompts.

- (a) The ordering of the major types of work presents first the ones least likely to be reported, and the standard type at the end of the list (to establish some degree of importance to these other activities). While the list may be adequate for more advanced labour markets, further clarification with activities specific to less organized labour markets will be required.

177. Combining a basic activity question with the collection of ‘hours worked’ data may have influenced data quality on one of these topics. These kinds of approaches need

to be tested if there is interest in trying them out.

**Example A.6**

Type of activity during past 30 days	
<p><b>A.6.1</b> During the past 30 days did ... work for cash?</p> <p>1 Yes for someone else } GO TO</p> <p>2 Yes for self } <b>xx</b></p> <p>3 No GO TO <b>A.6.2</b></p>	<p><b>A.6.2</b> Then what did ... do during the past 30 days?</p> <p>1 Family business</p> <p>2 Work at Lands/Farms/Cattle Post</p> <hr/> <p>3 Actively seeking work</p> <p>4 Housework</p> <p>5 Student</p> <p>6 Retiree</p> <p>(Other specify)</p>

178. The questions used here are rather short because they were used with a “landscape layout” questionnaire format, common in some parts of Africa – allocating one page (generally) to collecting details for all family members.

179. The combination of the basic activity status question with other topics, in this case status in employment, as presented in A.6.1 should be avoided if at all possible. In this particular case, it is unclear what the effect of having just two categories —working for someone else or for self —will be on the responses of those in unpaid work, producer cooperatives, home-based workers, etc.,. The answer categories for question A.6.2 are useful in that they prompt for unpaid work particularly in agriculture and family businesses. However, strong interviewing skills would be required to ensure that the priority rule is strictly applied, if only one category is to be checked for each eligible person. The principles for good interviewing would suggest that all possible responses should first be read out, before a response is accepted. On the other hand, unless it is clear to the respondent that work done in the first two categories takes precedence over the non-economic activity categories, it is likely that many of the responses to this question for women will be “housework”, especially for those for whom this is a time consuming activity. Since these are not mutually exclusive categories, this problem can be addressed by having interviewers check all the applicable categories, as in example A.8 below.

180. With the landscape layout there is always very little room for prompts or explanation on the questionnaire, and in this case the manual and training would have to have thoroughly covered the meaning of ‘Work’ as well as the other categories of ‘Work’. Prompt cards could be used to provide further clarification with a list of country-specific types of activities. There is no question to determine whether those who did no “work” during the period and were not looking for work, were attached to a job or business (and therefore economically active).

181. It has been common to determine the economically active population through a question with response alternatives that allow the derivation of particular categories, e.g., for ‘status in employment’ the categories for ‘status in employment’ would be used as

prompts. For example “In the last seven days did you do any work as (a) a wage earner/paid employee, or (b) an own account worker with employees, or ... etc.” To do this may save space on the questionnaire, but it would almost certainly help to underestimate the economically active population. This problem was accentuated by the use of technical terms such as ‘own account worker’ that are difficult for non-statisticians to understand correctly. If at all possible the initial question on employment should be left to stand alone, and questions for other topics should be asked separately. The census training of enumerators must emphasize the basic importance of the key question(s). The importance of doing so has been shown in quality investigations of both censuses and surveys.

**Example A.7**

<p>24 ¿Durante la semana pasada, trabajó aunque sea por pocas horas?</p> <p>Si <input type="checkbox"/></p> <p>No <input type="checkbox"/></p> <p>Ignorado <input type="checkbox"/></p>	<p>A.7.1 During last week, did you work even for a few hours?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Don't know</p>
<p>25 ¿Hizo algo en su casa para afuera o ayudó a alguien en un negocio, chacra o trabajo?</p> <p>Si <input type="checkbox"/></p> <p>No <input type="checkbox"/></p> <p>Ignorado <input type="checkbox"/></p>	<p>A.7.2 Did you do something in the house for the outside or did you help someone in a business or job?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Don't know</p>
<p>26 Tenía trabajo pero, ¿estuvo de licencia por enfermedad, vacaciones, etc.?</p> <p>Si <input type="checkbox"/></p> <p>No <input type="checkbox"/></p> <p>Ignorado <input type="checkbox"/></p>	<p>A.7.3 Did you have a job, but were you off from work due to illness, vacations, etc.?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> Don't know</p>

182. The first of the three questions asked in the Example A.7, is a simple formulation that includes both the reference period and the idea of “any” work. The second serves as a prompt for some types of work that tend to be overlooked by the respondent or the interviewer. These categories could be extended to cover specific activities that should be included as economic. The third draws in those who hold jobs even though they were off.

**Example A.8**

<p><b>A.8.1</b> Last week were you doing any of the following? (<i>X all the boxes that apply</i>)</p> <ul style="list-style-type: none"><li><input type="radio"/> In paid work</li><li><input type="radio"/> Retired from paid work</li><li><input type="radio"/> In full time education or a student</li><li><input type="radio"/> Working on a government sponsored training scheme</li><li><input type="radio"/> Looking after family or home</li><li><input type="radio"/> Permanently sick or disabled</li><li><input type="radio"/> Temporarily sick or injured</li><li><input type="radio"/> Looking for work and available to start within two weeks</li><li><input type="radio"/> Looking for work but not available to start within two weeks</li><li><input type="radio"/> Waiting to start a new job</li><li><input type="radio"/> Doing unpaid work in own/family business</li><li><input type="radio"/> Doing unpaid voluntary work</li></ul> <p><b>A.8.2</b> Are you currently in paid work?</p> <ul style="list-style-type: none"><li><input type="radio"/> Yes</li><li><input type="radio"/> No</li></ul>
---

183. The question in Example A.8 was included on the “short form” and intended for self-enumeration. The first question lists broad categories in the range of activities that could define an individual’s status in the reference week. It combines paid work (at the top of the list) and unpaid work (economic and non-economic, at the end of the list). Unemployment categories and non-economic activity categories are also included. Though not all the categories may be suitable for other national situations, the technique of providing multiple answers is an interesting approach that census planners in other countries might consider. The identification of the status of employed, unemployed and not currently economically active is then done during data processing making use of the priority rules of the labour force framework. The second question both validates the response to the first category in question A.8.1, and for some, might also show attachment to a job (but only if it is paid). In countries where unpaid work in family enterprise or self-employment is highly prevalent, this question may not yield the desired results for these groups of workers unless it can be modified accordingly.

184. The first question (A.9.1) in the pair presents many of the categories presented in the previous set of questions (example A.8). They are limited to the main economically and non-economically active categories. In this case only one category is to be checked. Presumably once a category is checked the interviewer is instructed to proceed to the next applicable question. Technically these can be seen as representing eight questions (though fewer than eight are posed as the sequence may stop after the first “yes” in the priority order). Question A.9.2 is more elaborate than the corresponding example A.7. This question serves to prompt those who do not report themselves as employed (i.e., did not get registered as either “worked” or “... had a job”) and are not incapable of working due to a disability, about non-standard economic activities (income-generating and unpaid economic activities, such as in a family enterprise, helper on a farm). These are important categories for capturing women’s work more completely, for they include

activities that many do but who would be reported as homemakers in the first question. Other categories that will be missed are those whose only economic activities are in the production of goods for own household consumption, fetching water and collecting firewood, and those engaged in volunteer work for profit making institutions.

**Example A.9**

A.9.1 CONDICIÓN DE ACTIVIDAD	A.9.2 VERIFICACIÓN DE ACTIVIDAD
<p>¿La semana pasada (NOMBRE)</p> <p>Marque con "X" un solo círculo</p> <p>trabajó? ..... <input type="radio"/> 1</p> <p>no trabajó, pero sí tenía trabajo? ..... <input type="radio"/> 2 } Pase a X.X</p> <p>buscó trabajo? ..... <input type="radio"/> 3</p> <p>¿Es estudiante? ..... <input type="radio"/> 4</p> <p>¿Se dedica a los quehaceres de su hogar? .... <input type="radio"/> 5</p> <p>¿Es jubilado (a) o pensionado (a)? ..... <input type="radio"/> 6</p> <p>¿Está incapacitado (a) permanentemente para trabajar? ..... <input type="radio"/> 7 } Pase a Y.Y</p> <p>¿Realiza otras actividades? ..... <input type="radio"/> 8</p>	<p>Además de (CONDICIÓN DE 16), ¿la semana pasada (NOMBRE)</p> <p>Marque con "X" un solo círculo</p> <p>ayudó a trabajar en un negocio familiar? ..... <input type="radio"/> 1</p> <p>vendió algunos productos (ropa, cosméticos u otros)? <input type="radio"/> 2</p> <p>hizo algún producto para vender (alimentos, artesanías u otro)? ..... <input type="radio"/> 3</p> <p>ayudó a trabajar en actividades agrícolas o en la cría de animales? ..... <input type="radio"/> 4</p> <p>a cambio de un pago realizó otro tipo de trabajo (lavó, planchó o cosió)? ..... <input type="radio"/> 5</p> <p>¿No trabaja? ..... <input type="radio"/> 6 } Pase a Y.Y</p>

A.9.1 Activity status		A.9.2 Verification of activity	
Last week (Name) Mark with "X" one circle only		Besides (the situation in A.8.1), last week (Name) Mark with "X" one circle only	
Worked?	<input type="radio"/> 1 } Go to	Helped work in a family business?	<input type="radio"/> 1
Did not work, but had a job?	<input type="radio"/> 2 } xx	Sold products (clothes, cosmetics, or others)?	<input type="radio"/> 2
Looked for work?	<input type="radio"/> 3	Made something to sell (like food, handicrafts, and others)?	<input type="radio"/> 3
Is a student?	<input type="radio"/> 4	Did you help work in agricultural activities or in animal husbandry	<input type="radio"/> 4
Is a homemaker?	<input type="radio"/> 5	In exchange for payment did other type of work (washed, ironed or sewed clothes)?	<input type="radio"/> 5
Is retired or receiving pension?	<input type="radio"/> 6	Did not work?	<input type="radio"/> 6 Go to yy
Is incapacitated permanently to work?	<input type="radio"/> 7 Go to yy		
Did other activities?	<input type="radio"/> 8		



**Example A.10 (used in a long form on a 10 per cent sample)**

A.10.1 *Did you/did ..... work for at least one hour during the first week of September 2001 ?*

Yes ( Go to V.V)     No     Not Stated

---

A.10.2 *Did you/did .....do anything like farming, buying and selling during the first week of September 2001 ?*

Yes ( Go to V.V)     No     Not Stated

---

A.10.3 *Did you/did ..... do any type of odd job or hustling during the first week of September 2001 ?*

Yes ( Go to V.V)     No     Not Stated

---

A.10.4 *What were you/was ..... doing for most of the time during the first week of September 2001 ? (READ CATEGORIES)*

Working in Agriculture or any other business without pay

With job not working ( Go to W.W )

Seeking first job ( Go to Y.Y )

Seeking a job which was not the first ( Go to Q X.X

Did not seek work but wanted work and was available ( Go to X.X )

Student ( Go to Q Z.Z

Did Home Duties ( Go to Q Z.Z

Retired did not work ( Go to Q Z.Z

Disabled unable to work ( Go to Q Z.Z )

Not interested in work ( Go to Q Z.Z

Other ( Go to Z.Z

Not Stated

185. The first three questions in the set of four used to determine activity status are simple questions presenting one clear idea, requiring a “yes” or “no” response. The first, (A.10.1) is the basic and commonly used question, especially when the landscape format is adopted for the questionnaire. It includes the main elements of the concept of employment, minimum hours worked and a reference period specifying the date. The

next two questions focus on different types of activities that are likely to be poorly reported. In the third question (A.10.3) a local term is used, which is bound to increase the recognition and understanding of what is entailed in the term “work” used in A.10.1. The fourth question, (A.10.4) in the set is similar to the first questions in examples A.8 and A.9, and combines non-economic activities with unemployment categories and the residual from the economic activities dealt with in the first three questions. Like the questions in example A.9, this set covers all the major activities except production of goods for own household consumption, fetching water and collecting firewood, and it seems likely that persons who do only such activities will be excluded from the employed population.

186. There are four questions in this set (A.11.1 to A.11.4), each dealing with a specific aspect of the work that qualifies as economic activity. The first is generally that of ownership of land, and though a positive response to this question does not necessarily mean that the individual is employed, the response forms a basis for probing further if at the end of the series of questions the individual is to be classified as non-economically active. The second (A.11.2), is the usual standard question that many countries use to determine if the individual worked during the reference week. The third question (A.11.3), probes for specific activities that tend to be missed (in many other countries as well, especially where there is a large informal sector or subsistence activities). It is important to note that in A.11.2 and A.11.3 both the reference week and the principle of “any work” are reflected, as respondent cannot be expected to assume that the two questions are connected. The fourth question (A.11.4), represents the group of people who had a job or enterprise but were not at work during the reference week.

**Example A.11**

<p>A.11.1 Do you/Does . . . own any land for farming?  <input type="radio"/> Yes    <input type="radio"/> No    <input type="radio"/> Don't know/Not stated</p> <p>A.11.2 Did you/. . . do any work for pay, profit, or family gain for at least one hour during the week ending May 12th? This includes helping in a family business or farm, street vending, or work at home.  <input type="radio"/> Yes (SKIP TO Q 11.5)    <input type="radio"/> Don't know/Not stated  <input type="radio"/> No</p> <p>A.11.3 Did you/. . . engage in any economic activity on the following list, for pay, profit or family gain, for at least one hour, during the week ending May 12th?  <input type="radio"/> Yes (SKIP TO Q 11.5)    <input type="radio"/> Don't know/Not stated  <input type="radio"/> No</p> <p><b>INTERVIEWER: READ THE FOLLOWING LIST</b></p> <table border="0"> <tr> <td>Babysitting</td> <td>Nurse's Aid for pay</td> </tr> <tr> <td>Laundry, ironing for pay</td> <td>Subsistence farming</td> </tr> <tr> <td>Cleaning yard/cutting yard</td> <td></td> </tr> </table>	Babysitting	Nurse's Aid for pay	Laundry, ironing for pay	Subsistence farming	Cleaning yard/cutting yard		<p>A.11.3 list continued.</p> <table border="0"> <tr> <td>Selling food or snacks at market/bus-stop/school</td> <td>Sewing for pay</td> </tr> <tr> <td>Bicycle/cart deliveries</td> <td>Cleaning of offices</td> </tr> <tr> <td>Selling food from home</td> <td>Car washing</td> </tr> <tr> <td>Selling sweets from home (fudge, etc.)</td> <td>Taxi-driver</td> </tr> <tr> <td></td> <td>Lottery vendors</td> </tr> <tr> <td></td> <td>Any other similar activity</td> </tr> </table> <p>A.11.4 Did you/. . . have a job during that week, from which you were temporarily absent?  <input type="radio"/> Yes  <input type="radio"/> No (SKIP TO Q 1 . . . X.X)  <input type="radio"/> Don't know/Not stated (SKIP TO Q 1 . . . X.X)</p> <p>A.11.5 How many hours did you/. . . work during that week?  <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>  <input type="radio"/> Don't know/Not stated  <input type="text"/> (SKIP TO Q . . . Y.Y</p>	Selling food or snacks at market/bus-stop/school	Sewing for pay	Bicycle/cart deliveries	Cleaning of offices	Selling food from home	Car washing	Selling sweets from home (fudge, etc.)	Taxi-driver		Lottery vendors		Any other similar activity
Babysitting	Nurse's Aid for pay																		
Laundry, ironing for pay	Subsistence farming																		
Cleaning yard/cutting yard																			
Selling food or snacks at market/bus-stop/school	Sewing for pay																		
Bicycle/cart deliveries	Cleaning of offices																		
Selling food from home	Car washing																		
Selling sweets from home (fudge, etc.)	Taxi-driver																		
	Lottery vendors																		
	Any other similar activity																		

**5 Employed: Suggestions for enhancing effectiveness of questions**

187. One important consideration in any decision to adopt the extended scope of economic activities is how comparability with results from previous censuses and with other sources can be maintained, while setting a new base year for the new concept. On

the other hand, if the previous questions were inefficient in capturing specific groups of workers, it may be more critical to improve coverage than to maintain comparability. It is, however, still important that the approach adopted permit separating these household economic activities from the other activities, especially for those who have concerns about how including some types of household production, such as fetching water and collecting firewood, might dilute the policy relevance of the measures of employed and unemployed.

188. The following steps can be adopted for this purpose:

- (i) Keep the basic questions used in the previous census (with improvements as deemed necessary, based on the census experience);
- (ii) Add one question that prompts for the main activities comprising the difficult-to-measure household economic activities, including fetching water and collecting firewood.

189. Several points to note about introducing the second question:

- (a) Countries have to make fewer adjustments to previously used questions:
  - (i) If they used two or more questions in their census.
  - (ii) If the questionnaire format allows more space for writing questions (for example one or more pages of questions per individual, or several columns for each person with ample space in the rows to enter the questions; as opposed to one page to all or almost all members of the household).
  - (iii) If they have regular labour force or household surveys which can be used to determine the effect on employment and unemployment figures of adopting the broader scope of economic activities. The census could then be limited to the main set of questions.
- (b) It is preferable that the second question be confined to probes on specific economic activities (relating to work done), and not be mixed with categories of job attachment.
- (c) These probes can either be presented as a set of categories in a question, and the response(s) entered against the appropriate category(ies), or as a question with a “yes” and “no” response categories, and the activity list entered on the questionnaire or on a prompt card, that will be read by the enumerator. An example of the prompt card is presented below in Diagram 3. This should be modified to suit national circumstances.
- (d) A decision is required on whether there is to be only one (as in example A.8) or there can be multiple responses (as in example A.7). It is useful to allow multiple responses, so that different types of aggregation can be possible at the data processing and tabulation stage. In this context, the implications for coding and data processing should be carefully considered.

**DIAGRAM 3: CURRENT ECONOMIC ACTIVITY PROMPT CARD**

<p style="text-align: center;"><b>ECONOMIC ACTIVITY</b></p> <p style="text-align: center;"><b>DID YOU DO ANY OF THE FOLLOWING ACTIVITIES IN THE LAST 7 DAYS</b></p> <ul style="list-style-type: none"><li>• <b>Any activities on your own or the family farm, garden, cattle post/kraal of any kind.</b> <i>(Products can be for sale or family use)</i> <i>Examples:    Planting Crops</i> <i>                  Weeding</i> <i>                  Harvesting Crops</i> <i>                  Chasing Birds or other pests off Crops</i> <i>                  Looking after Cattle or other Animals</i> <i>                  Moving Crops to storage or sale</i></li><li>• <b>Fetch any water or collect any firewood</b></li><li>• <b>Any kind of business activity big or small</b> <i>Examples:    Small shop/Kiosk/Tuck shop</i> <i>                  Street or Market Selling</i> <i>                  Shoe Cleaning/Repair</i> <i>                  Beer Brewing/Repair</i> <i>                  Making/Selling Hats, Mats, Baskets, Caps</i></li><li>• <b>Catch, collect, cultivate any fish, prawns, seaweed, shells or other food from the sea, river or fish/prawn farm</b> <i>(Fish etc. can be for sale or family food)</i></li><li>• <b>Make anything from farm or natural products for sale or own use</b> <i>Examples:    Making Handicrafts, Mats, Hats from Straw or timber</i> <i>                  Making Charcoal</i> <i>                  Making Butter from Milk</i></li><li>• <b>Do any construction or major repair work on your house or farm or for anyone else</b> <i>Examples:    Build/Repair Fences/Kraals</i> <i>                  Make/Repair Irrigation Channels</i></li><li>• <b>Have any type of wage or salary job, paid in money or by food/clothing/accommodation</b> <i>Examples:    Full or Part time jobs</i> <i>                  Casual or Piecework Jobs</i> <i>                  Permanent or temporary Jobs</i></li><li>• <b>Do any volunteer work to produce any building or other physical good</b> <i>Example:     Self-help community construction of bus shelter</i></li><li>• <b>Do any volunteer work to provide a service for a market enterprise</b> <i>Example:     Volunteer work at private hospital</i></li></ul> <p><b>Including any temporary absences for leave, sickness or family/village responsibilities</b></p>
--

*Example A.12 (Untested, hypothetical question)*

190. The final example is an untested hypothetical example to show how coverage of the new activities to be regarded as employment can be ensured. Some other questionnaire techniques are also suggested.

<p>A.12.1: <i>Did you/... do any work for one hour or more in the last 7 days?</i></p> <p>A.12.2: <i>Did you/... do any unpaid work in a family business or farm in the last 7 days?</i></p> <p>A.11.3: <i>Did you/... do any work at all in the last 7 days?</i></p> <p><b>Prompts: Any work on your farm or kraal!</b> <b>Any Fishing or seafood collecting!</b> <b>Make anything for sale or your own use from farm or natural products!</b> <b>Fetch any water or collect any firewood!</b> <b>Any work in a business of any type!</b> <b>Any type of wage job (Full time or part time)!</b></p> <p>A.12.4: <i>Do you/... usually work but happened to be absent last week because of leave, sickness, bad weather, industrial troubles or other reasons</i></p> <p><b>Interviewer: If Yes to any of A.12.1 to A.11.4 skip to next Section</b></p>
--

191. The major points to note are:

- (i) Prompts are included on the major employment activities. It is preferable to make them specific and include the most important activities likely to be missed;
- (ii) Separate questions emphasise the one-hour limit to pick up short-time work, unpaid work and temporary absences. These are the situations commonly missed when only one question is asked on whether the individual did any work during the reference week.

192. A possible extension would be to make each of the prompts in A.12.3 a separate question with their own 'Yes/No' answer, or a multiple response question. One drawback of this approach is that it will extend the sequence of questions. It would however supply additional information and could be used for creating comparability with the results from previous censuses. If compatibility with historical data is a priority objective, the building-block approach may be adopted as earlier mentioned; to keep the questions that have been used to enumerate the employed, and add a combination of A.12.1, A.12.2 and A.12.3 to get the other groups of employed that may have been missed, as in example A.9. The temptation to incorporate status in employment variable here should be resisted. An alternate approach would be to use a *Prompt Card* in place of the prompts to Question A.12.3. A possible prompt card is given as Diagram 3, above. Note that the examples of activities will differ for each country, therefore the examples on the *Prompt Card* need to be varied and country-specific. In addition, relevant activities can be added.

193. As stated earlier, prompt cards have proved to be very effective in surveys and

certainly should be examined and tested for censuses. The list and ordering of categories, and specific activities need to be tested with respect to (i) ease with which the respondent can identify with some of the activities; (ii) spontaneity with which the respondent reacts to the categories as they are read out; (iii) effect of the order of questions and of the list of activities; and (iv) significance of the numbers added to the employed, from the use of a multiple-category question compared with the use of the prompt card to probe for important activities most likely to be missed.

## **6      *Employed: Issues of measurement***

194. The following aspects may need pre-testing on current activity:

- (a) Any change in length or type of reference period from that used in previous censuses;
- (b) Success or otherwise in recording difficult groups e.g., part-time and unpaid work etc.;
- (c) Effects of inclusion of all additional groups from the extended production boundary on all classifications e.g., effect of inclusion of fetching water and collecting firewood on activity status, industry and occupation groups;
- (d) Effect of prompts, additional questions or prompt cards with the previous point;
- (e) Possible additional categories/wording/examples for prompts or prompt cards based on persons wrongly being recorded;
- (f) Temporary absences - what is being taken into account?
- (g) Evaluation of 'seeking work' compared to 'available for work' for the unemployed;
- (h) Categories being recorded in 'Other' for the 'not economically active';
- (i) If a simple layout is employed, its success in recording both employed and unemployed.

## **C      *The Unemployed Population***

195. In many developing countries, unemployment as defined with three criteria (without work, available for work and seeking work) is not particularly relevant as an indicator of 'oversupply' in the labour market. A low general unemployment rate certainly does not mean that the developing countries or regions do not have employment problems. Lacking the unemployment benefits and other social services common in most developed countries, most people in developing countries cannot be totally 'unemployed' or they would not survive. Few families can support totally unemployed youngsters or adults. Many people in developing countries generate incomes from a variety of informal activities or survive by growing their own food, and these activities represent 'employment'. In developing countries unemployment in the strict sense is therefore mostly found in urban or peri-urban areas where a person can be totally unemployed while obtaining support from other employed family members or relatives. Among young persons unemployment may also tend to be concentrated among those who are

better educated, as they tend to come from better off families that can afford to support periods of ‘unemployment’ for them. Thus the unemployment rate, as conventionally defined, is a useful indicator of labour market problems only for these particular groups in these countries.

196. The problems in these countries are more complex than this simple rate can show and this is why better and broader statistics on economic characteristics are needed, so that the situation of both the economically active population and the population not currently economically active can be examined in some detail. However the census may not be the best source to obtain these more refined measures of the labour force and they should be left to labour force and other household surveys. The major value of the census measure is that it supplies statistics for small areas and small population groups.

**1. Unemployed: The operational definition**

197. The “unemployed” comprises all persons above a specified age who during the reference period were:

- (a) Without work, i.e., were neither in paid employment nor self-employment,
- (b) Currently available for work, whether paid employment or self-employment; and
- (c) Seeking work, i.e., had taken specific steps in a specified recent period to seek paid employment or self-employment.

**2. Unemployed: Requisites for applying the definition**

198. The first major requirement to be counted as an unemployed member of the economically active population is that the person should not have been employed during the reference period. The second major requirement is that the person could have worked in the relevant reference period, or could have started working almost immediately afterwards if offered a job during that period. The third requirement is that the person in the recent past should have been actively seeking work.

199. The requirement to be ‘not working’ applies to the reference week in such cases, following the priority rules. For the ‘available’ criterion the international recommendations state that the period should be the current reference week (although some countries use the current reference week and the next two weeks). The extension of the availability period into the future allows a person to make the arrangements necessary to cope with other commitments after having received a job offer, e.g., women engaged in childcare or housework. The international recommendations do not specify a precise reference period for actively seeking work, but the previous month or four weeks has been used most commonly. Lengthening any of the latter two periods will give broader measures of unemployment and add to the numbers of unemployed persons. It is important, as earlier mentioned, to be aware of and try to avoid including any public holidays and other major events of more than a few days duration in reference periods for

search or availability.

200. The means to ‘actively seek work’, as understood in developed countries, include: registration at a public or private employment exchange; application to employers in writing or in person; checking at work sites, farms, factory gates, market or other assembly places and placing or answering newspaper advertisements. Steps towards self-employment include looking for land, building, machinery or equipment to establish own enterprise; arranging for financial resources; applying for permits and licenses etc., or other steps to start a business or, similar steps to start an agricultural activity. These “active” steps are often a problem to identify in most developing countries, particularly in rural areas. There are often no employment officers or they are known to be ineffective and few bother to contact them. There are also often few businesses or government offices in the area at which to apply for work. Funding or suitable land for businesses or agriculture is often also very limited. However, the person has no work and may obviously be available for work if any type of economic activity was open for him/her in the area. Women in particular may keep themselves busy with housework and take no active steps to find employment, but if the opportunity presented itself they could easily be available for that work.

201. The traditional strict measure of ‘unemployment’ that requires persons to have been actively seeking work will therefore generally mean that the measured ‘unemployment’ rates for many developing countries will be low. The international recommendations do recognize this problem and allow countries to relax the actively seeking work criterion if necessary in their situation. Very few countries seem to have done this in their censuses and this may partly explain why very low unemployment rates are reported from most developing and transition countries.<sup>67</sup> It is not only the requirement that a person should have been actively seeking work that might act to lower the ‘unemployment’ rate, but also the broad definition of ‘employment’ if it is effectively applied.

202. Ideally questions on both ‘actively seeking’ and on ‘available’ should be asked to make it possible to give statistics for both ‘strict’ and ‘relaxed’ unemployment measures. It may, for example, be useful to ask both questions in one census to provide both strict and relaxed estimates of unemployment, particularly for comparing the results, and then possibly continue in future censuses with one concept only. It would also be useful to distinguish first time job-seekers from other job-seekers in the classification of the unemployed.

203. The priority rules of the labour force framework give precedence to unemployment over economic inactivity. Thus even persons who during the reference period were mainly not economically active should be classified as unemployed if they satisfy the three conditions in the above definition of unemployment. Such persons could include students, homemakers and pensioners. Note however that, from these same

---

<sup>67</sup> The possibility and relevance of using only the ‘currently available for work’ criterion therefore should be explored.



priority rules, persons who had an economic activity during the reference period should be classified as employed even if they were mainly without work, available for work and looking for work during this period.

### **3. *Unemployed: Review of questions used in national censuses***

204. Questions on unemployment have been formulated in a variety of ways, and they are differentiated by (i) the length of the reference period, that is they may either use the same reference period as the employed, (for example, one week), or a reference period longer than but including the reference period for the employed (for example one month, including the week for the employed); (ii) whether only one of the two criteria of “seeking work” and “available for work” or both are used; (iii) if both criteria are adopted, whether they are applied in one question or in a two-part or two separate questions; and (iv) whether to use the same reference period for all three (or two) criteria.

205. The following types of questions, with emphasis on the use of the two criteria of “seeking work” and “available for work”, are presented and discussed:

- (a) One question that uses one criterion only, either applying the short, or the long reference period (examples B.1 to B.6);
- (b) One question with both criteria (i) applying the same reference period for both criteria, or (ii) applying a different reference period in either criterion (example B.7);
- (c) Two or more questions or parts to a question addressing each criterion in turn, (i) applying either the same reference periods, or (ii) two different periods corresponding to each criterion (examples B.8 to B.13).

**Example B.1**

**B.1.1 Are you: \***

**employed:**  
 full-time work (33 and more hours per week) — }  
 part-time work (12 to 32 hours per week) — } Please answer questions 11 to 16!  
Employers, the self-employed, farmers, as well as workers in the family business are considered to be employed if they work 12 hours a week or more.

**not economically active, but:**  
 unemployed — } please answer questions 11 to 13 on your last occupation!  
 parental leave, maternity leave — } if you have never worked before and are looking for a job, answer question 12 with "no profession yet!"

Were you unemployed at the time you entered maternity leave:      yes      no      —      —

military service, replacement service — } Please answer questions 15 and 16 for your journey to the barracks or to your place of service.  
 retirement benefits from own employment — } Please answer questions 11 and 12 on your last occupation!  
 widow's pension —

homemaker —

pupil, student — } Please answer questions 14 to 16 on the school you are presently attending!  
 child presently not attending school —  
 other livelihood — e.g. rent, alimony, social aid, financial support through relatives, etc.

**Example B.2**

<b>B.2.1 Activity status</b>	
Last week (Name)	
Mark with "X" one circle only	
Worked?	<input type="radio"/> 1 } Go to
Did not work, but had a job?	<input type="radio"/> 2 } XX
Looked for work?	<input type="radio"/> 3
Is a student?	<input type="radio"/> 4
Is a homemaker?	<input type="radio"/> 5
Is retired or receiving pension?	<input type="radio"/> 6
Is incapacitated permanently to work?	<input type="radio"/> 7 Go to YY
Did other activities?	<input type="radio"/> 8

<b>18. CONDICIÓN DE ACTIVIDAD</b>	
¿La semana pasada (NOMBRE) B.2.1	
Marque con "X" en solo círculo	
trabajó? .....	<input type="radio"/> 1
no trabajó, pero sí tenía trabajo? .....	<input type="radio"/> 2
buscó trabajo? .....	<input type="radio"/> 3
¿Es estudiante? .....	<input type="radio"/> 4
¿Se dedica a los quehaceres de su hogar? ....	<input type="radio"/> 5
¿Es jubilado (a) o pensionado (a)? .....	<input type="radio"/> 6
¿Está incapacitado (a) permanentemente para trabajar? .....	<input type="radio"/> 7 a Y.Y
¿Realiza otras actividades? .....	<input type="radio"/> 8

206. The meaning of 'unemployed' is left for the interpretation of the respondent. It could therefore reflect individuals who fit either one of the criteria or both. Also, the structure of the question suggests that the "unemployed" are considered among the "not economically active" which is conceptually not the case, according to the international recommendations.

207. The basic question seeks to determine activity status of individuals, and uses the reference period of the “past week” which is applied to all categories presented, one of which refers to whether the individual “looked for work”. The next question acts as a filter to exclude from the groups reporting that they did not work in the reference week, those that performed some kind of economic activity (see para.27 on page 12, relating to example A.8.2). This gives an approximate estimation of those who may have been employed. The reference period is probably too short, since some people who may be looking for work do not keep at it continuously. The criterion on “being available” is not applied. These two elements can produce opposite effects on the total number of persons classified as “unemployed”, but do not necessarily cancel out.

**Example B.3**

Type of activity during past 30 days	
<p><b>B.3.1</b> During the past 30 days did ... work for cash?</p> <p>1 <input type="checkbox"/> Yes for someone else } GO TO XX          2 <input type="checkbox"/> Yes for self }          3 <input type="checkbox"/> No GO TO B.3.2</p>	<p><b>B.3.2</b> Then what did ... do during the past 30 days?</p> <p>1 <input type="checkbox"/> Family business          2 <input type="checkbox"/> Work at Lands/Farms/Cattle Post</p> <hr/> <p>3 <input type="checkbox"/> Actively seeking work          4 <input type="checkbox"/> Housework          5 <input type="checkbox"/> Student          6 <input type="checkbox"/> Retiree          7 <input type="checkbox"/> (Other specify) _____</p>

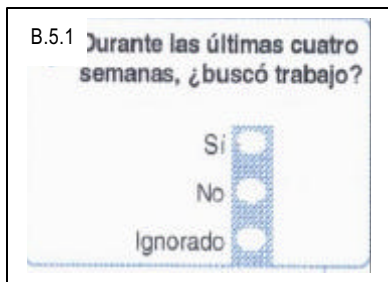
208. The question in example B.3 also uses the same basic question for some types of activities of the employed, the unemployed and the not economically active groups. The “30 days” reference period applies to all activity statuses but is a reasonably long period to consider measuring the unemployed. It is the duration typically used “for actively seeking” even for those that apply the “one week” reference to the employed. With the landscape layout used in this country’s questionnaire there is always very little room for prompts or explanations on the questionnaire, and in this case the manual and training would have to have thoroughly covered the other categories of ‘Work’ and the meaning of ‘Actively seeking work’ in particular. Note also that in practice the ‘Actively seeking work’ activity will be combined with one or more of the others mentioned, therefore, allowing for multiple replies would have been an advantage.

**Example B.4**

<p><b>B.4.1</b> If you/. . . . had been offered a job that week what would have prevented you/. . . . from taking up that job?</p>	
<p>① Nothing</p> <p>② School</p> <p>③ Home duties</p>	<p>④ Retirement</p> <p>⑤ Not interested in working</p> <p>⑥ Other (Specify _____)</p> <p>⑨ Don't know/Not stated</p>

209. The question (B.4.1), is addressed to all who neither worked nor had a job from which they were temporarily absent during the reference week. The main criterion applied in this question is “available for work” therefore applies the relaxed definition. The question is complicated by the fact that the reasons why the individual would not have taken a job offered in the reference period, is sought without first establishing whether the individual was available for or interested in taking a job. Only one category gives the clear idea that the person is not seeking work. Moreover, the response category “nothing” would seem to suggest that this is the group that can be taken as the unemployed, but it presents some ambiguity as to what “nothing” really represents. People in this category may still not take a job that is offered to them, though nothing prevents them from taking it. The reference period of one week is also short, because some people with “home duties” may need more time to arrange for help, even if they wanted to accept the job offered.

**Example B.5**



**B.5.1** During the last four weeks did you (or *Name...*) look for work?

Yes

No

Don't know

210. The question (B.5.1) posed to all who neither worked nor were on leave or temporary lay-off from their job during the past week, is straightforward and should be understood by most people. It does not state what is entailed in “look for work”, but presumably the respondents could take the smallest form of effort as seeking work. The duration is the four weeks preceding the census, as recommended in the international standards.

**Example B.6**

**B.6.1 Did you actively look for work over the last four weeks**  
 For instance: through employment bureau, wanted ads, contacting potential working places, etc.

No      2                      Yes      1

211. Question B.6.1 is preceded by two others on employment that refer to “last week” as the reference period and the unemployment question refers to last four weeks (that is incorporating the last week as well) following the common approach. One question is used to determine the unemployed status, and only the actively seeking criterion is applied. The question includes some clarification of what “actively seeking” entails. However, not all those who were seeking employment may be available to take a job offered to them at short notice. The measure from this single question may therefore overestimate the number of people who fit the strict definition of unemployed, i.e., “who are seeking and are available” for work.

**Example B.7**

<b>B.7.1</b> If Marginal Worker or Non-Worker, is the person seeking/available for work?	
Yes-1	No-2

212. In Example B.7 ‘Marginal’ and ‘non-worker’ are defined on the basis of previous questions. The concepts “seeking/available” need to be defined for this to be an effective question, but such definitions are not presented. It cannot be deduced from their question whether respondents, when they indicated “yes” had sought work, were available for work, or both. The brevity of the question stems from the very limited space, which the “landscape” format of presentation offers.

**Example B.8**

<b>B.8.1</b> What were you mainly doing in the last 7 days?	
Worked - Paid non-seasonal	1
Worked - Unpaid non-seasonal	2
Worked - Paid seasonal	3
Worked - Unpaid seasonal	4
On leave	5
Unpaid work on household holding or business	6
Unemployed and seeking work	7
Not seeking work but available for work	8
Full-time housewife/homemaker	9
Full time student	10
Not available for work for other reasons	11

213. Question B.8.1 combines several elements of both activity status and status in employment in the list of possible response categories. Only one response is expected given the key word “mainly” in the basic question. The reference period of one week applies to the definitions of both ‘employed’ and ‘unemployed’, and therefore some of the ‘unemployed’ may be lost to other categories such as “homemakers”, e.g., those who had searched for a job a bit earlier and/or would be available for work a bit later. On the other hand, the use of “mainly” might inflate the number that could be considered as ‘unemployed’, if for example, more time was spent in the reference week seeking work,

than working (e.g., if the person was working one day and searching for work the other four to six days). Thus, it is the respondent that decides the order or priority to assign to the different activities that s/he was engaged in during the reference period. The two criteria are reflected in the categories (i.e., responses 7 and 8). However, the response category “Not seeking work but available for work” is not a direct response to a question on what one was mainly doing, since there is overtly no action involved.

**Example B.9**

**B.9.1(a)** Has This person been looking for work to earn money during the last 4 weeks?  
 1  Yes (**GO TO X.X**)  
 2  No (**SKIP TO B.9.2**)

**B.9.1(b)** Last week could this person have started a job if offered one, or returned to work if recalled?

1  Yes, could have gone to work  
 2  No, because of temporary illness  
 3  No, because of all other reasons (in school, etc.)

**B.9.2** When did this person last work even for a few days? Do not include subsistence activity

1  2000  
 2  1999  
 3  1998  
 4  1995 to 1997  
 5  1990 to 1994 skip to  
 6  1989 or earlier skip to  
 7  Never worked or did subsistence only skip to

214. As can be noted in Example B.9 only the strict concept of unemployment was used. Response 7 of question B.9.2 should preferably distinguish those who have never worked before from those who have only worked in subsistence agriculture/fishing.

**Example B.10**

B.10.1 Last week did you have definite arrangements to start a new job within the next four weeks?  
 No  
 Yes

B.10.2 Did you look for work during the past four weeks?  
*For example, did you contact a National Employment Centre, check with employers, place or answer newspaper ads?*  
*Mark one circle only*  
 No  
 Yes, looked for full-time work  
 Yes, looked for part-time work (less than 30 hours per week)

B.10.3 Could you have started work last week had a job been available?  
*Mark one circle only*  
 Yes *could have started work*  
 No, Already had a job  
 No, temporary illness or disability  
 No, personal or family responsibilities  
 No, going to school  
 No, other reason

215. The three questions in Example B.10 are related to seeking and being available for work, but the second and third are more specific to measuring unemployment. The first, B.10.1, targets those who are waiting to start work within the next four weeks whereas the other two apply to those who did not work the previous week, and have been actively looking for a job. The reference period for the search (B.10.2), is reasonably long and the time within which the job seeker could start work were one to be available is just one week as recommended in the international standards

*Example B.11 (included in a short form questionnaire)*

B.11.1	<p>Last week were you doing any of the following?  <i>(X all the boxes that apply)</i></p> <ul style="list-style-type: none"> <li>• In paid work</li> <li>• Retired from paid work</li> <li>• In full time education or a student</li> <li>• Working on a government sponsored training scheme</li> <li>• Looking after family or home</li> <li>• Permanently sick or disabled</li> <li>• Temporarily sick or injured</li> <li>• Looking for work and available to start within two weeks</li> <li>• Looking for work but not available to start within two weeks</li> <li>• Waiting to start a new job</li> <li>• Doing unpaid work in own/family business</li> <li>• Doing unpaid voluntary work</li> </ul>
--------	---

216. As indicated earlier in Example A.7, the categories used in question B.11.1 are not all applicable or relevant for other countries, and testing may show that some adjustments may be necessary if this approach is adopted. The two categories dealing with the unemployed each has two parts and are differentiated with the conjunctions “and” or “but”. This formulation could work out reasonably well with a self-administered questionnaire (as in this particular case) but may not work well in the interview mode of data collection. The two categories may sound the same, if the respondent is not very attentive. (Highlighting, or for the interview mode, stressing the words **and** in the first and **but not** in the second may help.) If the technique of multiple answers is also adopted, interviewers will need to be made aware that the respondent may only check one of these two categories, with any others. Such a restriction in a multiple-choice question would need to be tested out. Also, there is no category for “available for work, but not looking for work”, which would permit using the relaxed definition of unemployment.

**Example B.12**

<b>B.12.1</b> Indicate whether, during the last 4 weeks, the person has actively tried to find a job or attempted to open own activity			
Yes...	01		
No ...	02		
<b>B.12.2</b> Indicate whether, should the opportunity arise, the person would be willing to start work within 2 weeks			
Yes...	01		
No ...	02		
<b>B.12.3</b> Indicate whether the person has ever held a paid job or worked as a family helper during their lifetime			
Yes ...	01		
No ...	02		
	<table border="1"> <tr> <td><b>B.12.4</b> Indicate the year when terminated</td> <td>go to point xxx</td> </tr> </table>	<b>B.12.4</b> Indicate the year when terminated	go to point xxx
<b>B.12.4</b> Indicate the year when terminated	go to point xxx		

217. The reference period for looking for work in question B.12.1 follows the international standard of four preceding weeks, but the reference period for availability for work in question B.12.2 is for the next two weeks only. This should be “last week” or “last week and the next 2 weeks” to be consistent with international practice. The third and fourth questions are used to distinguish the first time job seekers, from those who have recently lost a job or re-entrants into the labour market.

**Example B.13**

ACTIVE STEPS <b>B.13.1a</b>	AVAILABILITY <b>B.13.1</b>
<p>If NO to X.X <b>In the PAST FOUR WEEKS before 10 October has (the person) taken active steps to find employment?</b> Y = Yes N = No For example, (the person) went to visit factories or other employment places, placed or answered advertisements, looked for land or a building or equipment to start own business or farm.</p>	<p>If NO to X.X <b>If offered work, how soon could (the person) start?</b> 1 = Within one week 2 = More than 1 week, up to 2 weeks 3 = More than 2 weeks, up to 4 weeks 4 = Some time after 4 weeks 5 = Does not choose to work <b>Go to Y.Y</b></p>



218. Two questions are used in Example B.13 to determine the unemployed status. The first gives “the past four weeks” as the reference period, as in most of the previous examples, and gives examples of what constitute “active step” used in the question. The “yes/no” response categories help to simplify the question. Availability to start work is assessed in the next question. Instead of applying a specific reference period, four time intervals are used, three of which correspond to reference periods generally used with this question. This approach provides the flexibility to define and subsequently adjust the reference period through aggregation of these responses. It also permits analyses and some policy questions related to unemployment that could be explored. If this formulation is to be adopted in countries that have not previously used it, it is important that the question is tested. In particular, the effect that presenting the range of options might have on people’s willingness to start work offered within a short period should be ascertained. It is for example possible that many people who could start work within two weeks, might have the higher time range, if the option were presented. If a reference period was then imposed at the tabulation stage, the results could be misleading.

#### **4. *Unemployed: Suggestions for enhancing effectiveness of questions***

219. To try to obtain a better insight into the unemployment problem one might try to test the possibilities for relaxing the ‘actively seeking work’ criterion. Household surveys often include questions on ‘actively looking for work’ and on ‘available for work’. This provides a possibility for investigating the effects of relaxing one of the criteria when defining ‘unemployment’. Testing should be done to establish the effect before the relaxed definition is introduced into the census questionnaire. The ‘available for work’ criterion is better as a separate question than as a simple response category. However, it is probably best to aim for a simple, crude unemployment measure from the census and leave it to household surveys to apply more precisely defined concepts. Census planners should definitely investigate carefully this often thought simple topic to establish exactly what they are measuring.

#### ***Example B.14 (Untested, hypothetical question)***

220. This question is the follow-up to that presented earlier in the section on employment as Example A.11. The focus in this section is on questions B.14.1 and B.14.2, which deal with unemployment. Both the strict and the more relaxed definition of ‘unemployed’ can be used on the basis of the responses given to questions B.14.1 and B.14.2.

<b>B.14.1:</b> <i>Did you/... actively look for work sometime during the last four weeks?</i>	
<b>B.14.2:</b> <i>If offered a work opportunity could have started work last week?</i>	
<b>B.14.3:</b> <i>What were you doing, what was your situation, last week?</i>	
Studying .....	1
Performing Housework.....	2 CODE
<u>Not working</u> and disabled .....	3
Sick.....	4
Retired/aged.....	5
Pension, rental or other income recipient.....	6
Other .....	7

**5. Unemployed: Issues of measurement**

221. The possible variations on the unemployment question(s) should be thoroughly tested before being introduced. Pre-testing should also focus on the respondents' interpretation of the word 'available', particularly when translation in the field by the enumerator is common.

**D Population not currently active**

222. As noted earlier, this group consists of all in the population who during the reference period are neither employed nor unemployed. It is useful to determine the reasons why these persons are in this group in terms of the alternative activities they were engaged in that are assigned less priority in the labour force framework than being employed or being unemployed. Question B.14.3 above elicits responses from those who were not employed during the reference period as to the nature of their actual activities during this period. This could permit the analysis of the reasons why some persons are not currently economically active. The categories used in the question should be checked in pre-tests to ensure they are being understood and are sufficient and useful for the prevailing situation in each country.

## Chapter 5. Usual Economic Activity Status

### A Introduction

223. The statistics obtained from the 'usual' approach are conceptually valuable to obtain a stable picture of economic activity over twelve months, as a basis for all aspects of economic and manpower planning, and are particularly valuable as a basis for national accounts estimates. Also, they are less dependent on the timing of the census date and fit in with other statistics determined on an annual basis, such as household income. The twelve months prior to the census date was used for at least one census question in 47 out of 181 countries that reported census results to the United Nations in the 1975-1984 period {ILO COMMENT: UNSD TO UPDATE}. However, even in many of these censuses, the major block of economic questions referred to the week prior to the census and the long reference period was used only for one or two questions, and not for the set of other economic characteristics items in the census.

224. Only two countries used the international standard concept of usual activity as described below in which the number of weeks or months of employment, unemployment and inactivity are separately collected. Most of the countries with questions on "usual" activity measured the number of weeks/months of employment only or asked about predominant activity over a 12 month period.

225. Compared to the short reference period used for 'current' activity the long reference period used for usually economically active leads to a considerably larger proportion of persons who have experienced different activity status situations. During a 12-month period many persons change between economic and non-economic activities, depending on the season as well as on economic and demographic factors and events, and they can do different types of economic activities at different times of the year. It therefore is very difficult to obtain meaningful responses by simply asking persons: "What was your usual activity in the last twelve months?" or "Were you usually employed?" The questions that are required need to probe for all types of economic activity and preferably aim to quantify the answers in terms of time spent on the various forms of activity. This can mean that it will be necessary to use several questions. There may also be some difficulty in identifying a main job that would serve as the reference for economic characteristics such as occupation and industry. The relative precision of the current measure and general ease of interpretation for interviewers and respondents makes this measure more preferable to the usual measure for the questions on other economic characteristics.

226. It has been recommended in the *Principles and Recommendations, Revision 2*, that countries consider measuring both the usually economically active population and the currently economically active population. It is also strongly recommended that any country starting to use usual activity for all their economic characteristics questions, conduct a very extensive pre-testing programme to identify all difficulties and to build

experience in applying the concepts.<sup>68</sup> The broad aim of measuring the usual activity is to obtain a 'total' or 'average' picture of economic activity over all seasons.

227. There is less need to measure usual activity in a population census, if a country has a:

- (a) reasonably even pattern of economic activity throughout the year;
- (b) very high proportion of employment in wage earning or formal business activities, which by their nature are not strongly seasonal; or
- (c) very good programme of household surveys to measure economic activity at different times of the year and thus covering the various seasons. However, this situation is rare in developing countries.

### ***B The operational definition***

228. The "usually active population" comprises all persons above a specified age whose activity status, as determined in terms of the total number of weeks or days during a long specified period (such as the preceding 12 months or the preceding calendar year) was "employed" and/or "unemployed" as defined within the labour force framework.

229. The "usually employed" comprises all persons in the usually active population who, during the period of usual activity, had a total number of weeks or days of employment that was at least half of this period. Conversely, the "usually unemployed" comprises all persons in the usually active population who, during the period of usual activity, experienced a total number of weeks or days of unemployment greater than half this period.

230. The "population not usually active" consists of all persons in the population who were not usually active during the long specified reference period, including those below the specified age.

231. Diagram 4 gives a pictorial illustration of these groups and the relationship between them.

### ***C Requisites for applying the definition***

232. One fundamental issue in the choice of a reference period, is whether the last twelve calendar months be used or a calendar year, e.g., 2010. Recall problems or memory loss is a major problem with such long periods, in particular concerning the timing of events and duration of activities, and it has been found best to use the twelve calendar months ending at the month prior to the census. However, some countries have

---

<sup>68</sup> Surveys in countries that have measured both current and usual activity, have shown considerable differences in the results even in the basic activity classification (as employed, unemployed etc.). Thus, any change from a current activity approach to a complete usual activity approach is likely to result in lack of comparability with previous census results.

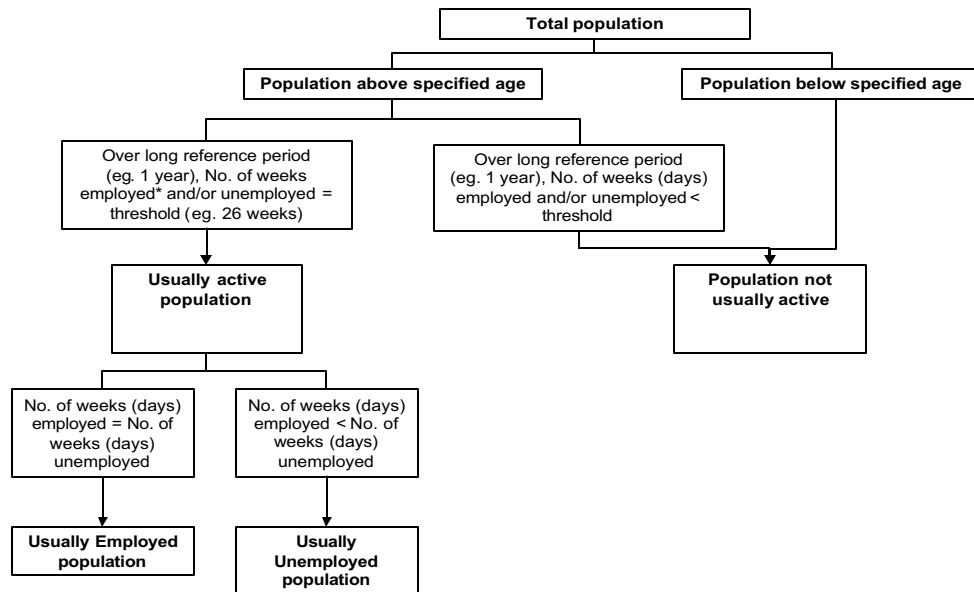
used the calendar year, as it is believed to give a more definite reference period to which respondents can relate. It also gives a reference period for the results, which can more easily be combined with other types of statistics for the year. If the census is early in the calendar year then there may be little difference and the calendar year could be preferred. However, this should be established by testing. A possibility for making recall easier is to fix the reference period by referring to a national event that took place just before the census enumeration (if there is one).

233. The basic priority rules regarding economic and non-economic activities, as discussed in the section on current activity, state that within the reference period, (i) economic activities take preference over non-economic activities, and (ii) within economic activities employed activities take preference over unemployed activities. Some people are both employed and unemployed during the same day, week or month as they are looking for other work. The above rules determine the question order. Questions to determine the length of all types of employment activity should always be asked first, followed by questions to determine the length of experiences of unemployment, and finally questions about the length of any non-economic activities. For usual activity the rules also determine the way in which individuals are to be classified to the main groups:

- (a) If a person spends more time as economically active (employment and/or unemployment) than as not economically active over the twelve months, the person is considered to be 'usually economically active', and the converse holds for the person to be 'not usually economically active';
- (b) If an economically active person spends more of the active time as employed than as unemployed, then the person is taken as 'usually employed', and the converse holds for the person to be 'usually unemployed'.

**DIAGRAM 4: USUALLY ACTIVE POPULATION**

**Diagram 4: Usually Active Population**



\* As defined for current activity

\*\* Adapted from Hussmans et al. 1990, p. 52

234. In applying the definition of usually active two procedures may be followed at the time of enumeration to determine the usual activity status of each person. One is to interpret it as the status that prevailed over most of the 52 weeks or most of the 365 days during the specified reference year. Another is to set a specific number of weeks or days as the cut-off point and classify anyone with at least that many weeks or days of economic activity as the "usually active population".<sup>69</sup>

235. Another decision in the usual activity measurement is what unit of counting is to be used, i.e., days, weeks or months, to record a person's activity and ultimately derive the major usual activity categories. A days-based measure would be very good if it can be obtained, but this would be very unlikely to be with any degree of accuracy. Respondents have great problems remembering the number of days they spend on any activity in a reference month and certainly will not be able to give accurate estimates for the last twelve months. Note that in principle the choice of unit is an issue of precision of measurement, (just as kilometre is less a precise measurement of distance than meter).

- (a) Weeks may represent a better degree of precision, which may be possible and would warrant testing.
- (b) Months is the easiest but would be a very approximate measure as activities may change significantly during a month.

<sup>69</sup> Hussmans, and others (1990). Surveys of the Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods. International Labour Office, Geneva 1990, p. 51.

However, partly depending on question formulations, reference to a week or a month may be interpreted as a 'whole' or 'full' week or month by the respondent. Thus it may be better to get an approximate (and possibly rounded) estimate of days than an estimate for a type of period, which one would not want. Be aware that the main activity status of a person could change substantially depending on this choice of measurement unit.

236. Moreover, in countries where employment is mostly of a regular and continuing nature and where a week of employment means by and large a week of full-time employment or, at any rate, employment for a major part of the working time, the question about main activity status may well refer to weeks of employment or unemployment. However, in countries where employment is largely of an irregular nature and where a week of employment does not generally mean a week of full-time employment or even employment for a major part of the working time, the question about main activity would better refer to days of employment or unemployment.<sup>70</sup>

237. In determining the length of periods of employment, unemployment and inactivity respondents may be asked to give values for the entire long reference period or to give these values on a month-by-month (or season-by-season) basis. The latter approach is likely to have less risk of memory recall and telescoping errors. However, it is more complex and takes up more space on the census questionnaire, which could be a major concern for 1-page landscape formats. Another approach is to formulate the questions in terms of the jobs held and their durations.

#### ***D Review of questions used in national censuses***

238. The types of questions on "usual" activity status are:

- (a) One question on "usual" activity with the main questions on activity status and economic characteristics covering the short reference period, i.e., current activity (example K.1);
- (b) Questions on both "usual" and "current" employment characteristics (examples K.2 and K.3);
- (c) All questions on usual activity and related employment characteristics (examples K.4 to K.5).

---

<sup>70</sup> This is discussed in detail in the ILO survey manual. Hussmans and others (1990). pages 50-58.

**Example K.1**

<b>K1.1</b>	<b>a</b>	<b>Last year (1999), did .....work, even for a few days, at a paid job or in a business or farm?</b>		
	<b>1. Yes</b>		<b>2. No ----- Skip to xx</b>	<input type="checkbox"/>
	<b>b</b>	<b>How many weeks did ..... work in 1999? Count paid vacation, paid sick leave and military service?</b>	_____ <b>Weeks</b>	<input type="checkbox"/> <input type="checkbox"/>
	<b>c</b>	<b>During the weeks WORKED in 1999, how many hours did ..... usually work each week/</b>	_____ <b>Hours</b>	<input type="checkbox"/> <input type="checkbox"/>

239. The question in Example K.1 came after the main set of economic characteristics questions, which referred to current economic activity, and which included reasonable detail on what to include as “work”. Thus respondents had already been exposed to these issues. The question in K.1.1 is a simple one which reflects the recommendation that “any work” should be measured for those who are employed; and the number of weeks worked is useful but not sufficient for determining the “usual” status. The reference to “even for a few days” in the first question may avoid the interpretation ‘full weeks’ in the second, but using the formulation ‘such weeks’ might have been better. This set of questions is however limited, as it does not include the number of weeks of unemployment, which is necessary to identify those who are usually active and usually unemployed.

240. There is also a simple measure of the number of weeks worked with an additional question on the usual working time (without repetition of the prompts on what to include). Note that there is no attempt to measure the number of weeks of unemployment, and therefore no possibility of measuring “usual economic status”.



Example K.2

SECTION 12 ECONOMIC ACTIVITY																									
<b>K.2.1</b> What did....do most during the <u>past 12 months</u> -for example, did you/he/she work, look for a job, keep house or carry on some other activity																									
<input type="radio"/>	1	Worked										(Go to Q.XX)													
<input type="radio"/>	2	Had a job but did not work										(Go to Q.XX)													
<input type="radio"/>	3	Looked for work																							
<input type="radio"/>	4	Wanted work and available																							
<input type="radio"/>	5	Home Duties																							
<input type="radio"/>	6	Attended School																							
<input type="radio"/>	7	Retired																							
<input type="radio"/>	8	Disabled, unable to work																							
<input type="radio"/>	9	Other(please specify.....)																							
<input type="radio"/>	10	Not Stated																							
<b>K.2.2</b> Did you/he/she do any work at all in the <u>past 12 months</u> ? Include work at home, for example, piece work, decorative stitching, handicraft, sewing, etc.																									
<input type="radio"/>	1	Yes										(Go to O.XX)													
<input type="radio"/>	2	No																							
<input type="radio"/>	3	Don't Know																							
<b>Have you/he/she ever worked or had a job?</b>																									
<input type="radio"/>	1	Yes										(Go to Q.YY)													
<input type="radio"/>	2	No										(Go to Q.YY)													
<b>K.2.3</b> How many months did you/he/she work in the <u>past 12 months</u> ?																									
<i>Number of months</i>																									
<input type="radio"/>	0	<input type="radio"/>	1	<input type="radio"/>	2	<input type="radio"/>	3	<input type="radio"/>	4	<input type="radio"/>	5	<input type="radio"/>	6	<input type="radio"/>	7	<input type="radio"/>	8	<input type="radio"/>	9	<input type="radio"/>	10	<input type="radio"/>	11	<input type="radio"/>	12

241. In this example, the questions on activities in the past 12 months are followed by more detailed questions in respect of the current economic activity in the last week. There are few probes or prompts as to what to include as “work”, and the use of months (rather than weeks) is a less precise measure, as already stated.

242. The set of questions in example K.3 was drawn from a landscape layout. As can be noted one question on current activity was asked, followed by the same question with the long reference period ‘since ...1999’. Both Questions K.3.1 and K.3.2 asked “What was ... mainly doing ...?” in the two differing reference periods. The long reference period was then used for the status in employment, occupation and industry questions, but the question on status in employment has a different description (“since August 1999”

rather than “for the past 12 months”) although the periods are the same. It is

recommended that the same working be used in all questions to avoid any confusion for the respondent.

243. For current activity the ‘mainly doing’ approach is not recommended, since the respondents do not know the priority rule and “mainly” would be interpreted differently by different groups of people. It is assumed that the manuals and training gave details of definitions for ‘work’, ‘leave’, ‘unemployed’, ‘available’, etc. Details of how successful this approach was, are not available. There would be problems of comparability with respect to the statistics on the other economic characteristics, if this approach was adopted and the previous census used the more common current activity questions. When considering this type of approach it will be necessary to pre-test it well and be aware of likely problems, as explained earlier.

**Example K.3**

<p><b>K.3.1</b> What were you mainly doing in the last 7 days?</p> <p>Worked - Paid non-seasonal : 1:1          Worked - Unpaid non-seasonal : 2:          Worked - Paid seasonal : 3:          Worked - Unpaid seasonal : 4:          On leave : 5:          Unpaid work on household holding or business : 6:          Unemployed and seeking work : 7:          Not seeking work but available for work : 8:          Full-time housewife/homemaker : 9:          Full time student : 10:          Not available for work for other reasons : 11:</p>	<p><b>K.3.2</b> What have you mainly been doing for the past 12 months?</p> <p>Worked - Paid non-seasonal : 1:          Worked - Unpaid non-seasonal : 2:          Worked - Paid seasonal : 3:          Worked - Unpaid seasonal : 4:          On leave : 5:          Unpaid work on household holding or business : 6:          Unemployed and seeking work : 7:          Not seeking work but available for work : 8:          Full-time housewife/homemaker : 9:          Full time student : 10:          Not available for work for other reasons : 11:</p> <p style="text-align: right;">GO TO XX</p>	<p><b>K.3.3</b> Employment Status</p> <p>Since August 1999 what have you been mainly:</p> <p>An employer <input type="checkbox"/>          an employee <input type="checkbox"/>          self-employed <input type="checkbox"/>          an unpaid family worker <input type="checkbox"/></p>
<p><b>K.3.4</b> Occupation</p> <p>What was your main occupation for the last 12 months. Write name of occupation and enter code.</p> <p>_____</p> <p>1 0 1 0 1 0 1          1 1 1 1 1 1 1          2 1 1 2 1 2 1          2 2 1 2 1 2 1          3 1 1 3 1 3 1          4 1 1 4 1 4 1          5 1 1 5 1 5 1          6 1 1 6 1 6 1          7 1 1 7 1 7 1          8 1 1 8 1 8 1          9 1 1 9 1 9 1</p>	<p><b>K.3.5</b> Industry</p> <p>What kind of main product or service is/was produced where you work? Write name of industry and enter code.</p> <p>_____</p> <p>0 1 1 0 1 0 1          1 1 1 1 1 1 1          2 1 1 2 1 2 1          3 1 1 3 1 3 1          4 1 1 4 1 4 1          5 1 1 5 1 5 1          6 1 1 6 1 6 1          7 1 1 7 1 7 1          8 1 1 8 1 8 1          9 1 1 9 1 9 1</p>	

244. In Examples K.4 and K.5, sufficient details are collected on the number of weeks/months in each of the three statuses (employed, unemployed and inactive) to be able to determine whether the person is “usually active, usually employed”, “usually active, usually unemployed” or “usually inactive”.

245. In both K.4 and K.5, interviewers and respondents are expected to apply fairly complex rules to determine whether the person was employed, unemployed or inactive within each week/month and then aggregate these periods. This is not a trivial task and there may be problems with the quality of reporting that would need testing and post enumerations checks.

**Example K.4**

Questions are asked to persons 10 years and above					
K.4.1 What work ..... usually doing during the last 12 months?	K.4.2 How many months did ..... work during the last 12 months? ( <i>Specify approx. the no. of months spent in each category</i> )	K.4.3 What work ..... usually do? ( <i>Occupation</i> ) ( <i>Describe the actual work done</i> )	K.4.4 Where did ..... do the work? ( <i>Industry</i> ) ( <i>Name the establishment or the organisation where worked</i> )	K.4.5 What was ..... employment status?	K.4.6 What was the ..... reason for usually not working during the last 12 months?
1. Agriculture 2. Salary/wage 3. Own Eco. Enterprises 4. Extended Eco. 5. Job Seeker 6. Household work 7. Student 8. No work	1. Work done for ..... (months) 2. Extended Eco. .... (months) 3. Work seeking ..... (months) 4. No work ..... (months) Total 12 months			1. Employer 2. Employee 3. Own account worker 4. Unpaid family worker	1. Student 2. House work 3. Aged 4. Pension 5. Physically and mentally handicapped 7. Sickness or chronic illness 6. Others .....

246. In Example K.4, questions are asked to measure both the “predominant” activity and the comprehensive “usual” status. In respect of the “usual” status, questions are asked on the “approximate number of months” employed (two categories called “Work done for” and “Extended Eco.” that are presumably described in more detail in the interviewers’ manual), unemployed (“Work seeking”) and the inactive (“no work”). The time spent in these four categories adds to 12 months. As already stated, the use of “approximate number of months” is a relatively imprecise measure, but reflects the reality of memory recall over this period.

247. Although it could be assumed that the questions on status in employment, occupations and industry relate to the main job in the last 12 months, this is not stated and should have been added into the questions.

**Example K.5**

10 years & over (for persons born ..... 1991)	<b>K.5.1</b>	<b>IN THE LAST 12 MONTHS ACTIVITIES AND THEIR DURATION IN WEEKS</b>	Activity	Weeks
		A. Employed: that is working for pay, working in own enterprise, business, profession, working as unpaid family worker or working as paid apprentice.	A. Employed	<input type="text"/> <input type="text"/>
		B. Unemployed: That is without work and available for work.	B. Unemployed	<input type="text"/> <input type="text"/>
		C. Engaged in own economic activity: That is student, engaged in household work, pensioner, unable to work or child not attending school, etc.	C. Non economic activity	<input type="text"/> <input type="text"/>
		If not engaged in any activity write "00" for that activity. The sum of all duration should be 52 weeks.		
		<b>K.5.2</b>	<b>WHETHER SEEKING EMPLOYMENT:</b> Whether took any action to find elf employment or paid employment in the last 12 months.	1. Yes
	<b>K.5.3</b>	<b>PRINCIPAL NON ECONOMIC ACTIVITY IN THE LAST 12 MONTHS:</b> If engaged in non economic activity (i.e. K.5.1C # 00) circle the main activity. If not go to K.5.4 1. Student 2. Household work 3. Income recipient/ Pensioner 4. Unable/too old to work 5. Child not attending school 6. Other non economic activity	1. Student 2. Household work 3. Income recipient/ Pensioner	4. Unable/too old to work 5. Child not attending school 6. Other non economic activity
	<b>K.5.4</b>	<b>PRINCIPAL OCCUPATION OR KIND OF WORK DONE IN THE LAST 12 MONTHS:</b> If did not work at all (i.e. K5.1A = 00) GO TO  State clearly exact occupation or kind of work done during the last 12 months, eg. Paddy Cultivator, Tea Plucker, Accounts Clerk, Mathematics Teacher, Vegetable Seller, etc.  If the person had more than one occupation/work, the principal occupation/work on which most time was spent.	..... ..... <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	
	<b>K.5.5</b>	<b>INDUSTRY, BUSINESS OR SERVICE.</b> Describe the kind of products made or services rendered at the work place of the principal occupation.  Write the name of the work place. eg. Paddy Cultivation, Garment Factory, Grocery Shop, Education Department, etc.	..... ..... <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>	
	<b>K.5.6</b>	<b>EMPLOYMENT STATUS AND SECTOR.</b>	1. Govt. 2. Semi Govt. 3. Pvt sector employee	4. Employer 5. Own a. w. 6. Unp f.w.
		1. Government employee 2. Semi Government employee 3. Private sector employee	4. Employer 5. Own account worker 6. Unpaid family worker	

248. In Example K.5, the unit of measurement was “weeks”, and the sum of weeks in the three categories (employed, unemployed and inactive is 52.

249. With respect to the time spent as “unemployed”, one would need to clarify whether the criterion of available for work, seeking work, or both was/were being applied. Although it could easily be assumed that all the questions have a 12-month reference period, the question on status in employment do not state the reference period, and could be interpreted as the status in employment for the present job, rather than for the main job.

**E Suggestions for enhancing effectiveness of questions**

250. Example K.6 is a hypothetical example that illustrates many of the points which have to be considered in designing questions on this topic.

**Example K.6**

<p><b>K.6.1 During the last twelve months up to the end of last month did you-</b></p> <ul style="list-style-type: none"> <li>• Do any work of any kind on your own or the family farm, cattle post/Kraal, or other agricultural holding e.g. Chicken house, vegetable plot etc.?</li> </ul> <p><b>Prompts:</b> The work can be for cash income or for your own food. Including any form or house construction or major maintenance</p> <ul style="list-style-type: none"> <li>• Conduct any type of business big or small e.g. brew beer, sell cakes, make hats/clothes, etc.?</li> <li>• Help unpaid in a family business of any type?</li> <li>• Catch/Collect fish, shells or other sea or river products? Collect any water or firewood?</li> <li>• Make anything from your farm or natural products for sale or for your own use?.</li> <li>• Do any type of wage or salary job - Full time or Part time, Temporary, Casual,, Piecework or Permanent</li> </ul> <p><b>Interviewer: - Prompt on each group with examples</b></p> <p><b>K.6.2:</b> How many weeks in the last twelve months were you doing any of these activities in total?[AA]</p> <p><b>Interviewer: Include paid leave or paid training as work. Also include short absences when sick or temporarily absent from work.</b></p> <p><b>If AA = 52 Weeks Skip to next Section</b></p> <p><b>K.6.3:</b> When not working were you available for work if given the opportunity?</p> <p><b>K.6.4:</b> How many weeks were you available for work in the last twelve months?[BB]</p> <p><b>Interviewer:- If AA + BB = 52, skip to the next section</b></p> <p><b>K.6.5:</b> What were you doing when not working?</p> <table> <tr> <td>Student .....</td> <td>1</td> <td></td> </tr> <tr> <td>Housework.....</td> <td>2</td> <td>CODE</td> </tr> <tr> <td>Not working and</td> <td></td> <td></td> </tr> <tr> <td>Disabled .....</td> <td>3</td> <td></td> </tr> <tr> <td>Sick.....</td> <td>4</td> <td></td> </tr> <tr> <td>Retired/aged.....</td> <td>5</td> <td></td> </tr> <tr> <td>Income recipient..</td> <td>6</td> <td></td> </tr> <tr> <td>Other .....</td> <td>7</td> <td></td> </tr> </table>	Student .....	1		Housework.....	2	CODE	Not working and			Disabled .....	3		Sick.....	4		Retired/aged.....	5		Income recipient..	6		Other .....	7	
Student .....	1																							
Housework.....	2	CODE																						
Not working and																								
Disabled .....	3																							
Sick.....	4																							
Retired/aged.....	5																							
Income recipient..	6																							
Other .....	7																							

251. The sequence of questions in Example K.6 is complex but would give labour policy analysts and national accountants quantitative measures of the number of weeks in each usual economic activity category as well as the allocation of the population to the overall usual activity categories. The number of weeks unemployed is a particularly useful measure in countries with a seasonal unemployment problem. The three categories

usually employed, usually unemployed and usually not economically active can be derived later during tabulations, i.e., from the following algorithms: the decision-making criteria are as follows:

- (a) Usually economically active and usually employed = AA + BB equal to or greater than 26 and AA > BB
- (b) Usually economically active and usually unemployed = AA + BB equal to or greater than 26 and AA < BB
- (c) Not usually economically active = AA + BB less than 26 weeks

Where AA refers to the number of weeks employed; and BB refers to the number of weeks when available for work (i.e., unemployed). This sequence also makes it possible to classify the population according to different patterns of activities during the year, i.e., by mixtures of employment, unemployment and inactive. For many users, statistics describing such patterns may be even more interesting than statistics by usual activity.

252. A series of prompts are used for the first question (K.6.1). This is in line with the earlier recommendations and covers forms of 'employment' that are usually difficult to identify. Such prompts would need to be varied to suit the situation in a particular country. An alternative would be to use a prompt card, similar to the one suggested in Diagram 2 for current activity, but referring to the last twelve months.

253. The unemployment question (K.6.3) uses the 'availability' concept, but this should be subject to testing and 'actively seeking work' can be substituted if required. A question of this type would need thorough pre-testing. One major problem will be space on the questionnaire, but this concern needs to be weighed against the value of the information that would be obtained. This question has been worded "Did you..." to encourage interviewing the respondents in person. However, it is recognized that interviewers often will have to change the question to "Did ... do any of the..." when interviewing through another informant. It is better generally not to assume the use of informants and word all questions with "you". Interviewers should be able to manage the change necessary for the informant situation.

254. The basic categories of usual economic activity should not be combined with other variables, e.g. 'status in employment', as these will most likely change over the course of 12 months for some groups of people with changing jobs. Respondents would then need to make a decision on the 'main' status (or 'type of work' or 'type of activity') without the necessary guidance. . If it is thought essential to have the 'status in employment' or other variables for the usual activity, it is suggested that the necessary information should be collected with separate questions (e.g. as in Example K.5).

255. There are several possibilities of more complex questions. Once questions on separate activities are posed, as in K.6.1 above, there will often be demands, from national accountants in particular, that the number of weeks spent on each activity is recorded. Besides taking time and space this can be very difficult as often several activities are done in any one week or month and it would take considerable time and

effort to discuss and work out a ‘major’ activity in terms of time spent each week on each activity. However, it may be possible to allow a simple ‘Yes/No’ answer in a separate column for each of the prompts in K.6.1. This could supply a measure of the effect of each prompt and assist in achieving comparability if a briefer question had been asked in a previous census.

256. There are other possible extensions that could be asked and recorded for these basic usual activity questions: e.g., ‘Which is the major work activity in terms of time spent?’ ‘For all agricultural activities, were they mostly for cash or home consumption?’ However, it is strongly recommended that these more detailed measurements, which would be very useful for many types of analyses at the national level as well as for smaller regions, should be collected by household surveys or by detailed case studies because in such surveys more time is usually available with better trained interviewers and detailed checks can also be made. The overall accuracy of the census could be jeopardised by trying to collect too much. Some additional detail on current activity, where the reference period is shorter, and there is less of a problem of multiple activities, could be obtained.

257. As noted earlier it is strongly recommended that initial questions on usual activity be relatively restricted until experience is gained with the concepts.

**Example K.7**

ACTIVITY STATUS DURING THE LAST YEAR		
P E R S O N	Usually active	a) All the time 1. Employed 2. Unemployed b) Most of the time (26 weeks or more) 3. Employed 4. Unemployed 5. Employed or unemployed
N o.	Not usually active	c) Not active all/most of the time 6. Student 7. Home worker 8. Disabled/Sick 9. Retired/Income recipient/Aged 10. Other reasons
1.		Code ____
2.		Code ____

258. In a situation in which emphasis is on current activity but a rough measure of usual activity is required for cross-analysis, the simple question in Example K.7 could be placed after the questions on current activity. Its advantage is that, without requiring the

respondent to give a specific length of time during the long reference period for the different activity statuses, one can arrive at classifying the population into the usually active/inactive statuses. However it assumes that respondents are already familiar with the notions of employment and unemployment from the earlier questions on current activity or that this could be explained to them by the interviewer. It may be necessary to do some prompting for employment activities that are difficult to recognize. Note, however, that the information collected is not sufficient to determine if the usually active person is usually employed or usually unemployed.

259. As already noted, it is very difficult to ask a simple question to get good measures of usual activity is very difficult. Given the pressure on space on census questionnaires, some countries may accept such a question or something similar but they should be aware of its serious limitations. In any case, it would need to undergo thorough testing before trying to use it.

#### *F Issues of measurement*

260. Few countries have asked usual activity questions in censuses in the past, and only few countries have tried them in surveys. Thus experience with the topic is limited. It is strongly recommended that the first introduction of the question(s) only be done after thorough pre-tests, and that current activity should be the major target for the economic questions, also for reasons of comparability with other censuses and surveys in the country. Most characteristics of the economically active population (e.g., occupation, industry, status in employment) should thus be collected only for current activity. However, in countries such as Brazil, where the census already has used usual activity as the major concept in prior censuses, and where the users of the statistics are satisfied with the results, one should of course continue with these measures.

261. All countries should at least consider the possibility of including a usual activity question. Census planners should work closely with their national accountants and Manpower/ Labour Policy colleagues to establish a minimum desirable need for measures of usual activity. It is strongly recommended that any question on usual activity be put separately either before or after questions on current activity. Respondents will be very confused if they are asked to switch from a week/month reference period to a year period and then back to a week/month. It would however help if such questions came after a block of current economic activity questions, which probed on the categories and explained them.

262. Landscape layout works best with relatively simple questions and, as noted earlier, asking a simple question on usual activity is very difficult. It may nevertheless be possible to design a question about 'main usual activity' with interviewer training and manuals stressing all the problem categories and the rules to apply. Some notes on the questionnaire would definitely help. It may also be possible to ask for the number of weeks employed and unemployed in the last twelve months. (The number of weeks not economically active should be calculated as a residual.) The usual activity prompt card should differ from the current activity prompt card, for example when activities are very



seasonal and certain activities dominate at census time and others during other months of the year. Some countries may wish to try such questions to obtain a rough measure of usual activity.

263. Pre-testing will be particularly important for this topic to validate:

- (a) The use of calendar year or twelve months as the reference period;
- (b) Use of weeks or months (or even days) as a counting unit;
- (c) Use of a prompt card;
- (d) Use of actively seeking work or available for work as unemployment questions; and
- (e) General coverage of difficult activities, e.g., unpaid work, part time-work etc.

264. Pre-tests must be well designed to be effective. A quantitative method of testing would be to use one question for half a large sample and the alternative for the other half and then analyse the differing results. Two matching samples plus a good knowledge of anything likely to cause other differences would be necessary. Another approach would be to try the alternative questions on matching representative groups of the population and then subject all the respondents to a more in-depth re-interview on their activities over the reference period and their problems with the first question(s). This would also be followed by quantitative analysis. Note again the need to make the objective of any test clear and to ensure that the results are well documented for future planning.



**PART 3: MEASUREMENT OF  
CHARACTERISTICS OF JOBS,  
ESTABLISHMENTS AND PERSONS**



## Chapter 6. Descriptive Characteristics For The Main Job

265. Once the ‘Currently economically active’ persons have been identified, the remaining economic topics only refer to this subset of the total in-scope census population. There are two major issues to be decided: which ‘job’ or ‘jobs’ associated with the person should be the focus of the questions, and whether to put these questions to only those currently employed or to both the currently employed and the unemployed. The relevant characteristics of a job include ‘status in employment’, ‘occupation’, and place of work.

266. *The Reference Job:* Employed persons can and do sometimes have more than one job in any reference period. In general, the reference job should be chosen as the job on which most time is spent during the reference period. This is usually not difficult for the short reference period for ‘current’ employment activity. However, problems may arise with persons temporarily absent from a ‘main job’ in the reference period, and with selecting a previous job for the ‘unemployed’. The *Principles and Recommendations, Revision 2*, does recommend that jobs from which the person is temporarily absent should not be considered as the main job if the person is employed and at work in another job during the reference week. It is not stated, but implied that if the person does not have any other employment in the reference period, then the job from which s/he is absent should be the reference job, and it has usually been found that most economically active persons do report about any job from which they are temporarily absent and expect this to be recorded. This applies particularly to wage earners on leave, etc., but is also very common for the self-employed. For those identified as ‘unemployed’ the most common practice is to make reference to “the last job” or to “the main type of job” over a recent, longer period, e.g., last year.

267. Census planners need to review this issue for their own country and decide what procedure to follow. If it is decided to use the temporarily absent job as the reference job in the cases where the person is at work in another job during the reference period then a note will be needed on the questionnaires and/or manuals. If secondary activities are also being recorded, interviewers should be instructed to include any other current activity as a secondary activity. This will also affect the way “hours worked” for the ‘main job’ is recorded and clear guidance needs to be provided. It should be noted that systematic exclusion of temporarily absent ‘jobs’ could cause major problems of comparisons with other employment data e.g., from employers, as well as some anomalies. This can particularly arise when a census is conducted in school holidays, as is the case in some countries, e.g., when teachers are given the jobs, temporarily, as enumerators and supervisors for the census.

268. It is essential that once the ‘main job’ is identified all characteristics should refer to this same main job in all subsequent questions. The questionnaire design must be such as not to confuse respondents or interviewers on this point. When any secondary activities are identified they must be clearly separated and placed at the end of the set of questions relating to the main job, to avoid any confusion. This is usually not difficult with current activity, but can be quite difficult with usual activity.

269. *Employed persons only, or employed plus unemployed:* The *Principles and Recommendations, Revision 2*, does recommend the collection of employment characteristics for the unemployed as well as the employed. For clarity such questions for the unemployed should be a separate block of questions from those for the employed. However, because of pressure of space on census questionnaires in some cases only a simple note has been inserted to tell the enumerators to ask the same questions for the unemployed as for the employed, but for the unemployed to refer to the last job they held. This has led to some confusion and relies on uncertain interviewing skills to change the question for differing situations.

270. As noted in Chapter 4 on the ‘unemployed’, the numbers of unemployed recorded in censuses is often relatively low in developing countries, and concentrated in the younger age groups in urban and peri-urban areas. Very few of such persons have had previous job experience and what little experience they have had is usually of limited relevance for the longer term economic or labour market policy planning and implementation for which census results are most useful. In some countries a ‘job they are seeking’ approach has been adopted for the unemployed. However, it has been shown that few ‘unemployed’ have a clear idea about the exact type of job they are realistically seeking due to their limited experience. Thus very often the most appropriate answer to a question on the characteristics of the job they are seeking would be ‘any job’ and this is also of limited use for the formulation or implementation of economic or labour market policies. For these reasons it is generally thought that in censuses, questions on the characteristics of the last job of unemployed persons should be severely limited. Such information is better collected in specialised household surveys where more in-depth questioning is possible.

## **A Status in employment**

271. Status in employment is one of the more common economic topics in population censuses and one of the most important. To know the number of persons working for others (‘employees’) or working for themselves (‘self employed’) is very important for the understanding of the structure and functioning of any labour market. The variable ‘status in employment’ was recorded in 66 of the 181 countries that reported their results to the United Nations for censuses in the 1975-1984 period {PEMBER COMMENT: UNSD TO UPDATE PLEASE}. This type of information was also often collected only partially in several other countries with mixed questions. The most recent and thorough discussion of the topic was made at the 15th International Conference of Labour Statisticians (ICLS) in 1993. The resolution as passed by the 15th ICLS<sup>71</sup> gives a much stronger conceptual base for the classification and deals with many of the problem situations that arise when using the classification. Broadly ICSE-93 classifies jobs according to the type of economic risk the job entails and the control that the person in that job has over the establishment and other workers.

---

<sup>71</sup> See references - ILO (1993a).

272. The ICSE generally is applied as a classification of the status of the jobs of employed persons only. If an employed person has more than one job and only one job is to be recorded and classified (as is the case for most censuses) then it should be the main job, that is the one in which the most time was spent during the reference period.

**1. *The operational definition***

273. Status in employment refers to the status of an economically active person with respect to his or her employment, that is the type of explicit or implicit contract of employment with other persons or organisations that 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 that the person has or will have in the job.

**2. *Requisites for applying the definition***

274. As indicated in Chapter 4, there is a major distinction between 'paid jobs' and 'self-employment jobs'. The key points are: (i) the paid jobs have some form of supervision and a fixed agreement of some sort regarding the amount of payment in cash or kind; while (ii) the income for self-employment jobs depends solely on the profits (or expectation of profits) of the activity which they undertake or the enterprise which they control.

275. The International Standard Classification of Status in Employment, ICSE-93 specifies six broad groups:-

- (a) Employees
- (b) Employers
- (c) Own-account workers
- (d) Members of Producers' Co-operatives
- (e) Contributing family workers
- (f) Workers not classifiable by status

276. The groups are virtually the same as those that were used before, except that behind the change of terminology from 'unpaid family workers' to 'contributing family workers' there is a modification of the definition of this group. The distinction between the first two sub-groups of the self employed jobs is according to whether the enterprise which the person singly or jointly controls, has paid employees on a continuous basis ('Employers') or not ('Own-account workers'). Note that the distinction is based on the presence of employees on a continuous basis: i.e., an employer must be employing and paying someone on a continuous basis not only ad hoc and for short period e.g., to solve a particular problem or workload. Payment may be in cash or kind (e.g., food, clothing, accommodation, or in some cases training only).

277. 'Contributing family workers' are by definition unpaid and usually live in the same household and are related to other self-employed family members who actually control the enterprise as either 'employers' or 'own-account workers' in this enterprise. However this restriction on living in the same household can be relaxed.

278. The last group of self-employed comprises 'members of producers' co-operatives' where the members jointly determine the organization of work and the distribution of proceeds of the enterprise. Members of producers' co-operatives' may be rare in many countries and this category is not used in censuses in countries where they are thought to be unimportant numerically. If used, care must be taken to ensure that a person merely selling or buying goods through a sales or retail co-operative, is not mistakenly categorized as members of a producers' co-operative. The person must actually work as a member of the co-operative producing goods or services. Employees of any form of co-operative should be excluded from this group and be classified as employees.

279. The ICSE-93 also provides for a number of additional, more detailed, status-in-employment categories, some of which may be both important in the country and possible to identify through separate response categories on the census questionnaire. It is strongly recommended that during the preparations for the census the need to do so is discussed with potential users of the census results and the feasibility is tested carefully. Among the possible categories one may expect users to be particularly concerned about e.g., 'owner-managers of incorporated enterprises' because they in many respects resemble employers; casual workers and seasonal workers; as well as outworkers (home-workers).

### **3. *Review of questions used in national censuses***

280. Two broad types of questions on employment status have been included on census forms. The characteristic is either combined with others, such as activity status, institutional sector, in a question or presented on its own. The following examples are discussed:

- (a) Combined with the determination of activity status (example C.1)
- (b) Combined with institutional sector (examples C2 to C4)
- (c) A question on employment status on its own (example C.5).



*Example C.1*

C.1.1		
ACTIVITY STATUS (During last seven days)		
ACTIVE:	UNEMPLOYED:	INACTIVE:
Mlimi	1. <u>Worked before:</u>	Home worker
Employee	(a) Seeking work :1A	Student
Family business worker	(b) Not seeking work :1B	Dependent
Self Employed	2. <u>Never worked:</u>	Independent
Employer	(a) Seeking work :2A	Other
	(b) Not seeking work :2B	

281. This example from comes from a landscape format questionnaire and bears the common problem of this format, limited space and quite cryptic instructions. The question does suffer considerably from the combination of the determination of current activity and status in employment in one question. It was stressed in the chapters on economic activity status that it is undesirable to combine topics in the same question. However it has been relatively common to combine the ICSE categories with the determination of economic activity status, whereby the status in employment categories are used as prompts. This has usually detracted from the accuracy of the data on both topics since, as noted in the chapter on measuring employment, a single question does not provide a satisfactory measure. Questions that start “Did you do any of the following in the last 7 days?” should be avoided. It has been shown in several tests that this immediate mention of paid employment focuses attention on this situation and some respondents then think the whole series of questions is about paid employment and answer ‘NO’, not considering whether their situation can be described by other status in employment groups. Such respondents are recorded as not employed and are therefore missing from all the following questions on occupation, industry etc.

282. The major point of interest in this question is, however, the first response category in Example C.1 called ‘mlimi’. This word means ‘working on own farm’ in the local language. Similar categories have been used in a few other countries and this technique can have major advantages. Respondents in countries with large traditional agricultural sectors rarely recognized themselves as being self-employed, employers, own account workers or in business in either the English wording or in local translations. However they usually have no problem recognizing that they were working on their own farm or ranch etc., but it will be necessary to ask another question to determine if they regularly employ anyone or not. Without this follow up, the employer category is bound to be underestimated with this kind of question.

283. No attempt is made to split the household members into own account worker, employer and contributing unpaid family helpers in two of the main categories, i.e.,

“mlimi” and family business worker. An advantage of having separate categories for these forms of employment is that one can then get a relatively clear count of those who are self-employed outside agriculture, and of the number of contributing family workers in these non-agricultural business activities. In the past, trying to make this distinction within a family often caused problems with lengthy discussions about whose farm it was, who owned the land, etc., and what to do if the head of household or farm owner was absent, etc. In reality such farms are family farms with all members contributing and sharing the risks and rewards of their labour. In these cases the separation into, e.g., own account worker and contributing family worker is not correct from a conceptual point of view, and the results from previous censuses were often difficult to interpret because of the different practices in the field.

284. In many countries the common meaning of ‘business’ does not include traditional agricultural activities. However, commercial agriculture is understood to be a business, as is the buying and selling of agricultural products. These are suitable interpretations also as the basis for economic policies. Thus, for example C.1, business activities with employees, and without employees would be a better formulation on the questionnaire than, e.g., using the ICSE-93 terminology. This illustrates the need to adjust to local situations and to use local terminology. Examples should be given if necessary.

***Example C.2***

<p><b>C.2.1</b> Were you self-employed or working for someone else in your (<u>main</u>) job last week?</p>	
<p>Self-employed</p> <p><input type="checkbox"/> With paid help (Employer)</p> <p><input type="checkbox"/> Without paid help</p>	<p>Worked for someone else</p> <p><input type="checkbox"/> As Government employee</p> <p><input type="checkbox"/> As employee of a Foreign Government</p> <p><input type="checkbox"/> As employee of private company/person</p> <p><input type="checkbox"/> As unpaid worker in family business/farm</p> <p><input type="checkbox"/> Not stated</p>

285. Combining status in employment with institutional sector seems to be an approach which works well for a census. The question is a reasonably standard status in employment question with paid employees split into two government categories and private for a sector classification. Unpaid worker in family business or farm is listed as part of working for someone else, even if the corresponding ICSE-93 category is part of self employed. It may have been shown in testing that respondents recognised the category better under the ‘worked for someone else’ heading. For tabulation and publication of the results they may easily be regrouped. The question does tend to use ICSE terminology and use of local wording might have been better, as mentioned above.

**Example C.3**

**C.3.1** Was ... *Read list. Mark (X) one box.*

1  Employee of PRIVATE FOR PROFIT company or business or of an individual for wages, salary, or commission.

2  Employee of a PRIVATE NOT FOR PROFIT, tax-exempt, for charitable organization.

3  Local or territorial GOVERNMENT employee (territorial/commonwealth, etc.)

4  Federal GOVERNMENT employee.

5  SELF-EMPLOYED in own NOT INCORPORATED business, professional practice or firm.

6  SELF-EMPLOYED in own INCORPORATED business. Professional practice, or farm

7  Working WITHOUT PAY in family business or firm.

286. The question provides more of the sector categories than usual. It does assume that respondents do know if their workplace is a ‘private, for profit’ company or a ‘private, not for profit’ organization. It is strongly recommended that respondents’ understanding of any such categories be thoroughly tested before being introduced. Government employees as a group is separately identified, but self-employed is not categorized by whether with or without employees. The distinction between employment and own-account workers has policy application in enterprise promotion and equal opportunity and gender issues. Instead, the self-employed are categorized by whether their business is incorporated or not. These categories have other uses (including as a component in the identification of informal sector enterprises – see Chapter 7, Section C on the informal sector), but the full ICSE-93 classification could not be produced from the response categories given here.

**Example C.4**

P 21	EMPLOYMENT STATUS AND SECTOR:		1.Govt.	4.Empl.
C.4.1	1 Government employee	4. Employer	2.Semi	5.Own
	2. Semi government employee	5. Own account worker	govt.	acc.wor.
	3. Private sector employee	6. Unpaid family worker	3.Pvt.sec.	6.Unpaid
			employ.	fam.wor

287. In this question, C.4.1, institutional sector of employment is also combined with employment status, but with a simple breakdown for the employees only. Three categories of institutional sector are used. Proxy respondents will most likely know if the person about whom they are reporting is in government or private sector. It may however be difficult for them to determine if the individual is in semi-government. Terms like “own account worker” will generally have to be explained so that the distinction with employer is clear.

**Example C.5**

<p><b>C.5.1</b></p> <p><b>Employment status</b>  Part-time workers are classified as Employee,  <i>Self-employed</i> (includes a proprietor of unincorporated  business or a person working on his/her own account).</p>	<input type="checkbox"/> Employee <input type="checkbox"/> Director of a firm or corporation <input type="checkbox"/> Self-employed, employing others <input type="checkbox"/> Self-employed, not employing others <input type="checkbox"/> Family worker <input type="checkbox"/> Doing home handicraft
--	--

288. The questionnaire is in ‘portrait’ format with separate columns for each person. As can be noted, “Directors” have been identified separately, illustrating the earlier point on owner/managers of incorporated companies (see sub-section 2). The other interesting variation is the separate identification of ‘Doing home handicraft’, which is presumably significant enough to warrant separate identification. This would cover both the production of such goods for sale, and the production for home use included within the SNA production boundary. To avoid any mix-ups “family worker” should be qualified as “unpaid” and make this category equivalent to the ‘contributing family worker’ category of ICSE-93, otherwise it should be clearly explained in the field manual. To revert to the standard categories in the ICSE-93 the answers could be combined, at the data processing stage, with the appropriate categories, the handicraft workers included with the own account workers and the directors classified as employees or employers. An assumption would have to be made that the directors’ enterprises employ at least one person on a continuous basis (for the director to be classified as employer), and that the home handicraft workers do not employ anyone on a regular basis (to be classified as own-account workers).

## **B Suggestions for enhancing effectiveness of questions**

289. The major point to note about a question on status in employment is that the ICSE categories as worded do not have to be given in the census and the order of the ICSE categories does not have to be strictly followed. The aim is to be able to derive these categories from the replies to the question or questions. Not recognizing this creates a major problem as considerable effort may be wasted on training in the meaning of self-employment, own-account worker, etc. Interviewers then in the field often have used these words as prompts to respondents who did not understand the terms. Time will be wasted with lengthy explanations but there will still be confused replies from respondents (e.g., self employed in Government jobs). It would be far better to use locally recognized terminology.

## **C Issues of measurement**

290. This topic has been included frequently in the past but it is strongly recommended that tests be carried out in the light of the recommended wider interpretation of economic activity and some of the suggestions above. There are several groups of jobs on the margin between employees and self-employed, which pose problems for correct classification. Some of these are discussed below and census planners should identify those that are common in their country and decide whether they should be identified separately and how they should be classified. Some research is often necessary, and field

manuals and the training of field staff should cover the commonly found situations.

291. *Owner managers of incorporated companies:* These persons are usually paid a salary by the enterprise in which they or members of their family also own a controlling part of the shares. They may also receive a part of the profits through the shares they own or may purchase shares at a favourable price (e.g., through stock options). They act as the top managers of their enterprise and thus exercise functions similar to those of employers. Because they receive wages as remuneration for their work and not for the capital investment, they are for national accounts purposes defined as employees. ICSE-93 recommends that they be identified separately if possible, to allow users of the statistics to choose whether to group them as self-employed or employees, because other users of these statistics will tend to prefer to group them with the self-employed. If the separate identification of persons in this group is not possible, then it is recommended that they be recorded as self employed.

292. *Outworkers:* Common for a significant number of persons, particularly women, in many countries, the outworkers usually work from their homes, or at another location away from the factory, with some agreement with a factory or entrepreneur who supply the materials and/or tools worked with. The outworker agrees to complete the set task with the materials supplied by a set time for a set payment. When such written or oral contracts exist, the outworker is in effect a paid employee similar to a factory-based pieceworker. They have little control over their work and do not share either the risks or rewards of the entrepreneurial part of the activity. Note that they should not be confused with other home-based workers who buy the materials to work with and bear the risk of selling the product to others. Such home-based workers are in effect self-employed, even though they may have a continuing and long-term relationship both with suppliers of raw materials and buyers of their end product. With no formal agreement or guarantee on the supply of raw materials and the sale of the product these home-based workers are bearing the risk and gaining the reward of self-employment. Often other family members become involved in the activity either as partners or as unpaid helpers. The sub-contracting of part of the work to others is also quite common. Research needs to be done in each country to determine the situation(s) of any home-based workers, how to identify such activities and, if necessary, how to differentiate the paid employee work from the self-employment work. This must be stressed in manuals to ensure that not only are these activities recorded as economic activities, but also that all field staff classify the activities in the same manner for status in employment.

293. *Business Partners:* Workers may enter into a partnership with another person either long term or just for a short-term contract for a particular job. These arrangements should pose no problem. Since the persons in such an arrangement are sharing the risks and rewards of the business activity each should be recorded as self-employed, and classified as employer or own-account workers depending on whether they have paid employees on a continuous basis.

294. *Commission workers:* These workers are often paid a small basic retainer and then a commission based on their sales. Usually the basic retainer in such cases is very small

and the major component of the person's remuneration is based on commissions on sales. Such person should usually be recorded as self employed. They are bearing most of the risks and gaining the rewards of the job. Another group of workers in a similar situation as commission-based workers are workers who rent factors of production such as land, a piece of equipment, a place of operation, etc., for operating a business (e.g., taxi drivers, street hawkers, hair dressers, shoe shine boys). These workers often see themselves as employed by the owners of the assets they rent. However, since they are bearing the risk of production and generating income they should be considered self-employed and not employees. This should be stressed in manuals, and illustrative examples used in training in countries where such work arrangements are common.

295. In each country it is likely that there will be other categories which are either important or problematic, or both. Their classification should be based on the degree of risk the person is exposed to for the major part of his/her remuneration from the job. If the person bears most the risk (and gets the reward) then the person is in self-employment. If the enterprise takes most of the risk and pays most of the remuneration then the person is a paid employee. Some additional marginal cases are discussed in the ILO references.<sup>72</sup>

296. It is not unusual for data users to want the census to provide them with sub-classifications under the major general groups of ICSE-93. In addition to those already mentioned some of the more common requests are:

- (a) Paid employees by whether permanent/regular or 'temporary/casual/piece work'. These types of sub-classifications of employees are also specifically mentioned in ICSE-93. In ICSE-93 it is stated that employees with stable contracts, or regular workers could be separately identified. Census planners will have to decide whether there is sufficient demand for such sub-groups and which groups.
- (b) Self employed by whether in agricultural activities or not, and those in agriculture by whether the agriculture is mostly for cash or for home consumption. These sub-classifications of self-employed of particular interest to national accountants, in some countries even further subdivisions have been made.<sup>73</sup> However, it is generally recommended that such detailed sub-categorizations should not be collected in censuses but through special surveys or case studies.

In the case of the self-employed, cross-tabulation of the status in employment variable with the appropriate agricultural categories of the industry variable (discussed in the next chapter) can give an approximate distribution. Since a complex set of questions is normally needed to establish the other distinctions, these issues are better left for labour force or other household surveys.

---

<sup>72</sup> ILO (1988, 1993a) and Hussmanns and others (1990).

<sup>73</sup> See Bain D. (1996).

297. Some aspects to be investigated are:

- (a) Recording for status in employment for those engaged in the broad range of agricultural economic activities now included in production using either suggestions above or with other formats;
- (b) Examples of problem groups (mentioned above) in the local situation;
- (c) Local interpretation of or expressions for employees on a continuous basis, own business, and unpaid help in family business;
- (d) Any sub-groups of paid employees or self employed if it is decided to introduce them;
- (e) Producer co-operatives – if they exist and how are they currently recorded, and is a separate category needed for this group?
- (f) Any problems with persons temporarily absent from their main activities.

298. A good critical review of the past and proposed use of statistics on ‘employment status’ groups by national accountants and labour market analysts is strongly recommended before the current round of censuses to guide the development and testing on this topic. 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".

## **D Occupation**

299. ‘Occupation’ is a topic that almost all countries have included in earlier censuses. In fact, it is the most common economic topic asked in censuses, except for the basic activity status (i.e., employed/unemployed). As the census will be the only source for statistics on the occupational profile of the whole employed population, in many countries, the occupational distribution from the most recent population census and the changes since the previous one are major components of a country's Labour Market Information (LMI) and the changes taking place in its economic and social structure. These statistics are of great value for many forms of labour market description and analysis and are needed as a basis for policy formulation and implementation. The major target of occupational questions is the employed population, but it is common also to ask about the occupation of the last job for the unemployed population. It is recommended that the question to be asked of employed persons should be separated from that to the unemployed. Interviewers will have to be relied upon to change the wording of questions if only one question is applied to both groups. This is usually confusing, strains the capability of many census interviewers and leads to poor results.

### **1. The operational definition**

300. Occupation refers to the type of work done during the time reference period by the person employed (or the type of work done previously, if unemployed), irrespective of where (i.e., the industry) or under what economic conditions (i.e., the status in employment) this work is being done.

## 2. *Requisites for applying the definition*

301. The 1988 International Standard Classification of Occupations (ISCO-88, to be updated in 2008 as ISCO-08)<sup>74</sup> and national occupational classifications, developed with the same principles, are structured so that jobs requiring similar skills are grouped together. This does not mean, however, that the response to be written on the census questionnaire should provide information about these skills. The information to be recorded on the questionnaire should reflect the type of tasks and duties undertaken in the work, as this is the type of information needed to determine the correct occupation code for the respondent's job. As explained in Part 5, this information is usually most effective if it consists of a job title, supplemented with a few words to indicate the main tasks and duties of the job. With properly constructed coding tools this type of information will allow detailed, reliable and effective coding.

## 3. *Review of questions used in national censuses*

302. Three types of approaches to the question(s) on occupation have typically been used in national censuses.

- (a) (One or two pre-coded questions on the jobs that the individuals have (or had) (example D.1).
- (b) One question to obtain occupation relevant information about an individual's job (example D.2);
- (c) Two or more questions, a basic question on the title of the position held with follow-up on main duties of the individual in the job (examples D.3 to D.5).

---

<sup>74</sup> At the time of writing (September 2007), ISCO-08 had not been adopted and therefore this section refers to ISCO-88 methods and groupings.



**Example D.1**

D.1.1. Indicate your professional position for your actual employment		
• Manual worker, specialized worker	• .....	<input type="checkbox"/> 1
• Qualified worker or highly qualified	• .....	<input type="checkbox"/> 2
• Master agent supervising workers	• .....	<input type="checkbox"/> 3
• Master agent supervising technicians	• .....	<input type="checkbox"/> 4
• Technician, designer	• .....	<input type="checkbox"/> 5
• Teachers, social workers, nurses and staff of B category in public services	• .....	<input type="checkbox"/> 6
• Engineer, Professionals (Persons who do not have professional skills are not considered even if they contribute for social security under professionals)	• .....	<input type="checkbox"/> 7
• Professors and all persons of A category of the public service	• .....	<input type="checkbox"/> 8
• Employed in office, employed in commerce, agent of service, assistant nurses, nursery staff, persons of C or D category of public service	• .....	<input type="checkbox"/> 9
• Other cases. Give details:	• .....	<input type="checkbox"/> 0
D.1.2 What is your main function?		
• Production, manufacturing	• .....	<input type="checkbox"/> 1
• Installation, maintenance, repair	• .....	<input type="checkbox"/> 2
• Cleaning, caretaking, housing services (servant)	• .....	<input type="checkbox"/> 3
• Lifting, storage, transport	• .....	<input type="checkbox"/> 4
• Secretary, data entry, teller	• .....	<input type="checkbox"/> 5
• Management, accounts, administration	• .....	<input type="checkbox"/> 6
• Commerce, sales, technical business	• .....	<input type="checkbox"/> 7
• Research, study, method, information technology	• .....	<input type="checkbox"/> 8
• Managing directors or direct deputy	• .....	<input type="checkbox"/> 9
• Other cases: (Give details)	• .....	<input type="checkbox"/> 0

303. This example shows the application of two pre-coded questions. Given that very broad categories are presented in the questions, considerable training would be necessary to ensure interviewers are sufficiently aware of the contents of all the groups in order to guide respondents. It is likely that more detailed occupational categories can be derived from combining the answers to the two questions. The international classification (ISCO-88) has 10 major groups, which are subdivided into 28 sub-major groups, with 116 minor groups and a total of 390 unit groups.<sup>75</sup> {ILO COMMENT: NUMBER OF GROUPS AND FOOTNOTE 75 TO BE REVISED WHEN ISCO-08 IS FINALISED}

304. It is also important to emphasize the point made earlier, that pre-coded answers for occupation will not give the type of statistics needed by the large majority of data users. However, there may be censuses where this solution is unavoidable because of limited funding; in which case the pre-coded categories should be carefully thought out. Occupations for which information is definitely required should be given separate codes and others may be grouped together. There will be very severe limitations on the number

<sup>75</sup> International Labour Office (1990b); p. 3.

of separate codes possible, and even with good training only rough estimates of the numbers in these groups will result. Note that the resulting response categories normally will not be represented by the broad groupings presented in the occupational classification in use in the country. One should definitely not try to simply use the first digit groups of ISCO-88 or the national standard classification. The ISCO-88 was not constructed for use in this way, nor was the national classification. It is also very unlikely that the meaning behind the titles of some of the groups, e.g., ‘Associate professionals’ will be recognizable to interviewers.

**Example D.2**

D.2.1 What occupation did you perform in your principal job during the week of 23 to 29 June 2000?  
 (The criteria for determining principal job were: the most number of regular hours; the most time spent during the reference week; and the job that provides the most income).

305. In the question above (D.2.1), sufficient space was provided to record the description given by the respondent. The interviewer will need to probe to get the level of detail required since there is no indication given in the basic question what level of detail is required. Also, the use of the word ‘occupation’ will present some difficulty for some categories of workers (for example labourers and some unskilled workers), who may not be accustomed to referring to the work they do as an occupation.

**Example D.3**

To be asked of all persons aged 10 years and over	
Occupation	Name and address of workplace
a) What kind of work did ... do? (b) What are ...’s main duties at this job?  <div style="text-align: center;">D.3.1</div>	What is the name and address of ...’s workplace?  <div style="text-align: center;">D.3.2</div>
(a)..... (b).....	..... .....

306. The two questions asked are as suggested earlier. The questionnaire was in the landscape format, which usually has restricted space for both questions and answers. This is however, (in the case of occupation and of industry), less of a problem for the questions than for the answers. The questions are quite standard, although in this case the blank space in the question area could have been used for more explanation and

examples. The very limited space for the two answers is however the major problem, since it severely limits the amount of information (or description of activities) that can be provided, and is bound to affect the quality of data. The data are affected by interviewers' need to summarize the responses and then being precluded from probing for additional information.

**Example D.4**

<p>OCCUPATION</p> <p>D.4.1</p> <p>If YES to X.X</p> <p><b>What is the main occupation of (the person) in this workplace?</b></p> <p>Occupation refers to the type of work (the person) performed in the seven days before 10 October.</p> <p>Use two or more words.</p> <p>For example, street trader, cattle farmer, primary school teacher, domestic worker, fruit vendor, truck driver, warehouse manager, filing clerk, etc.</p> <p>Use CAPITAL LETTERS only.</p>
---

307. This example is from a questionnaire that was presented in a portrait format. The suggested two questions were asked, with explanation and examples given to show not only the type but also the desired level of specificity. Space was available for the two answers. The use of the word occupation in the actual questions is however not recommended. Very similar questions were also asked in a separate block for the unemployed referring to their last job, which is good practice.

**Example D.5**

<p>D.5.1 b What was your occupation?</p> <p>What are the main tasks or duties you have to perform in that occupation</p> <p>What educational or professional qualifications are required by that job?</p>	<hr/> <hr/> <hr/> <hr/>
---	-------------------------

308. This example has an interesting variation — a third question on the qualifications needed for the job. This information may have been useful for coding to the national classification, but note that unless this information was used with great care (as described in Part 5) it may have contaminated any analysis of the connection between people's qualifications and the type of work they are doing. It is not known how the responses to this element were used in the coding process. The term “occupation” is used in the first question without explanation or examples. The other two questions complement the first question, but all the questions would function better with some examples.

#### 4. *Suggestions for enhancing effectiveness of questions*

309. Questions on occupation have been asked for many years in population censuses and there has been a considerable amount of research carried out on how to best formulate these questions and code the responses. Most of this research has been done in English-speaking developed countries, but much is still likely to be relevant for developing countries and other languages. Some of the major findings are:

- (a) The word “occupation” can be misleading and may be best left out of any actual questions on the topic. For developing countries where translation in the field is very common, the terminology being used by interviewers in the local language should be carefully checked during testing and training periods. As always, the aim is to get the meaning of what is being asked across to the respondent, consistently, rather than insisting on the terminology in the language used in the questionnaire.
- (b) Space allowed for recording has been shown to be a major factor in poor descriptions of jobs and thus poor coding. In some countries good questions have been asked, but the space allowed for recording the response has been very limited.
- (c) It is most effective if two separate questions are asked - A description of the job and then a question on the tasks or duties actually performed. A combined question has proven to be far less effective.
- (d) The quality of occupational descriptions is quite poor when given by an informant other than the jobholder. Every effort should be made to interview all employed respondents in person for these questions, but this is clearly difficult to achieve in particular in a large-scale exercise like the census.

#### 5. *Issues of measurement*

310. Using pre-coded response alternatives with the occupational question(s) has the major attraction that it is by far the least costly solution, as one does not have to code the answers after interviewing. However, there are major disadvantages to pre-coding. Occupation is a complex topic and even with a very well thought out set of response categories and with excellent training there is great difficulty for interviewers (or the respondents themselves) to place a job in the correct category. Both the validity (correct categorization) and reliability (same categorization made by different interviewers of equivalent responses) of pre-coded occupational categories has been shown to be very poor. In addition, only a limited number of categories are possible with pre-coding and thus the details needed by many users of such statistics cannot be provided. Pre-coding is therefore not a reasonable alternative for this major variable in a census. The expense of a coding operation may therefore be well justified in light of the quality and relevance of the data that are generated.

311. *Agricultural activities* are often relatively easy to describe, e.g., with terms like ‘vegetable farmer’, ‘maize or rice farmer’, ‘goat raising’ or ‘cattle rancher’. A brief description of the actual tasks done in the reference period can be given with two or three words as response to the second question or second part of the question. A problem arises when the classification makes a distinction between those for whom the same activities are mainly for sale (referred to in ISCO-88 as market-oriented skilled agricultural and fishery workers) or mainly for home consumption (referred to in ISCO-88 as subsistence agricultural and fishery workers), which most classifications based on ISCO-88 do if subsistence farming is of importance in the country. To make this distinction may require an additional question, as it cannot be expected to be reflected spontaneously in the reply given by the respondents. A possible supplementary question may then be: “Do you produce mainly for sale or is it mainly for your own or family use?” As a general rule such questions should actually be written into the questionnaire and not only given as notes to interviewers. In developing countries with a large number of farmers, it has generally been found that interviewers have been able to manage this addition to the occupation question with reasonable accuracy and that they usually have a good understanding of subsistence farming. More precise estimates or complex subdivisions such as ‘all cash crops/part cash/all subsistence’ are best done by separate questions for this group or preferably in specialized surveys or the agricultural censuses rather than in the population censuses.

312. *Other non-market production*: As noted in earlier chapters there are many activities, which are regarded as economic production even if they are not marketed or exchanged. Examples are: fetching water or firewood, processing of farm or natural produce (e.g., making butter, brewing beer, extracting oil, making hats, mats or clothing, whether for sale or only for the household’s own consumption), own house repair/construction; farm construction work such as digging irrigation ditches or building fences; volunteer building of communal roads. When responding to questions such as “What kind of work do (...) do? What are the main tasks and duties?” words reflecting these activities should be recorded as for any other job, i.e., with a ‘title’ and some description of the tasks performed. However, it will help overall coding if interviewers are told to record that such activities are mainly for own use, similar to the distinction between agricultural activities which are mainly for sale and mainly for own use. If some of these activities are very common and there is no other way of identifying them separately it may be necessary to create special occupational codes similar to that for ‘subsistence farmer’. It should be noted, however, that on farms in particular, these activities are relatively rare as a main activity separate from the general farm work. They are often carried out as secondary activities of farming families, although they can be a main activity in the agricultural off-season and in non-farm settings. In countries where they are likely to occur reasonably often, it is very important that they be allowed for and interviewers should be given guidance about how to record such activities for the coding of occupation and industry.

313. Interviewer training is the key to good occupational descriptions for subsequent coding. Interviewers should be briefed on the need for accurate and relevant information about the job. They should be thoroughly drilled on adequate and inadequate

descriptions through series of examples. The examples should be local and come from the queries and problem cases in recent censuses and surveys. Descriptions such as ‘Manager – managing’; ‘Public Servant - Office work’; ‘Supervisor – supervising’; ‘Clerk - clerical work’; ‘Apprentice – helping’; ‘Labourer - manual work’; and similar uninformative responses must be stressed as inadequate or they will occur frequently. Field practice should focus on the responses to the occupation questions. Supervisors should be instructed to send interviewers back to households when inadequate responses are given. Owing to time constraint in the census, however, this may not always be possible to do. It is only with such a concentrated effort that the quality of occupational responses will improve and, with a similar effort on coding, the quality of the final statistics.

314. Even if an occupation question has been asked several times before it is worth testing for possible improvements to the question. Better wording, more examples, more ample space should all be tested. As always, the test(s) should have clearly stated objectives and be quantitatively evaluated. Any questions with pre-coded answers must be thoroughly tested to minimize errors and in order to ascertain how interviewers are going to interpret the summary groupings given, in realistic conditions, and to correct any problem.

## **E Place of work**

315. In the past, a question on the geographic location of the place of work has been included in censuses in some European and North American countries, but will have been a new topic in most other countries if included in the 2010 round of censuses. Where this topic has been included the actual address of the place of work has often been recorded when relevant and used to code the location to a fine geographic level. This has allowed detailed tabulations and mapping of place of residence by geographic location of place of work.

316. Additional questions have often been asked on the method of travel to work and these and other variables have enabled the production of statistics on important travel-to-work patterns. This information has then been very valuable as a basis for transportation planning.

317. While the information on place of work can be used to develop area profiles with respect to 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 geographic location of 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 residence to the place of work.

318. The type of place of work has also been included as a topic in many censuses and provides interesting information in its own right as well as providing supplementary information in the analysis of informal employment.

1. *The operational definition*

319. "Place of work" is the location in which a currently employed person performed his or her job.<sup>76</sup>

2. *Requisites for applying the definition*

320. As presented in *Principles and Recommendations, Revision 2*, both the geographic place of work (mentioned above) and the type of place of work is collected. For "type of place of work", three categories are sought:

- (a) Working at or from home: This includes persons whose work is directly at their home, e.g., home based work for a wage, sewing/tailoring or such professional activities as accountancy or computer programming, non-agricultural production for own use at home, including that which takes place on the farm; informal sector activities, such as production for sale at home or sales at or adjacent to the home (small shops etc.) doing agricultural work on the farm near their home. Persons staying in work camps or in military compounds as well as paid employees in agriculture living on the farm and domestic employees living with or near their employer would also be taken as 'working from home', even if they travel to another residence, e.g., every weekend would also be included here.
- (b) No fixed place of work: This category is restricted to persons who travel from place to place for their work and do not report daily in person to a fixed address as a work base: e.g., hawkers, travelling salesmen, long distance truck drivers, seafarers, fishermen and some taxi drivers. Excluded are construction workers, etc., as they should be taken as working at their current work site, if based there longer than a week, and be classified to group (3).
- (c) With a fixed place of work: All other employed persons should be included in this category, including all persons who move around in their job but have a fixed base location to which they report daily, e.g., bus/taxi drivers (with a base), train and airline staff and the construction workers mentioned above.

321. It is for those classified to the last group that a supplementary question on the precise address has been added in many countries. Note that these categories only provide a geographic reference for the place of work, and do not make the type of distinctions between the physical characteristics of the place of work, e.g., whether at a street corner, with a roof, in a shack or in a building, which may be relevant when determining whether the place of work is in the informal sector or not.<sup>77</sup>

---

<sup>76</sup> Or where a "usually employed" person currently or last performed the job.

<sup>77</sup> These issues are further discussed in International Labour Office (2002). *On Measuring Place of Work*.

322. There are no major conceptual problems with the above categories. It is however necessary to define how close to home 'Work at home' is, especially for informal sector activities. Those taking place on the same plot of land as the dwelling should definitely be taken as 'at home'. However, the definition is to some extent a function of the uses of the information and national circumstances; and in some cases 'at home' has been defined to be at some distance/time from the home. It is preferable, however, to limit this definition to the structure and area around the home. The classification of those working on farm fields that are at a distance from the home require careful consideration. Pre-testing should help to establish all such problem categories and their treatment will have to be included in training and manuals for field staff.

### 3. *Review of questions used in national censuses*

323. For those countries that have included a question on place of work, the form of the questions suggest the following objectives:

- (a) to determine if the activity is undertaken in the home, its vicinity or some other location, as in the international recommendations (example E.1);
- (b) to obtain specific information on about the actual location of the work place (example E.2), and
- (c) a combination of both objectives, (examples E.3 to E.6).

#### *Example E.1*

E.1.1 *Where is your/is ..... 's place of work located ?*  
(READ CATEGORIES)

<input type="radio"/> <i>In own home or yard</i>	<input type="radio"/> <i>On a Farm</i>
<input type="radio"/> <i>In another home or yard</i>	<input type="radio"/> <i>Not Stated</i>
<input type="radio"/> <i>Not in a private home</i>	

---

E.1.2  
4.10 *In which parish do you/does ..... work ?*

<input type="radio"/> <i>Kingston</i>	<input type="radio"/> <i>Hanover</i>
<input type="radio"/> <i>St. Andrew</i>	<input type="radio"/> <i>Westmoreland</i>
<input type="radio"/> <i>St. Thomas</i>	<input type="radio"/> <i>St. Elizabeth</i>
<input type="radio"/> <i>Portland</i>	<input type="radio"/> <i>Manchester</i>
<input type="radio"/> <i>St. Mary</i>	<input type="radio"/> <i>Clarendon</i>
<input type="radio"/> <i>St. Ann</i>	<input type="radio"/> <i>St. Catherine</i>
<input type="radio"/> <i>Trelawny</i>	<input type="radio"/> <i>More than one parish</i>
<input type="radio"/> <i>St. James</i>	<input type="radio"/> <i>Not Stated</i>



324. In this question the emphasis seems to be on working from the home, either own home or in another's home. The second question relates to the administrative area in which the work place is located. There is however no category of "no fixed location". The category "more than one parish" provides some indication of those who may work from no fixed location, but a specific category would be needed, in order to obtain a more accurate count of those working from no fixed location (such as street vendors).

**Example E.2**

PLACE OF WORK

E.2.1a & E.2.1b  
If YES to X.X

Does (the person) work in the same sub-place in which s/he usually lives?  
Y = Yes  
N = No  
Dot the appropriate box.

If NO, where is this place of work?

If NOT the same place, write PROVINCE P R ,  
MAIN PLACE (city, town, tribal area, administrative area) and  
SUB-PLACE (suburb, ward, village, farm, informal settlement).

If another country, write the name of the country in the boxes below.

325. Example E.2 presents a two-part question in which the first part seeks information about the place of work relative to the place of residence. The use of the term "sub-place" here is not specific and can be confusing as it is not clear if it means the dwelling unit/compound, locality or other geographical demarcation. However, some clarification of the term is given indirectly in the second part of the question, which is addressed to those who answer "no" to the first part of the question. The definition given for sub-place suggests that the intent of this question did not include determining the

number of people who work at home or have no fixed work location. It should not be too difficult to extend the first part of question E.2.1, to cover these categories. The reason for limiting the information about the place of work in another country to the name of the country is unclear. In areas where cross-border commuting for work is common it would be as useful to have the precise commuting patterns across the border, as it is to have them for commuting within the country.

**Example E.3**

E.3.1 Place of work or location of school

- For a person who worked and also attended school during the week, give the place of work.
- If the place is other than one's own home, write the address of the place of work or the location of the school (i.e., the name of the *to, do, fu* or *ken* (prefecture), and the *shi* (city), *ku* (ward), *machi/cho* (town), or *mura/son* (village). Do not omit the name of the *ku*, if the address is in one of the 13 major cities.
- The major cities are ....

Home (including living-in-employee)

Outside home

Same *ku, shi, machi/cho, mura/son*

Other *ku, shi, machi/cho, mura/son*

Go to question 1 X.X

(Write the address)

to, do, fu, ken

shi, gun

ku, machi/cho, mura/son

326. A format similar to the one being proposed was used in this example (E.3), where the first question related to whether the individual worked at home or elsewhere, in which case the full address of the place of work was obtained, as a second part to the question. The question used is more specific about the relative location and also clear about how live-in employees of the household are to be classified.

***Example E.4***

**E.4.1** What is (was) the full address at which you actually work(ed)?


1  Work mainly at or from home      2  No fixed place of work

327. The example given has as the main question the “full address” of the place of work, with the two categories to be checked if worked at home or at no fixed location. These questions are simple enough to be used in almost all countries, but when/where street addresses are not commonly used, then the locality and related geographical information can be used, as in example E.2.

***Example E.5***

**E.5.1** At what address did this person **usually** work?

*Example:*  
365 Laurier Ave. West  
Number     ↑     ↑     ↑     ↑  
Name         ↑     ↑     ↑     ↑  
Type            ↑     ↑     ↑     ↑  
Direction      ↑     ↑     ↑     ↑

If direction (e.g., North, South, East or West) is a part of the street address, please include it.  
If street address is unknown, specify the building or nearest street intersection.

**Note:**  
Some large cities are made up of smaller cities or towns called municipalities. Where applicable, identify the municipality rather than the larger city, for example,

07  Worked at home (including farms)  
Go to Question Y.Y

08  Worked outside Canada  
Go to Question Y.Y

09  No fixed workplace address  
Go to Question X.X

10  Worked at the address specified below:

**Specify complete address**  
Street address (see example)

11 →

City, town, village, township, municipality or Indian reserve

12 →

Province/territory

Postal code

13 →

328. The example E.5 is from a self-enumeration schedule. It is an example of the more detailed questions used in some developed countries mentioned earlier. Note however that the three recommended categories are there. In other countries the three basic categories could simply be asked without further elaboration, but in developing countries it will be essential to give examples or notes on the questionnaire for the common situations in that country. Presumably, “including farms” means residential farms. If a broader coverage is intended, it would be preferable to include a separate category for “farms” since “work at home” category is of increasing interest and value to analyst and labour legislators, especially in the context of home-based work and informal employment arrangements. Again it can be noted that the question treats work places abroad differently from those within the country.

**Example E.6**

<b>E.6.1 Branch of industry of place of work</b>	
Name and place of work ..... .....	
District <input type="checkbox"/> <input type="checkbox"/>	Commune/Municipality <input type="checkbox"/> <input type="checkbox"/>
<b>Describe in detail the economic activity at your place of work</b> ..... .....	
	NACE <input type="checkbox"/> <input type="checkbox"/>
<b>E.6.2 Type of your place of work</b>	
Fixed premises outside home	<input type="checkbox"/> 1
Home (not a farm)	<input type="checkbox"/> 2
Farm	<input type="checkbox"/> 3
Not a fixed place of work	<input type="checkbox"/> 4

329. In the question E.6.2, special attention is given to “farms” as a place of work. The question design would imply that farms which are not the person's home should be classified to the first response category and that the third response category is only for farms that are the person's home. However this should be clearer. A possible alternative might have been: “home - not a farm,”; “home – located in farm”; “farm located outside the home”; “other fixed premises outside the home”; and “not a fixed place of work”.

330. More guidance is needed in both E.6.1 and E.6.2 on the treatment of taxi drivers, tradesmen and others who operate in a mobile way from a fixed business location. As indicated earlier, these should be classified as “with a fixed place of work”.

#### 4. *Suggestions for enhancing effectiveness of questions*

##### **Example E.7 (Untested, hypothetical question)**

###### **E.7.1 Working at home**

- Include*
- *Farmers and their employees living on that farm*
  - *Domestic paid workers living on their employers plot*
  - *Home based wage jobs*
  - *Home production for own use or sale*
  - *Fetching Water and Firewood*
  - *Other informal activities based at or near the home*

###### Working at or from another fixed location (not home)

- Include*
- *All activities with a fixed location or base e.g. Bus drivers, Construction workers as well as normal shop/office/factory workers*

###### Working without any fixed location.

- Include*
- *Only activities with no fixed base*
  - *Hawkers, travelling salesmen or long distance drivers, seafarers*

331. This information may be presented on a prompt card or included on a questionnaire. For a landscape questionnaire layout the notes or examples would have to be considerably reduced because of the limited space and reliance have to be placed on the interviewers'/supervisor' manual and on training. The *Principles and Recommendations, Revision 2*, para 2.348 provides a more limited recommendation on content, which is also more reasonable given the resources for most countries' census operations.

#### 5. *Issues of measurement*

332. This is a relatively simple topic but as a new topic for most countries it will need thorough testing that focuses on problem groups (especially the group working from home) and how to assign them to the proposed categories. The response categories should be determined according to national needs.

## **CHAPTER 7. Descriptive Characteristics Of The Establishment**

333. As earlier stated, it is essential that once the main job is identified all characteristics should refer to this same main job throughout the questions that follow. The questionnaire design must be such as not to confuse respondents or interviewers on this point. When any secondary activities are supposed to be recorded they must be clearly identified separately and placed so as not to confuse the flow of questions relating to the main job. This is usually not difficult with current activity (see later examples) but can be quite difficult with usual activity.

### **A Industry**

334. As a census topic, industry is only slightly less common than occupation and has been included in several census rounds in many countries.

335. Industry is a key classifying item for most employer- and establishment-based employment and wages surveys. However, most of these surveys only cover larger formal employers and therefore a population census is usually the only possibility for comprehensive and detailed statistics of all employment in a country by industry. The census statistics on industry is thus a major basis for most countries' economic statistics as well as for their labour market and social statistics. It is essential for comparability that the industry classification used in the census, is consistent with that used in other relevant surveys. Many countries have been using in previous censuses the International Standard Industrial Classification of all Economic Activities (ISIC, Revision 2) approved in 1968, but it was hoped that all countries would have changed to a classification compatible with ISIC, Revision 4 by the time of the 2010 round of censuses.

#### **1. *The operational definition***

336. Industry refers to the activity of the establishment in which an employed person worked during the time reference period established for data on economic characteristics (or last worked, if unemployed). It describes what the establishment does, not what the individual does when working for that establishment. (But for those who are self-employed without any employees on a continuous basis, the activities of the establishments that they represent will be those that they undertake themselves.)

#### **2. *Requisites for applying the definition***

337. There are generally two components to the desired industry question(s). The first part of the question should ask for the name of the place of work with details of division or branch for larger private enterprises and government institutions and establishments. To ask for the full name and address may also be useful, in particular when access can be had to an establishment register. The second part must request a good description of the main products or the main functions of the establishment.

### 3. Review of questions used in national censuses

338. The questions that have been used in censuses are grouped under three categories:

- (a) Questions with pre-coded categories, as in the case of occupation.
- (b) (One question that asks about the type of activity undertaken by the establishment (examples F.1 and F.2).
- (c) (Two or more questions, a basic question on type of activities and one or more other questions requesting supplementary information on other characteristics, such as, place of work, or address of the workplace (examples F.3 to F.5).

339. As with occupation, the use of pre-coded response categories is not desirable, as it is very difficult for field staff to allocate employed persons to the correct industry category. In addition this strategy will result in only very limited details on industry being available for this variable. It is, however, easier to draw up a reasonably sound pre-coded list for industry based on the ISIC than for occupation from ISCO-08. The list of categories will still be severely limited and the resulting groupings are likely to be of very limited usefulness.

#### Example F.1

BRANCHE D'ACTIVITE
<p>Marquer le genre d'activité de l'établissement où travaille le recensé ou le travail que fait le recensé indépendant.</p> <p>Cette colonne est "Sans Objet" Pour les personnes de moins de 10ans Pour ces personnes inscrire 888 dans les grilles</p>
F.1.1

Branch of activity
<p>Mark the type of activity of the establishment where the respondent works or the kind of work if s/he works for her/himself.</p> <p>This column is not applicable to persons younger than 10 years. For these persons write 888 in the cells.</p>
F.1.1

340. In the column for the branch of activity the space allocated for recording activities is not adequate for providing any details. The space provides for entering a three-digit code, which implies that the enumerator is to match the description given by the respondent with the codes for the activities, select the appropriate code and enter it on the questionnaire. This approach not only prolongs the interview, but also it is prone to errors and there are no means of checking or correcting them after the interview is completed, since the actual response to the question would not have been recorded. The problems associated with interviewer coding, discussed in Part 5, are also inherent in this approach.

**Example F.2**

EN EL TRABAJO PRINCIPAL, EN EL QUE TRABAJA MAS HORAS...											
28	El lugar o establecimiento donde trabaja, ¿a qué actividad se dedica o qué servicio presta?										

**IN THE ESTABLISHMENT IN WHICH YOU WORKED THE MOST HOURS**

**F.2.1** The place or establishment where you work, what activity is it mainly engaged in or what does it provide?

---

341. The example presents one question about the type of activity (what the establishment is dedicated to or produces). Sufficient space is given to record the response.

342. A stand-alone question requesting a description of the activity is unlikely to elicit the details needed to accurately classify industry to the third or fourth digit level codes presented in ISIC. However, it is recognized that census planners may be forced, due to space limitations, to use only one question for industry, particularly with the landscape questionnaire layout. In such cases it will still be better to ask a combined question on the name of the location and the activity, rather than solely asking about the activity.

**Example F.3**

<p>COMPANY/BUSINESSACTIVITY</p> <p><b>F.3.1</b></p> <p>IF YES to F.X.X</p> <p><b>What does the business do (main economic activity)?</b></p> <p>Write the MAIN INDUSTRY, economic activity, product or service of (the person's) employer or company. For example, gold mining, road construction, supermarket, police service, healthcare, hairdressing, banking.</p> <p>OR</p> <p>Write the activity of the person if self-employed. For example, subsistence farming.</p> <p>If doing PAID domestic work in a private household, write DOMESTIC SERVICE</p> <p>Use CAPITAL LETTERS only.</p>
---

343. This is also an example of a fairly comprehensive industry question, but does not attempt a combination with place of work. It has very useful notes regarding the self employed and paid domestic work. The Manual and training for interviewers should give guidance on how to record persons working from home and persons with no fixed place of work. It would have been best to record these statements for the answers to F.3.1 in these cases. The word industry has been used. This would be unacceptable for a single question, but in this example alternates are given in case the term was not understood, and this is good practice. Question F.3.1 does refer to the person's employer or company whereas it is the actual establishment or place of work that is recommended as the reference for industry. It is uncertain how large multi-

establishment companies or similar government departments would be recorded.



**Example F.4**

<p>40.a.    i.....i           geo-code</p> <p>40.b.</p> <p>      i.....i           isic-code</p>	<p>40. Where do you work?</p> <p>work address: .....</p> <p>name of company/organization: .....</p> <p>type of activity engaged in by company/employer: .....</p> <p>.....</p> <p>.....</p>
--	---

344. The three-part question is captioned by the main question – “Where do you work”? The description of activities engaged in by the company of the employer is presented with the work address and name of the company. The information requested should be easy enough for the respondent to provide, and there are no technical terms that would require explanations.

**Example F.5**

<p><b>F.5.1</b> What is the name of your employer? Please do not use abbreviations.</p> <ul style="list-style-type: none"> <li>➤ Give the trading name if one is used.</li> <li>➤ Civil Servants - give names of departments.</li> </ul> <p>Local Government Officers -give name of employing authority.</p> <p>_____</p> <p>_____</p> <p>_____</p> <p><b>F.5.2</b> What is the main thing which the organization you work for makes or does, at the place where you work?</p> <ul style="list-style-type: none"> <li>➤ Describe clearly what the employer (or yourself if self-employed) makes or does?</li> <li>➤ Civil servants - write “Government department”.</li> <li>➤ Local government officers - give departments in which employed.</li> </ul> <p>_____</p> <p>_____</p> <p>_____</p>	<p><b>F.5.3</b> What is the address of the place where you work?</p> <ul style="list-style-type: none"> <li>➤ If you are not reporting a fixed place tick “no fixed place”.</li> <li>➤ If you work mainly at home, tick “mainly at home”.</li> <li>➤ If you are employed on a site for a long period, give the address of the site.</li> <li>➤ If you do not work regularly at one place but report daily to a depot or other fixed address, give that address.</li> </ul> <p><input type="checkbox"/> No fixed place</p> <p><input type="checkbox"/> Mainly at home</p> <p>Please remember to give the Postcode</p> <p>_____</p> <p>_____</p> <p>_____</p>
--	---

345. The above questions cover both industry and place of work topics. They show how the problematic groups mentioned earlier, i.e., persons with no ‘fixed place of work’ and ‘working from home’ can be identified and thus avert possible confusion when coding industry. The notes on the questionnaire are extensive and do take considerable space.



#### 4. *Suggestions for enhancing effectiveness of questions*

346. Industry questions have been asked in several censuses in many countries and thus there is usually some experience with this topic. However, there are still many problems being identified with the resulting statistics. The following issues have been found to be of concern:

- (a) Lack of space on the questionnaire for recording the relevant information has been found to be the major factor resulting in inadequate industry descriptions. This is particularly so if the landscape format is used for the questionnaire.
- (b) The word industry should not be used in the question, as respondents' understanding of the word frequently is not what is being aimed for. However, the word may be given as a heading for the question, as has been done in some cases, to guide interviewers. One of the objectives in training of field workers should be to eliminate any confusion that interviewers may have about the meaning of this term.
- (c) The temptation to give too many examples should be avoided as it will clutter the questionnaire and may confuse rather than enlighten the interviewer. Examples should be selected on the basis of previous experience, in particular, those whose identification has often caused problems for enumerators. Similarly instructions to enumerators have to be restricted to the most critical ones only.

#### ***Example F.6 (Untested, hypothetical question)***

<p><i>F.6.1 What is the name of your place of work and what are the main Activities/Services/products of this place of work?</i></p> <p>(a) Name</p> <p>-----</p> <p>-----</p> <p>(b) Main activities</p> <p>-----</p> <p>-----</p>
---

347. It is useful to give good examples and it may be necessary to give effective training to enumerators for such a complex question to succeed.

#### 5. *Issues of measurement*

348. The industry question should always be included in any pre-testing programme because of its importance, even if the topic has been asked many times before. Most developing and many transition countries are still searching for a means of improving the

canvassing of this important topic. Aspects that need to be tested include:

- most suitable order for this block of questions;
- means of obtaining good industry descriptions;
- most effective examples and notes to be included on the questionnaire;
- identification of establishments with different industry codes when they belong to the same large enterprise or government department.

349. Good interviewer training is the key to the collection of accurate industry descriptions. Interviewers should be briefed on the need for accurate and relevant information about the activities of the place of work. They should also be thoroughly drilled on the difference between adequate and inadequate descriptions through a series of examples. The examples should be local and come from the queries and problem cases in recent censuses and surveys. Unsuitable descriptions that occur often, such as ‘Manufacturing’, ‘Public service’, ‘Teaching’, ‘Farming’, ‘Selling things’, must be stressed as inadequate. Field practice should also give sufficient emphasis to the descriptions given for the industry and occupation questions. Supervisors should be instructed to send interviewers back to households when inadequate descriptions are given. It is only with such a concerted effort that the quality of industry and occupation descriptions will be adequate, and with a similar effort on coding that the high quality of the resulting statistics can be assured. Any lists of major enterprises (private and government), if used, should also be handed out when training interviewers and be explained, so that the interviewers are also fully aware of why it is necessary to record the location description for the different units.

350. Special attention should be given to explaining difficult to classify categories such as production for own consumption, paid household staff and work at home or with no fixed location.

- (a) Production for home consumption can be significant in some countries: Subsistence agriculture is usually the most common relevant activity but other activities are also possible, particularly in the agricultural off-season. These activities were discussed in more detail earlier in relation to occupational classification of such activities (see paras.259 and 260). The same type of problem arises for industry. It is very important to have a strategy for identifying these activities separately when they are significant. ISIC, Revision 3 did not have industry codes for subsistence agriculture nor corresponding codes for manufacturing for home consumption or own account construction, but a strategy was developed, and is presented in United Nations (2002). As a result, ISIC, Revision 3.1 has a separate code 9600 for “Undifferentiated goods-producing activities for private households for own use” and a similar code (9800) appears in the draft ISIC, Revision 4.

- (b) Paid staff for private households: These workers include mostly domestic servants, gardeners or watchmen working directly for private households. Their situation needs to be identified on the questionnaire. They can then be coded to a separate industry code, e.g., 9700 for draft ISIC, Revision 4. Generally the first part of an industry question should either serve to identify them as working from home, when they live on the premises, or interviewers should be instructed to record the private household as their place of work. This will differentiate them from persons working for security companies, gardening/landscape companies or domestic service supply companies. The procedure must be explained well to interviewers, as this group is usually of a significant number in developing countries.
- (c) Problems with persons working at home or with no fixed location: It is often confusing to ask for details about the place of work for persons working from home or with no fixed location. If an address is recorded for such cases, this may only be that day's location for a hawker, or some description of the home location. There are two alternatives: (i) continue to ask the question for all the employed but tell interviewers to write in home or 'no fixed location' for these cases, or (ii) try to design the questionnaire so that the place of work is not required to be recorded for these cases.

## **B Institutional sector**

351. The institutional sector of employment is broken down into five main groups/sectors (see para. 302 below). The nature of and conceptual differences between these sectors makes it quite difficult for a population census to display the distinctions between them precisely. Only 3 of the 181 {ILO COMMENT: UNSD TO UPDATE PLEASE} countries that reported census results to the United Nations in the 1975-1984 period attempted to collect information on sector of employment. However, several countries have attempted partial collection of these categories, particularly for the 1990 round of censuses, and these replies have provided very valuable statistics.

### **1. The operational definition**

352. "Institutional sector of employment" relates to the legal and social organisation and institutional status of the establishment in which a job is located. Following the definitions provided in the *System of National Accounts (SNA)*<sup>78</sup> it is recommended to distinguish between the following types of institutional sector categories, which are noted in the *Principles and Recommendations, Revision 2*. These are:

- The non-financial corporations sector
- The financial corporations sector

---

<sup>78</sup> See chapter IV of Commission of the European Communities and others (1993). *System of National Accounts 1993*. United Nations Sales No. E.94.XVI.4.

- The general government sector
- The non-profit institutions serving households sector
- The household sector.

353. Detailed descriptions of each of the sectors are presented in the *Principles and Recommendations, Revision 2*, (para. 2.335) and also in the System of National Accounts (1993), (Chapter Four).

## 2. *Requisites for applying the definition*

354. Due to the complexity of the definitions and the difficulty in applying them in a large-scale exercise, the main relevant distinction to be made in a population census is to split paid employees into government paid employees, and other paid employees, and this split will be the primary focus of this section. This split is very important for labour policy analysts as almost all employment estimates and policy plans attempt to provide separate statistics for the government and the other (mostly private) sectors. This is because of the perceived difference in relevant policies for the two sectors.

355. The definition of government comprises:

- All branches of government, national, state/provincial or local;
- Social security funds imposed and controlled by those units; and
- Non-profit institutions (NPIs) engaged in non-market production that are controlled and mostly financed by the units in the above two categories.

356. The employees of all branches of government are generally not difficult to identify in a population census. Employees of social security funds can be identified, but the issue is whether extra steps need to be taken to ensure they are included in the government sector rather than with other financial or non financial sectors, in which case a separate question/category or instruction will be required.

357. Employees of the non-profit institutions are also quite difficult to identify separately in a census. The NPIs include schools, hospitals and trade associations, some of which are government controlled but others are partly or fully privately financed and controlled. When mixed types of NPIs exist in a country it is essential that clear guidance be given to field staff whether they are to be included or excluded from the government sector. This issue should be discussed with national accountants and labour policy analysts. Obviously local terminology rather than national accounts terminology should be used in Manuals or notes to interviewers (e.g., “include all employees of government aided schools, colleges and hospitals as government”).

358. The situation for all semi-government or parastatal (i.e., partly government and partly private)<sup>79</sup> organizations needs to be made clear. In many developing countries

---

<sup>79</sup> The term ‘para-statal’ may also need a locally known definition, as in some countries the term means commercial enterprises in which the government has a controlling interest or ownership.

there are many such organizations, though they are now being partly or fully privatised. For simplicity it would be best to exclude all such organizations from government and provide a separate category for them among the possible responses to the appropriate census question. A wide range of examples of such institutions should be included in the field manuals and covered during training of field staff. The status and ownership of the Post Office, national railways, Harbour Boards, National Airline, etc., vary from country to country and over time in any country. It should be made very clear whether such organizations/institutions are to be classified as government or in semi-government category (as the case might be). Often interviewers and respondents think of some of these as government organisations even when they are partly or fully privatised.

359. *Sub-divisions of government:* Some countries will want to identify separately different categories of government, e.g., national government, state/provincial government and local government. Steps should be taken to ensure that the distinctions between these categories are clear and well understood by field staff and respondents. In some countries there are complex situations, for example, where the central government pays and controls the teaching staff of schools but local governments pay and control other staff and the general management of the schools. Such situations seem to always lead to confusion even if explained well to field staff. If there are too many such complex situations, it is recommended that these sub-divisions of government not be attempted, or the procedures should be pre-tested.

360. The industry codes for financial services will allow further split of those employees in the financial corporations sector. Employment in non-profit institutions (NPIs) is very difficult to measure separately in a population census as the required information on the characteristics of the establishment (their non-profit status and who they service) generally cannot be expected to be known by many of the persons employed in them. Thus it is expected that paid employees of establishments in this sector will mostly be grouped together with those employed in private non-financial corporations, although some may also consider themselves to be government employees.

361. *Sub-divisions of other:* Some countries have sub-divided the 'other paid employees' into categories for missions/churches, Non-Governmental Organisations (NGOs) or parastatals, when such organisations are significant employers in the country concerned. With any such set of sub-categories it is recommended that the relevant questions and response alternatives should be well tested to ensure they are easily and accurately understood. For example, the NGO category has produced some surprising results when tested. It may also be noted that (as indicated in United Nations, 2003)<sup>80</sup>, some non-profit institutions are serving in the market sector.

---

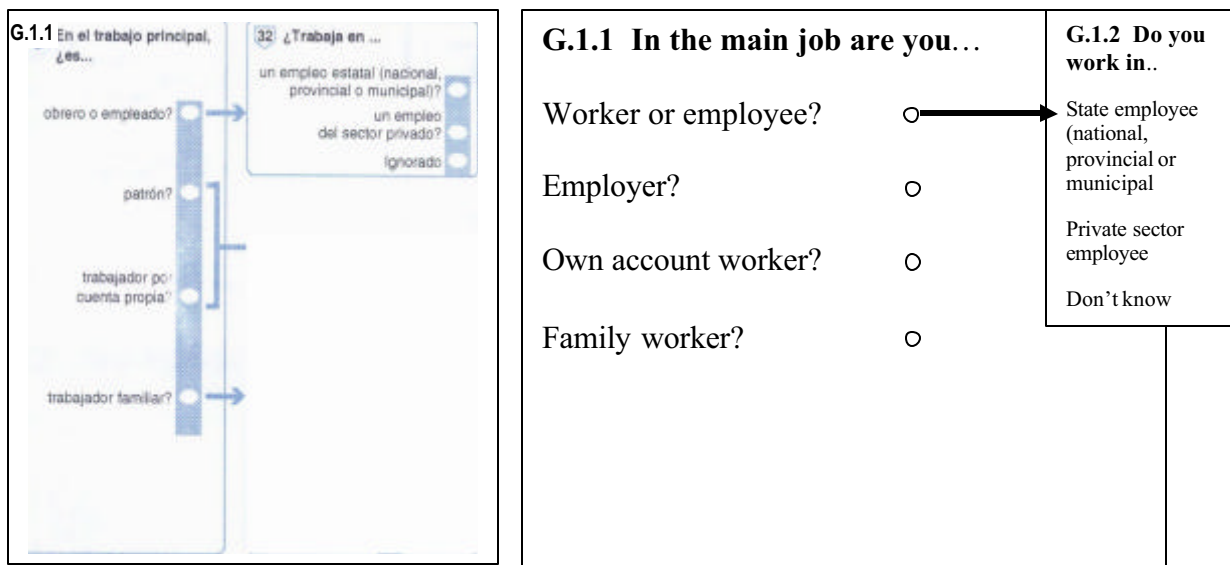
<sup>80</sup> United Nations (2003): Handbook on Non-Profit Institutions in the System of National Accounts, Series: F, No.91, Sales 03.XVII.9

### 3. *Review of questions used in national censuses*

362. Two types of questions have typically been asked:

- (a) those that treat this as a distinct topic investigated on its own (example G.1); and
- (b) those that combine the institutional sector with the status in employment (examples G.2 to G.4).

#### *Example G.1*



363. The question on institutional sector is very simple with two main categories, state/government and private, and is addressed to employees only. For many purposes this will suffice. However, as the above discussions indicate, there are sub-categories under government that users might be interested in.

#### *Example G.2*

364. The question in this example combines employment status and number of employees in the establishment, with institutional sector. This unduly complicates the question and could increase the time that it would take to complete the interview. The quality of responses should also be of concern; employees do not always know how many workers there are in the establishment. It would simplify the task of enumerators and help improve the accuracy of responses if separate questions were presented for the different characteristics, as was done in example G.1.

**Example G.3**

P44. ARE/WERE YOU SELF-EMPLOYED OR WORKING FOR SOMEONE ELSE IN YOUR MAIN JOB?				
1	<input type="checkbox"/>	SELF-EMPLOYED (NO PAID HELPER)		
2	<input type="checkbox"/>	SELF-EMPLOYED (1-4 PAID HELPERS)		
3	<input type="checkbox"/>	SELF-EMPLOYED (5 OR MORE PAID HELPERS)		
4	<input type="checkbox"/>	EMPLOYEE (GOVT/GOVT CORP)		
5	<input type="checkbox"/>	EMPLOYEE (PRIVATE - 1-4 WORKERS)		
6	<input type="checkbox"/>	EMPLOYEE (PRIVATE - 5 OR MORE WORKERS)		
7	<input type="checkbox"/>	UNPAID FAMILY WORKER		
				(Skip to P46)
P45. DO YOU MOVE ALL YOUR GOODS DAILY; E.G. FRUITS, PEANUTS, NEWSPAPERS, CLOTHING, EQUIPMENT?				
1	<input type="checkbox"/>	YES (INFORMAL TRADER)		2 <input type="checkbox"/> NO

G.3.1	Was ... Read list. Mark (X) one box.
1	<input type="checkbox"/> Employee of PRIVATE FOR PROFIT company or business or of an individual for wages, salary, or commission.
2	<input type="checkbox"/> Employee of a PRIVATE NOT FOR PROFIT, tax-exempt, for charitable organization.
3	<input type="checkbox"/> Local or territorial GOVERNMENT employee (territorial/commonwealth, etc.)
4	<input type="checkbox"/> Federal GOVERNMENT employee.
5	<input type="checkbox"/> SELF-EMPLOYED in own NOT INCORPORATED business, professional practice or firm.
6	<input type="checkbox"/> SELF-EMPLOYED in own INCORPORATED business, professional practice, or firm
7	<input type="checkbox"/> Working WITHOUT PAY in family business or firm.

365. The question provides more of the institutional sector categories than usual. It assumes that respondents (who may be reporting on behalf of other household members) do know if their workplace is a 'private, for profit' company or a 'private, not for profit' organization. It is strongly recommended that the understanding of any such categories be thoroughly tested before being introduced. 'Government employees' as a group is separately identified, but 'self employed' is not categorised by whether with or without 'employees'. Instead the 'self employed' are categorised by whether their business is incorporated or not. Therefore other categories in the employment status classification, such as employer and own account worker, cannot be derived from these response categories.

#### ***Example G.4***

G.4.1	Were you self-employed or working for someone else in your ( <u>main</u> ) job last week?
Self-employed	Worked for someone else
<input type="checkbox"/> With paid help (Employer)	<input type="checkbox"/> As Government employee
<input type="checkbox"/> Without paid help	<input type="checkbox"/> As employee of a Foreign Government
	<input type="checkbox"/> As employee of private company/person
	<input type="checkbox"/> As unpaid worker in family business/farm
	<input type="checkbox"/> Not stated

366. The question is a reasonably standard ‘status in employment’ question with ‘paid employees’ split into two government categories and private for an institutional sector classification, in line with the earlier discussion. It however uses the international classification of status in employment (ICSE) terminology and examples or local wording might have been better, as mentioned earlier. ‘Unpaid worker in family business or farm’ is listed as part of ‘working for someone else’, though the corresponding ICSE-93 category is part of self employed. It may have been shown in testing that respondents recognised the category better under the ‘worked for someone else’ heading. They can easily be regrouped for tabulation and publication of the results.

#### ***4. Suggestions for enhancing effectiveness of questions***

367. The institutional sector categories used in combination with employment status seems to work well for a census. The major categories suggested for this topic, i.e., those of government and other, are sub-categories of paid employees from the status in employment classification, and when formulating the relevant questions the categories of these two variables can be combined with little difficulty as shown in the set of example G.2 to G.4 above. Although this approach should not generally cause problems, it should be tested if introduced for the first time.

368. There is a tendency, when institutional sector is combined with other characteristics, such as employment status, to use technical terms, as this helps to shorten the sentences or response categories. This temptation should be resisted and an effort should be made to use locally understood terminology to improve the quality and consistency of responses. Additionally, when two characteristics are combined in a question, some categories of each characteristic are lost. These questions need to be carefully tested for their validity and for their potential uses in future analyses.

#### ***5. Issues of measurement***

369. If the suggestion above to combine questions for institutional sector and status in employment is adopted the main points noted for testing on institutional sector are:



- (a) Respondents' interpretation of government, particularly regarding 'semi-government' or 'parastatal' organisations. Identify necessary prompts and notes on the questionnaire, or instructions for Manuals and training, to try to record the meaning of these terms as desired in that particular country.
- (b) Test any sub-categories of 'government' or 'other' (particularly 'profit/non-profit' as in example G.3) that are to be identified separately.

### **C Informal sector and informal employment**

370. At the 17<sup>th</sup> ICLS in 2003, a new concept of informal employment was added to the existing concept of employment in the informal sector<sup>81</sup>. Both of these concepts are described in this section. The difference between the two concepts may be summarized as follows:

- (a) The informal sector is enterprise-based, and therefore informal sector employment relates to all people working in enterprises of a certain size, household ownership and other characteristics.
- (b) Informal employment is job-based and relates to all those who have a job with certain characteristics (generally these are workers with no written employment contracts, and those in employment that is not subject to labour legislation, pension/social security regulations, collective agreements, etc.

371. The size and structure of the informal sector is a topic of considerable interest and importance in many countries. The sector is considered to engage 50 per cent or more of the persons in non-agricultural employment, either as a main or as a secondary activity. A number of countries are obtaining reasonable national estimates for the size and structure of the informal sector, mostly using a household survey methodology. As such statistics are in high demand also for different regions and special population groups, data analysts are very likely to attempt to develop measurements of the informal sector from census results, even if those responsible for the census do not plan for such estimates to be prepared.

372. The 15th International Conference of Labour Statisticians (ICLS) gave the first international guidelines on the development of statistics on the informal sector,<sup>82</sup> including a detailed review of definitions and data collection procedures. Precise identification of those working in the sector cannot be achieved with a census, in part because of the limitation on the number of questions that can be included in a census and also because of heavy reliance on third party reporting for the census data collection. It should be possible, however, to obtain some selected indicators (the size, distribution of activities and of occupations within the sector) on at least some segments of the informal

---

<sup>81</sup> See ILO: 17<sup>th</sup> ICLS 2003 Guidelines on measuring informal employment (<http://www.ilo.org/public/english/bureau/stat/download/guidelines/defempl.pdf>) and Hussmanns: Measuring the informal economy: From employment in the informal sector to informal employment, ILO Bureau of Statistics Working Paper No 53 December 2004.

<sup>82</sup> International Labour Office (1993a).

sector and this is important because of the significance of informal sector activities in many local communities.

373. As summarised in the *Principles and Recommendations, Revision 2*, the informal sector is part of the household sector,<sup>83</sup> one of the main categories of the institutional sector of employment variable. Informal sector units are neither incorporated businesses, which have a legal entity separate from their owners or shareholders, nor are they quasi-corporations, i.e., production units which though neither incorporated nor registered do keep a reasonably complete set of accounts so that the finances of the business are kept separate from those of the household(s) which owns or controls the business.

374. The informal sector was a recommended topic for the 2000 round of population censuses, but now an additional topic, namely informal employment, may also be considered for inclusion in the census.

375. As stated above, informal employment is a new concept in the international standards. As shown in Diagram 5, components of informal employment include:

- (a) Own-account workers and employers employed in their own informal sector enterprises.
- (b) Members of informal producers' cooperatives, i.e. cooperatives not formally established as legal entities, etc.
- (c) Own-account workers engaged in production of goods exclusively for own final use by their household (e.g. subsistence farming, do-it-yourself construction of own dwellings).
- (d) Contributing family workers working in formal or informal sector enterprises

376. Employees holding informal jobs, whether employed by formal sector enterprises, informal sector enterprises, or as paid domestic workers by households. Employees are considered to have informal jobs if their employment relationship is, in law or in practice, not subject to national labour legislation, income taxation, social protection or entitlement to certain employment benefits (advance notice of dismissal, severance pay, paid annual or sick leave, etc.).

#### **DIAGRAM 5: CONCEPTUAL FRAMEWORK: INFORMAL EMPLOYMENT**

---

<sup>83</sup> See also Commission of the European Communities and others (1993); paras. 4.132-4.150, 4.159-160, and Annex to Chapter 4).

## Conceptual Framework: Informal Employment

Production units by type	Jobs by status in employment									
	Own-account workers		Employers		Contributing family workers	Employees		Members of producers' cooperatives		
	Informal	Formal	Informal	Formal	Informal	Informal	Formal	Informal	Formal	
Formal sector enterprises					1	2				
Informal sector enterprises <sup>(a)</sup>	3		4		5	6	7	8		
Households <sup>(b)</sup>	9					10				

- (a) As defined by the Fifteenth International Conference of Labour Statisticians (excluding households employing paid domestic workers).
- (b) Households producing goods exclusively for their own final use and households employing paid domestic workers.

Source: Hussmanns: Measuring the informal economy: From employment in the informal sector to informal employment, ILO Bureau of Statistics Working Paper No 53 December 2004.

377. Therefore in order to measure informal employment in the population census it is necessary to collect information, not only on the status in employment (whether paid employee, employer, own account worker without employees, contributing family member, etc), but also on the conditions of employment for employees (security of tenure, access to pension/social security, paid leave, etc).

378. Both informal sector employment and informal employment are needed because both types of employment exist and are of growing importance in the economy. The concept of informal employment is particularly important to consider the quality of employment conditions (which is an aspect of decent work).

379. There is some informal employment outside the informal sector, including employees holding informal jobs in formal sector enterprises, paid domestic workers employed by households, contributing family workers working in formal sector enterprises, and own-account workers engaged in production of goods exclusively for own final use by their household.

380. For the 2010 round of censuses any attempt at measuring informal employment and, to a lesser extent informal sector employment, may to a certain extent be experimental. It will therefore be useful to compare among countries experiences with these topics in order to improve measurement methodology over time.

### 1. *The operational definition*

381. The population employed in the informal sector comprises all persons, irrespective of their status in employment, who during a given reference period, were employed in at least one business unit or establishment, whose business activities are carried out by households, and for which no complete set of accounts is kept, i.e., their business transactions cannot be totally and clearly separated from other household transactions.<sup>84</sup>

382. The population who are informally employed comprise those with the status in employment indicated in para. 373 above and for those who are paid employees, it is necessary to further identify those with reasonably secure employment contracts (that is, of long duration or even permanent contracts), whether they have social security protection and whether they are entitled to certain employment benefits (advance notice of dismissal, severance pay, paid annual or sick leave, etc.).

## **2. *Requisites for applying the definition***

383. The informal sector should not be confused with the concept of the 'hidden' or 'underground' economy or with illegal activities. In most countries, most if not all informal sector activities are legal and 'above ground', at least in the sense that the only laws and regulations broken will be those relating to operational licenses or the payment of taxes and social security contributions.

384. The 1993 International Conference of Labour Statisticians (ICLS) resolution recognised a sub-division of the informal sector into:

- (a) Informal own-account enterprises, which can employ contributing family workers and the occasional paid employees but do not have paid employees on a continuous basis;
- (b) Enterprises of informal employers, who can employ paid employees on a continuous basis, but usually only on a small scale, e.g., with less than five employees.

Thus, there are both own account workers, contributing family workers and paid employees in the informal sector. It is how to identify, separately, individuals working in such activities, which is the problem when measuring employment in the informal sector.

385. The informal sector activities are for many people secondary activities in the sense that they may also have another job, and this may often be considered to be the main one, even if it does not necessarily provide the highest income. Therefore, in order to obtain a more complete measure of its size, both the main and secondary jobs of the individual should be covered. Where possible, the population employed in the informal sector should be sub-classified into two categories: (i) persons exclusively employed in the informal sector, and (ii) persons employed both in and outside the informal sector.

---

<sup>84</sup> *Principles and Recommendations, Revision 2*, para. 2.339.

### 3. *Review of questions used in national censuses*

386. There is little experience with collecting data on the informal sector in censuses and no experience in collecting informal employment.<sup>85</sup> One might expect that few countries will wish to burden their census questionnaire with the additional questions on the characteristics of the job contract and conditions of work that are needed to measure informal jobs of employees. However information from questions on some topics such as status in employment and institutional sector included in the census, are directly relevant. For the measurement of informal employment, additional questions would be needed on the employment contract (whether it exists and if so, the duration) and on access to certain employee benefits. The discussions on possible questions deal primarily with the person's main job (economic activity). The possibility of a question on secondary activities and a question at the household level respectively, both with the aim of obtaining a better measure of total informal sector employment and total informal employment, are discussed under the suggestions for enhancing the effectiveness of the questions.

387. *Questions relating to the person's main employment:* The informal sector is a characteristic of the private enterprise in which the person works. It is important to design and locate any proposed question so that it relates to the correct sub-group of the total population. It would be most efficient for the informal sector questions to follow an institutional sector question and only those working for private enterprises were directed to it.

---

<sup>85</sup> The first time that the topic of informal sector was included in the *Principles and Recommendations*, was in the Revision issued in 1998. Thus the United Nations has not had this item in its previous assessments of the content of census questionnaires.

### Example H.1

ANY WORK IN THE 7 DAYS BEFORE 10 OCTOBER  
H.1.1  
In the SEVEN DAYS before 10 October did (the person) do any work for PAY (in cash or in kind) PROFIT or FAMILY GAIN, for one hour or more?

1 = Yes: formal registered (non-farming)  
2 = Yes: informal unregistered (non-farming)  
3 = Yes: farming  
4 = Yes: has work but was temporarily absent  
5 = No: did not have work

If YES go to X.X

388. Combining a ‘yes’ response with several characteristics of the job complicates the basic question used to determine economic activity status. In defining the categories, both industry and institutional sector variables are combined. As earlier stated, terms like “formal” and “informal”, which are not easily or uniformly understood by field workers (or even by statisticians in general) should not be used in a question without providing explanations. Moreover the formulation “informal unregistered” may be incorrectly understood by some, who will suppose that there is an “informal registered” category somewhere. It is unlikely that this approach will yield reasonably accurate estimates of a sector whose measurement requires more elaborate combination of institutional sector, industry and legal requirements.

389. In order to identify informal employment, one must first determine the status in employment and then, for paid employees, ask about characteristics of the employment contract and employment conditions.

### Example H.2

H.2.1 Was ... *Read list. Mark (X) one box.*

1  Employee of PRIVATE FOR PROFIT company or business or of an individual for wages, salary, or commission.

2  Employee of a PRIVATE NOT FOR PROFIT, tax-exempt, for charitable organization.

3  Local or territorial GOVERNMENT employee (territorial/commonwealth, etc.)

4  Federal GOVERNMENT employee.

5  SELF -EMPLOYED in own NOT INCORPORATED business, professional practice or firm.

6  SELF -EMPLOYED in own INCORPORATED business. Professional practice, or farm

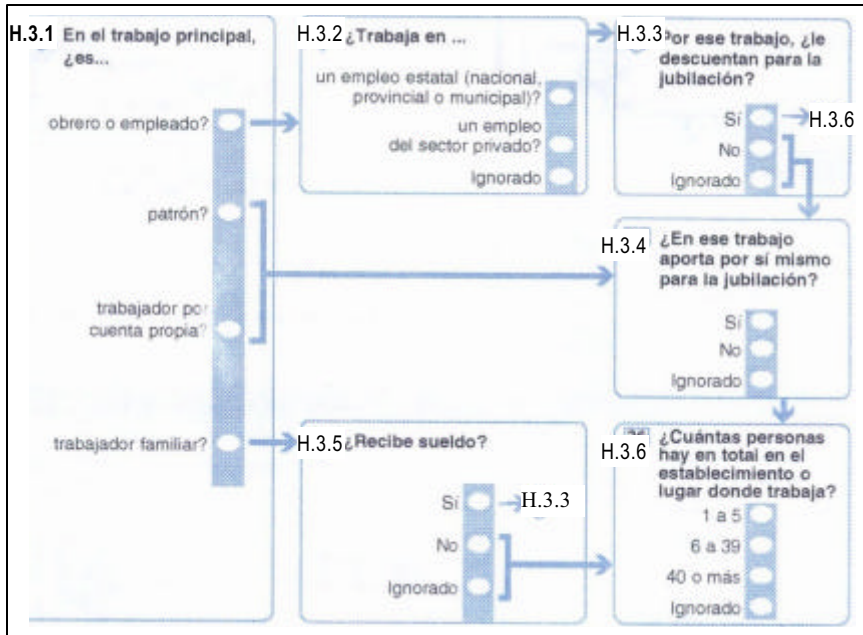
7  Working WITHOUT PAY in family business or firm.

390. The question in example H.2 combines a partial informal sector identification within the sector classification by sub-dividing the self-employed by whether they work in a NOT INCORPORATED business or an INCORPORATED one. This would assist greatly in identifying self employed in the informal sector but would not cover all the informal sector establishments, since unincorporated businesses keeping complete sets of accounts will not be identified; nor will any paid employees working in the informal sector be identified from this question.

391. Questions would need to be asked to determine whether an employment contract existed and if so its duration and whether the paid employee was covered by social security provisions, had paid annual leave, paid sick leave, etc.

392. Some have asked for the numbers of employees or paid employees working at a person's place of work. The results can be used for approximate informal sector identification.

**Example H.3**



<b>H.3.1 In the main job are you...</b> <input type="checkbox"/> Worker or employee?	<b>H.3.2. Do you work in...</b> <input type="checkbox"/> State employee (national, provincial or municipal) <input type="checkbox"/> Private sector employee <input type="checkbox"/> Don't know	<b>H.3.3. For this work, was allowance made for your pension?</b> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know
<input type="checkbox"/> Employer? <input type="checkbox"/> Own account worker?		<b>H.3.4. For this work, did you make any contribution towards your own pension?</b> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know
<input type="checkbox"/> Family worker?	<b>H.3.5. Did you receive any pay/wage?</b> <input type="checkbox"/> Yes <input type="checkbox"/> No <input type="checkbox"/> Don't know	<b>H.3.6. How many people are there in total in the establishment where you work?</b> <input type="checkbox"/> 1 to 5 <input type="checkbox"/> 6 to 39 <input type="checkbox"/> 40 or more <input type="checkbox"/> Don't know

393. In the flow of questions, all of the employed are asked to state the number of workers in the establishment where they work. Those working in the informal sector can then be approximated from a combination of responses to questions H.3.1, H.3.2 and H.3.6 (employers and employees of the private sector, own account workers, and family workers, in enterprises that have five or fewer workers). It is not possible with these set of questions to distinguish between regular employees and other types of employees or unpaid family workers; and the only other criterion that can be used to define the informal sector, besides the number of workers, is “whether deductions are taken for pension”.

#### **4. *Suggestions for enhancing effectiveness of questions***

394. Information about a number of basic characteristics of the establishment is required depending on the definition adopted and the degree of precision aimed for. For example, in addition to being privately owned, the unit (i) has to be unincorporated, (ii) not keep separate accounts from household expenditures, and (iii) be small in size, e.g., with respect to employment.

395. Any information that has been collected from population censuses can at best give only partial coverage of the informal sector employment. Some countries have tried to identify segments of the informal sector from characteristics other than those discussed above. One example is using the location of the place of work, specifically by identifying those persons who have no fixed work location or those who are working at home. Another approach is to ask a specific question to identify one segment of the informal sector.

#### ***Example H.4***

<p><b>H.4.1</b> Do you/does he/she move all your/his/her goods every night: e.g., fruits, nuts, lottery tickets, clothing/shoes, etc.?</p> <p>1 <input type="checkbox"/> Yes (Informal trader)      2 <input type="checkbox"/> No</p>
---

396. Example H.4 provides a very restricted view of the informal sector, and may not have been the intent of the question. Non-retail activities (including handicrafts, bicycle repairs from home, taxi operation, etc.) would be excluded by this question. The question may be useful for broader planning in some countries, with particular focus on a sub-sector of the informal sector.

397. Almost all informal sector activities can be done either as a main or as a secondary activity for the individuals involved. In some countries the number engaged in informal sector as a secondary activity is almost as large as those engaged in them as a main activity. With the difficult economic circumstances in many countries, it is only by engaging in secondary activities that many who work as paid employees as the main



activity can survive and support their families. In many cases one member of a family will engage in the informal sector activity as a main activity and other household members will assist as secondary activities. It should be noted, however, that very few countries have asked census questions about secondary economic activities. One country, Cook Islands, asked the complete block of economic questions again with regard to secondary activities, but for most countries there will probably be insufficient space for this approach to be used. If the complete set of questions were asked for any secondary activities, the same procedures discussed earlier, could be used for identifying persons engaged in the informal sector as a secondary activity.

398. Given the limit space and already curtailed number of questions even for better tested questions on economic characteristics, the topic of the informal sector is best left for sample surveys designed specifically to collect this type of data. It is definitely not possible to ask a simple question such as “Do you work in the informal sector?” and expect to obtain reasonable results. Thus, a probable approach to: (i) examine what can be obtained by using the other topics/variables and their classifications which are included in the census, and (ii) look into possible additional questions and response categories which may be used to give more refined or precise identification of those working in the informal sector. It is strongly recommended that countries use this approach, i.e., examine what can be obtained from their currently proposed questions and classifications and then examine possible further questions to provide refined measures of the informal sector. This should be done whether or not it is decided to make a serious, official attempt at informal sector measures from the census, as users of the census results are very likely to ask for estimates of informal sector employment from the census, nation-wide as well as for local labour markets.

399. An alternative approach would be to use a simple question such as the untested hypothetical example given below. A major aim would be to identify secondary informal sector activities, but the responses could have broader use than that. The question would come after all questions relating to the main activity, in order not to confuse the respondents.

***Example H.5 (Untested, hypothetical question)***

**PERSONS EMPLOYED IN PRIVATE BUSINESSES ONLY**

**H.5.1** *Is the business you work in:*

- A Registered Company (e.g. Pty or Ltd)?
- Keeping a complete set of accounts?
- Registered for Value Added Tax (VAT) collection?
- Having 5 or more paid employees on a regular basis?

**Interviewer:** For each of the above record one of the following codes

Yes .....1  
No .....2  
Don't know .....3

400. If employed in the informal sector, the individual would answer NO to all four questions above. For most informal sector small business operators, whether employers or own-account workers, the questions would not be difficult. For contributing family workers interviewers would have to be trained to ensure that the answers should be the same as for the person(s), usually in the same household, who is/are the self employed operator(s) of the same business. Paid employee respondents are much less likely to have the required knowledge about characteristics of the business activity, such as whether the business is registered or keeps a full set of accounts; and they may not correctly differentiate between regular paid employees and unpaid family workers, although they may know approximately how many workers are engaged on a continuous basis. Because the reply to questions other than that concerning the number of employees, will most often be “Don’t know”, it should be established through testing whether it is worthwhile to ask the complete set of questions of the employees. Problems are likely to be even more severe in third party reporting, that is, if another person is used to answer questions on behalf of the actual respondent.

401. The sequence of questions indicated above could become fairly complicated. Therefore it may be preferable to ask just one simple additional question on the topic of informal sector. A question about the number of paid employees would be the easiest and could be asked for all employed persons. The results from this question are also of general interest and the elimination of certain sectors and industries that are known not to be represented in the informal sector could be done during data processing or analysis. If this one question is to be applied to all the employed it may be best to place it after the question on place of work.

***Example H.6 (Untested, hypothetical question)***

<p><b>H.6.1:</b> <i>Besides your main economic activity covered in the previous questions, did you engage in any of the following forms of secondary economic activity in the last seven days?</i></p> <ul style="list-style-type: none"><li>• Working on any form of small business for yourself or with a partner</li><li>• Working unpaid in a family business?</li><li>• Working on your own or family farm or ranch?</li><li>• Working in a paid second job?</li></ul> <p><b>Interviewer:</b> For each of the above record one of the following codes</p> <p>Yes .....1</p> <p>No .....2</p> <p>Don’t know .....3</p>
--

402. The questions have been kept simple to obtain a rough broad estimate of secondary activities in these categories. ‘Yes’ answers to the first two categories could then be considered to identify informal sector self-employment jobs. Probably most of those answering ‘yes’ to the last categories of questions would be in an informal sector

paid employee jobs. If the first two categories were asked alone, a note would need to be inserted to exclude agricultural activities from them. The informal sector questions could be more detailed and specify, for example, that the business should have less than five continuing paid employees. However it has been shown in a number of surveys that secondary activities almost always are with units that have few employees. This, however, could be confirmed by testing. The responses to such questions would only provide a crude measure of secondary activities, principally informal sector activities, but the results may be better than having no estimates, in particular for small areas and sub-population groups, and could be used with higher geographic level survey estimates to impute other characteristics at local levels.

403. *Identification of informal sector activities at the household level:* In many cases informal sector activity is a household activity with several members of the household engaged, some full-time and some part-time. Identification of the number of households with some engagement in informal sector activity, preferably by type of household and activity, is extremely valuable information in itself and also for creating a sample frame for later informal sector surveys.

404. The household information can be obtained during listing preparations for the census, provided 'solutions' can be found to the following problems, if they arise:

- (a) It is often difficult to match the individual household information from the listing with the replies obtained for the household during the census enumeration.
- (b) Informal sector activities are often intermittent. People stop and start easily, particularly the smaller informal sector activities and especially those that are seasonal. Listing information collected over a year or more prior to the census (as is often the case) may not relate to economic activities at the time of enumeration.

405. Thus although listing data can be useful for both the informal sector identification and for other enterprise characteristics (see *Principles and Recommendations, Revision 2*, paras.1.51-1.52), it is far more useful if a question were asked during the actual enumeration. The following question is hypothetical and untested, but is provided only for illustrative purposes.

***Example H.7 (Untested, hypothetical question)***

**H.7.1: Did any member of the household engage in any of the following small scale business activities in the last seven days**

- Making or repairing any items (e.g. clothing, pots, hats, furniture, metal products)?
- Preparing food or beer for sale?
- Selling food, clothing etc. at the market, in a small shop etc?
- Transporting people or goods, using e.g. taxis, hand carts?

**Interviewer: For each of the above record one of the following codes**

Yes .....1

No .....2

**NOTE: Must be own or family business with less than five employees on a regular basis. Not working as a paid employee.**

406. The question targets particular types of informal sector activities. This is an effective approach and it can be extended to other categories. It would obviously have to be changed for each country to fit local examples and priorities. It could also be asked generally about small-scale business activities with a variety of examples.

407. Example H.8 is a set of questions that were designed to measure informal employment in one country's Labour Force Survey. As previously mentioned, it is suggested that such a detailed set of questions would generally not be appropriate for inclusion in a population census, but they may be useful if a programme of labour force surveys does not exist.

### ***Example H.8***

H.8.1	Is [Name] employed on the basis of a contract or agreement?	
	Yes 1 (continue)	
	No (go to H.8.4)	
H.8.2	Is the contract or agreement of a limited duration?	
	Yes 1	
	No 2 (go to next section)	
H.8.3	What is the duration of the contract or agreement?	
	Daily	1
	Less than a month	2
	1 to 2 months	3
	3 to 6 months	4
	7 to 12 months	5
	Over 12 months	6
H.8.4	Does [Name]'s employer pay social security contributions for [Name] (pension fund, unemployment fund)?	
	Yes, certainly	1
	Yes, possibly	2
	No	3
	Don't know	4
H.8.5	Does [Name] benefit from paid annual leave or get compensation for unused leave?	
	Yes	1
	No	2
	Don't know	3
H.8.6	Does [Name] benefited from paid sick leave in case of illness?	
	Yes	1
	No	2
	Don't know	3
H.8.7	(For women only) Does [Name] benefit from maternity leave if she wishes to have a baby?	
	Yes	1
	No	2
	Don't know	3

### **5. *Issues of measurement***

408. As a new topic in most countries, and in particular in a census, any questions related to informal employment or the informal sector should be subject to thorough testing. It is strongly suggested that a test of any such questions be followed by complete re-interviews using a detailed household survey questionnaire to check both whether the identified informal sector activities really are informal sector activities and whether any relevant informal sector activities were missed, and whether the employment was really informal in nature. Tests would need to cover a variety of rural and urban areas and various types of informal sector activities. They should be combined with test derivations of informal categories during processing and analysis as discussed below.

409. As indicated earlier, at present the only reasonable option in many censuses, is to derive estimates of employment in the informal sector estimates from answers to other questions as no specific question designed to identify the informal sector is known to be have been asked. The responses to the status in employment question, for example provide the most useful information for such derivations. The self employed with no employees, i.e., own-account workers, and contributing family worker categories are the most useful ones for identifying possible informal sector workers. If those engaged in

agricultural work can be excluded by using the appropriate industry codes or other means, then most of the remaining persons in employment in these two status categories will be working in the informal sector in many developing countries. Careful consideration of other industry and occupation categories, combined with local knowledge, may lead to the exclusion of other jobs and a more refined measure. However, similar possibilities are difficult to find for the self-employed with employees and for paid employees. Thus only for two components of the informal sector workforce can rough measures be obtained by this approach. However, for some countries this may be the only approach they can afford, at it may be better than making no efforts at all. The estimates can be further refined if the number of employees (i.e., less than or more than the size limit defined for the informal sector) were included in the employment status categories.

410. Similarly estimates of some of the components of informal employment as show in Diagram xx can be made from the answers to the question of status in employment, but the identification of those employees with informal jobs will not be possible without more detailed questions, and it will still be necessary to identify those employed in informal sector enterprises.

411. In most of the countries with a large informal sector, analysts (official or otherwise) have frequently tried to obtain statistics on the informal sector from previous censuses by making use of statistics classified according to sector, occupation, industry and status in employment, if available, and on this basis derive those groups which they think will approximate the informal sector or some of its activities: e.g., taking the self employed coded to the industry code for 'street sellers and hawkers' to approximate this segment of the informal sector.

412. It is easier initially to identify what is not in the informal sector, and to make sure that the persons engaged in these activities (i.e., formal sector and subsistence activities) can be identified separately and thus excluded from consideration either for additional questions and/or during the tabulation stage. There are two groups that can clearly be excluded from the informal sector or be separately identified, using information obtained with the status in employment, sector of employment, industry and occupation questions. These are the following.

- (a) Employees of recognisable formal institutions: All paid employees working for the government, or semi-government or parastatal enterprises (or NGOs or international organisations where significant). These employees can be identified principally from the appropriate sector and industry categories.
- (b) Persons engaged in agricultural or non-market production activities (as own account workers or paid employees) or as private household employees.
  - (i) Agricultural activities: All agricultural activities, including those for home consumption or for sale, and as an unincorporated enterprise, which are part of the household sector, are usually excluded from the informal sector. Population censuses can and usually do provide

valuable measures for the identification of persons engaged in these activities. Industry classifications based on ISIC, Revision 4, might be made to allow separate identification of subsistence or traditional agriculture from commercial agriculture, and occupational classifications based on ISCO-88 usually allow for a distinction between market oriented agricultural workers, and subsistence agricultural workers where this distinction is important. It is recommended that if possible persons classified to these categories should be identified and be excluded from the remaining discussion of the informal sector.

- (ii.) Non-market production activities: These activities include the processing of primary products for home consumption, e.g., making butter from milk, oil from oil seeds, making mats from straw or reeds, weaving cloth from fibres; and building or carrying out major repairs on the family farm or house, e.g., fence or irrigation ditch making; as well as fetching/collecting water and firewood. As earlier noted, persons engaged in these activities are to be included as employed if such activities are important in the national economy, and they are included in the household sector. They are not commonly included in the informal sector, however, because they are non-market as well as closely tied to agricultural activities. Note that the processing and building activities are often done partly for sale (which is informal sector) and partly for home consumption (which is not informal sector), but that they can become the only or main economic activity during certain periods, e.g., in the agricultural off-season. Note also that these activities usually cannot be separately identified by the use of standard industry and occupation codes, as persons with jobs consisting of only such processing or building activities for own use should be coded to the appropriate industry group (e.g. manufacturing, construction) and occupation (e.g., process worker, construction labourer), and will be difficult to differentiate from similar workers who are producing for a market and thus are to be considered for inclusion in the informal sector.<sup>86</sup> These problems should be carefully examined, particularly in countries where such non-market activities are significant. Interviewers will need guidance on how to record these situations, to ensure that special codes for them can be introduced for their classification. Note again that the status in employment question suggested earlier would include most of these activities among those working on own or family farm etc., and that this will assist in identifying them separately from the small scale non-agricultural businesses which are in the informal sector. This is another reason for considering such a status question.

---

<sup>86</sup> ISIC, Revision 3.1 was reviewed in this context and a new industry code 9600 was introduced to take account of own-account household production. The draft ISIC, Revision 4 has a similar code 9800.

- (iii.) Employees of private households: These paid employees, such as maids, cleaners, guards, etc., are included in the household sector and are in some countries considered to be part of the informal sector. The resolution recommends that they should then be shown separately. A special industry code is usually used for such employees, e.g. code 9500 in ISIC Revision 3 (code 9700 in ISIC, Revision 4). The occupational classification should ideally have separate codes for each of these occupations, so that this component of the household sector employment can easily be identified through cross-classification with the industry variable. Then the users of the statistics have the option of keeping them within or outside the informal sector according to descriptive or analytical needs.

413. The informal sector employment is derived as follows:

$$\text{Workers total} - \text{RIC}_p - \text{OBCA}_p - \text{COM}_{p=x} = \text{Workers in 'private businesses'}$$

Where:  $\text{RIC}_p$  = Persons working in registered incorporated companies

$\text{OBCA}_p$  = Persons working in other enterprises (businesses) with complete accounts

$\text{COM}_{p=x}$  = Persons working in companies with “x” or more paid employees on a continuous basis.

414. The specified exclusions cannot usually be determined from the answers to questions on other topics included in the census. Relevant questions would thus have to be asked, and the types of questions possible and whether respondents can be expected to have this knowledge were considered earlier in this chapter.

415. The upper limit on the number of employees in the own account enterprise to be regarded as an enterprise of the informal sector is usually decided in each country according to the local circumstances. If there is a business register that supports employer/establishment based statistical surveys, this often includes only establishments above a certain size, often expressed in terms of the number of paid employees. If this is the case then the informal sector should be complimentary to the scope of this register. In countries where there is a Value Added Tax (VAT) and a lower limit of total sales for VAT registration, then such registration may be a possible criterion for defining the scope of the informal sector, at least if there are no other reasonable formalization criteria available. (Note that with VAT non-registration as the criterion there will be informal sector units with total sales larger than the threshold for VAT registration, either because of deliberate tax evasion or because the registration requirement has not yet been complied with.) It should be noted that surveys have shown that the number of informal sector enterprises declines rapidly as the number of paid employees on a continuous basis increases, because most enterprises with high levels of employment are usually not informal for one of the reasons given earlier.



## Chapter 8. Person Totals

416. It is essential that once the main job is identified all characteristics refer to this same main job throughout the following questions. The questionnaire design must be such as not to confuse respondents or interviewers on this point. When any secondary activities are identified they must be clearly separated and placed so as not to confuse the flow of questions relating to the main job. This is usually not difficult with current activity (see later examples) but can be quite difficult with usual activity.

### A Working Time

417. 'Working time' is a very valuable topic for many countries, both developed and developing. Working time covers both working hours and working time arrangements. A growing proportion of employees and self-employed persons have non-standard working hours/arrangements. In these situations observing only the number of persons employed will roughly estimate the total amount of work done. One needs information about the hours actually worked by those employed to obtain estimates of labour input with any reasonable degree of precision and to provide comparability over time and between groups. Information about working time is also essential for studies of the relationship with income, underemployment and productivity, as well as for some national accounts measures, and for assessing new working time arrangements and the extent to which excessive hours are worked. Working time was asked in 84 of the 181 countries that reported census results to the United Nations during the 1975-1984 period. {ILO COMMENT: UNSD TO UPDATE PLEASE} It is recommended that working time on all economic activities, main as well as secondary, be recorded for all employed persons and not just for paid employees.<sup>87</sup>

418. Many countries use household surveys to obtain national estimates of working time, and it is true that conceptually better measures and a larger variety of measures can be obtained by the more thorough yet more time consuming methodology which such surveys can use, in particular to measure the key concept of hours actually worked. However, it is primarily through a census that measures for small areas and for small groups of the employed population may be obtained. Estimates of working time from surveys of establishments or enterprises usually only cover large, formal employers, and cannot give any breakdowns by age, sex or occupation.

#### 1. *The operational definition*

419. Working time is the total time actually spent producing goods and services or

---

<sup>87</sup> In 2008 at the 17<sup>th</sup> ICLS, the discussion of revised international standards concerning the measurement of working time will address extension of the production boundary as an alternative to also measuring working time for all production of services for own final consumption within households, such as cleaning, cooking, repairs, transport and care of dependent persons, which is a significant proportion of often unmeasured productive activity.

performing activities related to or enhancing production, including time spent on short pauses that may be in the nature of work-related breaks or rest periods during the reference period adopted for economic activity in the census. It is recommended that if the reference period is short, e.g., the week preceding the census, working time should be measured in hours, as hours actually worked. If the reference period is long, e.g., the 12 months preceding the census, working time should be measured in units of weeks, or in days where feasible. It is however not recommended that any attempt be made to establish the usual hours for such a long period, as recall and calculation errors are likely to be high.

## **2. *Requisites for applying the definition***

420. Working time should include the total time spent producing goods or services during the reference period, including any overtime, time spent in preparation for work (which includes training to facilitate production for the economic unit) or in maintenance or repair of equipment, as well as short tea breaks. Teachers should include their preparation and similar work related time in addition to the time actually in front of a class, and so should persons on-call in essential (health, security, etc.) services whose movements/behaviour are restricted even if not at work. Hawkers and other sales workers should include time waiting or looking for customers. However, long meal breaks should be excluded as should time off for holidays, vacation, sickness, or industrial disputes. Time spent commuting (to the work place and back) should be excluded – unless work is carried out, such as during a train ride, and travelling is part of the job, e.g., for taxi drivers, and workers travelling to or between variable work sites, such as field projects, fishing areas or meetings. As mentioned earlier, the total working time should include all activities included as economic activities as detailed earlier, and if practical, time spent on secondary activities should be recorded separately from the time spent on the main activity. This will allow for more accurate recording of time, and also provide a sound basis for analyses comparing wages and salaries for different categories of workers.

421. If the block of economic questions asks about current activity then the question(s) about working time should also relate to the current (short) reference period. The objective thus is to measure hours actually worked in the current reference period. Such measures will include some zero hours for persons temporarily absent from work (on holidays, sick leave, etc.) throughout the reference period. Reduced hours as compared to ‘usual’ hours will also be recorded if any part of the reference period is taken for holidays or off sick, etc. All overtime, paid or unpaid, should be included. Thus this concept measures hours actually worked rather than hours paid for or normal hours established by laws or regulations.

422. Although the *Principles and Recommendations, Revision 2* (para 2.323-2.325), recommends the measure of actual hours worked during the short period just before the census reference date, some data users may prefer usual hours worked in a corresponding short reference period. This measure would be of the hours worked during a normal or

typical week, including overtime hours regularly worked, whether paid or unpaid. Days and hours not usually worked should be excluded as should unusual periods of overtime. In some censuses, questions have been asked about both actual and usual hours worked. Users of the statistics should be consulted regarding the most useful measure(s) and reference made to the ILO international standards (expected to be updated in 2008).

### 3. *Review of questions used in national censuses*

423. Questions on working time have been:

- (a) a single question on actual or normal working time during the reference period, on the main activity or all economic activities combined (examples I.1 to I.3);
- (b) a set of questions on actual and normal working time during the reference period, on the main activity or all economic activities combined (example I.3); or
- (c) separate questions on working time actual or normal for the main activity and for other secondary activities (example I.4). When using only a single question, the accuracy of the answers for working time will usually be poor. Asking and recording the hours actually worked each day in the reference period and in each job can improve the accuracy. However this procedure does take more time and space on the questionnaire than is often possible in a census operation.

#### *Example I.1*

<p><b>I.1.1</b> Last week, how many hours did this person spend working for pay or in self-employment?</p> <p><i>Include:</i></p> <ul style="list-style-type: none"> <li>• working directly towards the operation of a family farm or business without formal pay arrangements (e.g., assisting in seeding, doing accounts);</li> <li>• working in his/her own business, farm or professional practice, alone or in partnership;</li> <li>• working for wages, salary, tips or commission.</li> </ul>	<p>I.1.2 <input type="radio"/> None → Continue with the next question</p> <p style="text-align: center;">OR</p> <p>I.1.3 <input type="text"/> ← Number of hours (to the nearest hour)</p> <p style="text-align: right;">Go to Question <input type="checkbox"/> X.X</p>
---	---

424. This set of questions combines a basic activity question with the collection of working hours data. Although I.1.1 prompts very well on some of the problem categories of economic activity, and the later questions identify temporary absence from work and probe for unemployment, it does not allow for recalling other activities in addition to the main economic activity. In this setup, the emphasis may do better at determining who did “any work” than at measuring working time.

**Example I.2**

I.2.1 How many hours did you/. . . . work during that week?

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

99 Don't know/Not stated

(SKIP TO Q X.X)

425. The question I.2.1 deals with hours actually worked. It is not clear however, whether all work done or only the main job is to be included. The responses to this question will most likely be mixed, with some individuals focusing on one job and others including all jobs. Training of interviewers will need to aim at standardizing the way respondents are guided to provide the intended information.

**Example I.3**

I.3.1 How many hours per week did you generally work, in all places of employment? 25

hours   I worked  
a week

426. This question, on the other hand makes it clear that all employment should be included. It deals with usual hours rather than hours actually worked. Depending on how common multiple job holding is, it may be more effective to separate the main job from other jobs, to help respondents' recall.

**Example I.4**

I.4.1 Last week how many paid jobs did you work at?

Not stated jobs

I.4.2 How many hours do you normally work in your (main) job in a typical week, including overtime whether you are paid for it or not?

**FOR PERSONS WITH 2 OR MORE JOBS, OTHERWISE GO TO 42**

I.4.3 Excluding your main job how many paid hours do you normally work in your other jobs in a typical week?

Not stated paid hours, other jobs

427. The questionnaire, from which example I.4 was taken, was designed for data processing using optical mark readers (OMRs). It has a relatively detailed question on working hours and asks for main and secondary activities separately. It asks for the normal hours rather than the hours actually worked in the last week. A complete series of questions would be to ask for the actual hours first for both main and secondary activities separately and then the ‘normal’ hours. The sequence of actual followed by usual hours has been shown to improve accuracy in labour force surveys. The questions do prompt for the inclusion of overtime, but might have benefited also by including prompts on types of economic activities, as in example I.1, above.

#### **4. *Suggestions for enhancing effectiveness of questions***

428. In some countries pre-coded ranges have been used for the answers to this question to help the respondents understand that precise estimates are not needed and to avoid that the replies cluster on particular values, such as 35, 40 or 45 hours per week. This procedure may, however, provide another source of error as some interviewers with inadequate training may record the wrong category. There will also be a rounding error, as the mid-point of each range has to be taken as the basis for the calculation of averages, and other descriptive statistics. It is generally best to ask and record an estimate of the actual number of hours. The answer recorded should normally be rounded to the nearest hour as fractions or decimals of an hour imply a very dubious accuracy. The asking and recording of individual hours also allows a review of the detailed distribution of working hour. This will be particularly useful if some users of employment statistics wish to set a higher limit for the minimum hours worked for those persons who are to be considered as employed in certain tabulations rather than the recommended one-hour criterion<sup>88</sup> or to analyse full- and part-time work, excessive working hours or to measure working time in relation to a broader production boundary that includes “unpaid care work in households”. It is not recommended that a question using the categories part-time/full time be used on the questionnaire. Even with notes to interviewers, respondents often will apply their own, inconsistent interpretations of these terms. Output categories for tabulations can of course be called full time and part-time after reviewing the overall distribution of usual hours worked.

#### **5. *Issues of measurement***

429. Where working time is a new census topic, it is strongly recommended that it be included in the pre-testing programme. Testing should focus on evaluating the accuracy of the answers by conducting a second interview using more detailed, household survey questions and the wording, layout, prompts and examples most suitable for improving accuracy. In some developed countries, it has been found that responses giving few

---

<sup>88</sup> The available evidence, usually from Labour Force Surveys, of the distribution of workers according to hours actually worked has generally shown very low numbers of persons working very few hours per week, e.g., below 10 hours.

hours worked often tend to be too low, whereas responses with many hours worked tend to be too high. Little testing has been done in other countries, and is needed to focus particularly on time spent on activities that represent self-employment. The respondents' estimates for such activities are likely to be even more problematic in these countries than elsewhere.

430. It is, however, worthwhile to train interviewers to ask about each day and then add up the hours for all days before recording the answer. For rural farmers in many countries, interviewers will have to impute the time from other information, as these workers often do not have clocks and watches and can only answer in terms of sunrise, sunset, half-days, etc. It is acknowledged that the time spent on non-market activities, such as production for home consumption and fetching firewood and water, will be very difficult to assess accurately. Some countries collect working time only for paid employees in their census and rely on other sources, e.g., time use surveys, for measuring the hours worked of those with less standardized or supervised activities.

## **B Income**

431. Income is not a common topic in population censuses. It was only asked in 40 of the 181 countries reporting to the United Nations about censuses in the 1975-1984 period. {ILO COMMENT: UNSD TO UPDATE PLEASE} The topic is recommended in the *Principles and Recommendations, Revision 2*, but it is advised that only very restricted categories of individual income should be collected. In particular it is recommended that only cash income be collected and that non-cash income be excluded. The value of home production and of income paid in kind (e.g., food, clothing and accommodation), are extremely important components of the total income for individuals and households in most developing countries, but measurement is extremely difficult and complex even in household surveys. It is not the most efficient use of resources in respect of interviewing time and therefore not recommended for a population census.

432. The measurement of cash income itself raises many problems and the major ones are discussed below. For economically active persons it is recommended that only income from employment be collected for the individual. However, it is also recommended that, if possible, an item of total household income be collected, including any cash income from interest, dividends, rent, social security benefits, pension and life insurance annuity benefits.

433. It should be noted that censuses in some developed countries have a long history of collecting detailed cash income information at the individual level with few problems, e.g., Canada, United States of America and Australia. When detailed data are collected by the source of the income, the results will allow the identification of economically depressed areas, as well as providing statistics on the source of the incomes in these areas. This information, used together with other characteristics from the census, will help in the development of social and economic policies to assist such areas and/or to monitor the effects of existing programmes to help them.

434. Income is often thought to be a sensitive topic, but the degree of sensitivity seems to vary significantly between countries as well as between groups within the same country. Some groups, e.g., owners of relatively large businesses, do sometimes refuse to answer questions on the topic in any format, but paid employees and informal sector operators have been generally co-operative in most countries, within the limitations of their capacity to provide accurate estimates of their income, e.g., because they do not have written records.

**1. *The operational definition***

435. In light of the conceptual underpinning for the new international standards concerning the economically active population, income is defined as (a) monthly income in cash and/or in kind from the work performed by each active person and (b) the total annual income in cash and/or in kind of households regardless of source. The income from employment of economically active persons should include wages and salaries of employees, income of members from producers' co-operatives and the entrepreneurial income of employers and own-account workers operating business and unincorporated enterprises. In addition to the income from employment of its economically active members, the total income of the household should include, for example, interest, dividends, rent, social security benefits, pensions, and life insurance annuity benefits of all its members.

**2. *Requisites for applying the definition***

436. What income to measure is a major conceptual and practical issue. It is useful to distinguish between:

- (a) Income from paid employment
- (b) Income from self-employment
- (c) Other income (e.g., investment, including rental income; income from social security and insurance schemes; pensions; etc.).

437. The concepts involved in determining income are not simple to grasp and respondents may be unable or unwilling to provide exact information. If the aim is to measure solely income from current or usual employment then a measure of the first two items above is sufficient and the questions need only be targeted at the employed population (current or usual). If welfare is more the concern, then the other incomes must be included, and the question addressed to all individuals. However, this will make the questionnaire more complex. Note that rental income is not uncommon in many low-income areas of developing countries where a significant number of the aged or widows often earn a living by renting of rooms, etc., obtaining estimates of the net rental income is, however, often very difficult.

438. One approach is to ask for income from paid and self-employment from individuals and to obtain a broad measure of total cash income at the household level.

The cash income figure should also always be the amount before income tax or other, similar deductions (e.g., for social security or pension contributions).

- In some cases an employee may not know this gross figure easily but it can usually be obtained by reference to a pay slip.
- For the self-employed the income should be the gross output or takings, minus any operating expenses (e.g., on staff wages, materials/supplies, electricity, gas, water). In theory depreciation of capital assets should also be deducted, but this will only be possible from the self-employed who operate formal businesses with formal accounts).

439. Ideally one would usually like to measure current income with current employment and usual income with usual employment, i.e., the income measure should correspond to the employment measure. Except for some regularly paid employees, very short-term income measures (e.g., for last week or last month) can often be unrepresentative. Rarely is the same reference period for income received suitable for all the employed. For owners of large businesses a year is often the most convenient as they can obtain this from their accounts or tax records. A year is also the best period for income from activities with significant seasonal variations in earnings, e.g., agriculture. However, for small businesses, particularly in the informal sector, such a yearly measure is almost impossible to observe with their limited records. For employees an annual figure is also difficult unless they have a copy of the tax return easily available. One possibility is to vary the period according to the type of employment with either a current or usual measure. However, this would be more complex and take more space than is available on most census questionnaires.

440. The approach that can be adopted is to take a current, longer period, usually a month, and ask for details about incomes from all types of employment with notes to interviewers to gross up daily earnings and divide annual earnings, e.g., to obtain a monthly average. This does of course add to the complexity and inaccuracy of the resulting estimates.

441. It is further recommended that the income from all work activities be collected rather than for the main activity only. However, in censuses that include questions about secondary activity, it is very desirable to collect separately cash income from both the main and the secondary activity. This is particularly so in countries where the informal sector is important and often provide a secondary activity.

### **3. *Review of questions used in national censuses***

442. Pre-coding of income ranges is often used, particularly in some developed countries. It is obviously easier to record the answers when pre-coded ranges are given for the weekly, monthly and annual figures. It is also thought that respondents are less reluctant to give a range for income than they are to state the precise value. However



there is a loss of accuracy in obtaining ranges only, as the calculation of average incomes etc., normally will have to be based on the mid-point of the range. If ranges are to be used then good base data on the distribution must be available from pre-tests, household surveys or other sources to ensure that the ranges given are suitable for the relevant income distribution. In some developing countries ranges have been found to confuse interviewers and thus be a source of error. Good pre-testing is essential before their introduction.

**Example J.1**

**J.1.1** What was ...'s last pay/income period?  
 1  Weekly  
 2  Fortnightly  
 3  Quarterly  
 4  Annually  
 5  Other (*Please specify*)  
 6  None  
 7  Not stated

**J.1.2** What was ...'s gross pay/income during the last pay period that is before income tax or other deductions? (Present flash card).

Interview: For self-employed persons obtain 'net income' i.e., receipts less business expenses.

INCOME GROUP	0	10	20	30	40	50	60	70	80	90
	0	1	2	3	4	5	6	7	8	9

Don't know

443. This sequence is interesting in that it allows a variety of periods in the first question and then asks about the income for that period in the second question. The income information is collected in ranges and a flash (or prompt) card is used to assist in determining the range from the respondent. The use of prompt cards is a common survey technique also being used in some censuses. The questionnaire was designed for processing using Optical Mark Readers (OMRs). It is also interesting that a 'Don't know' category was allowed on the questionnaire, as this is not always a good practice because interviewers may overuse it, or respondents when uncertain may prefer to use this option instead of selecting the category most likely to be correct. Question J.1.1 is definitely worded for paid employees and interviewers would have to rephrase it for the self-employed. There is also no allowance for secondary employment and it is uncertain whether total income from all jobs was to be recorded or only income from the main job. This should be explained in the interviewers manual and in training. Even if secondary activities are rare some guidance is necessary for when they do occur.

**Example J.2**

J.2.1 Rendimento bruto do mês de agosto de 1991 da ocupação declarada no Quesito X.X	0 <input type="checkbox"/>	Não tem	Nº de dígitos	Cr\$	.00
J.2.2 Rendimento bruto do mês de agosto de 1991 de outra(s) ocupação(ões)	0 <input type="checkbox"/>	Não tem	Nº de dígitos	Cr\$	.00

J.2.1 Gross income for the month of August 1991 from the occupation declared in Question xx?

No. of digits \_\_\_\_\_ Cr\$ \_\_\_\_\_

J.2.2 Gross income for the month of August 1991 from other occupations?

No. of digits \_\_\_\_\_ Cr\$ \_\_\_\_\_

444. The question differentiates between the main job on which other economic characteristics have been collected, and any others that the individual might have had in a specified month. The question also requests gross income, which not all respondents might know, even if they were reporting on their own incomes. In general, unless it is common practice for employees to receive a printed account of their salaries and corresponding deduction, it should not be assumed that individuals will know their gross income.

**Example J.3**

<p>J.3.1 Are you a visitor to the household?  <input type="checkbox"/> Yes (<i>do not complete the remaining questions</i>)  <input type="checkbox"/> No (<i>go to next question</i>)</p> <p>J.3.2 Which kinds of income do you receive?  <i>(Tick all the boxes that apply)</i></p> <p><input type="checkbox"/> No source of income</p> <p><input type="checkbox"/> Salary/wages</p> <p><input type="checkbox"/> Income or profit from self-employment</p> <p><input type="checkbox"/> Pension from a former employer or spouse's former employer</p> <p><input type="checkbox"/> Child benefit</p> <p><input type="checkbox"/> Disability benefits such as:</p> <ul style="list-style-type: none"> <li>* disability living allowance</li> <li>* disability working allowance</li> <li>* attendance allowance</li> <li>* severe disablement allowance</li> <li>* invalid care allowance</li> </ul> <p><input type="checkbox"/> Other state benefits such as:</p> <ul style="list-style-type: none"> <li>* retirement pension</li> <li>* Widow's pension</li> <li>* family credit</li> <li>* unemployment benefit</li> <li>* income support</li> <li>* sickness/invalidity benefit</li> <li>* maternity allowance</li> </ul> <p><input type="checkbox"/> Interest or annuity from savings or investments</p> <p><input type="checkbox"/> Other sources of regular income such as:</p> <ul style="list-style-type: none"> <li>* student grant or scholarship</li> <li>* rent from property</li> <li>* maintenance</li> </ul> <p><input type="checkbox"/> Other sources not already mentioned</p>	<p>J.3.3 What is the total gross income  <i>(before any deductions for Income Tax and National Insurance contributions)</i> that you usually receive from all sources?</p> <p>Count all income including:</p> <ul style="list-style-type: none"> <li>* salary or wages</li> <li>* income or profit from self-employment</li> <li>* occupational pensions</li> <li>* child benefit</li> <li>* disability benefit</li> <li>* state retirement pension</li> <li>* family credit</li> <li>* unemployment benefit</li> <li>* income support</li> <li>* sickness/invalidity benefit</li> <li>* maternity allowances</li> <li>* Interest or annuity from savings or investments</li> <li>* student grant/loan</li> <li>* rent from property</li> </ul> <p>Do not deduct:</p> <ul style="list-style-type: none"> <li>* taxes</li> <li>* national insurance contributions</li> <li>* superannuation payments</li> <li>* health insurance payments</li> </ul> <p>Per week                      or                      Per year</p> <p>Nil    Nil</p> <p>Less than £40    Less than £2,000</p> <p>£40 - £79    £2,000 - £3,999</p> <p>£80 - £119    £4,000 - £5,999</p> <p>£120 - £159    £6,000 - £7,999</p> <p>£160 - £199    £8,000 - £9,999</p>
---	---

445. In this example for self-enumeration the questions cover both ‘sources of income’ and ‘total gross income’ by ranges, but separately. In the USA and Canada it has been common to ask for the detailed income from each source. Questions with pre-coded ranges do take space for the ranges and, as shown, explanatory notes are extensive. The exclusion of visitors in J.3.1 is interesting. This could also be done during processing provided that visitors were identified elsewhere on the questionnaire. The aim presumably was to provide income data only for usual residents and to sum to household totals without visitors.

**Example J.4**

J.4.1 INGRESOS POR TRABAJO

¿Cuánto gana o recibe (NOMBRE) por su (s) trabajo (s) (actividad (es))?

*Anote en pesos la cantidad recibida y el periodo  
Si no recibe ingresos anote "0" en Periodo*

PERIODO: A la semana ..... 1  
A la quincena ..... 2  
Al mes ..... 3  
Al año ..... 4

\_\_\_\_\_  \_\_\_\_\_   
Anote con número Periodo

**J.4.1 Income from work**

How much did (Name) earn or receive for his/her work (activity(ies))?

Record in pesos the amount received and the period.

If received no income record ‘0’ for the period.

Period: A week .....1  
A fortnight .....2  
A month .....3  
A year .....4

\_\_\_\_\_ \_\_\_\_\_  
Record the amount Period

446. No mention is made of income in kind (e.g., food, clothing) in either of the above example of questions. This may have been excluded and was not stated presumably because most respondents do not think of this as income. However, the fact that this type of income is to be excluded should still be noted in the manuals and during the training for the rare cases when interviewers are asked about it. The prompts to be used on the questionnaire, and the exact questionnaire wording, needs to be established through pre-tests. Note that as with other questions every effort should be made to avoid technical jargon and concepts in the actual questions.

447. To include an income question when the questionnaire has a landscape format can be difficult because of the space needed for the essential notes to interviewers. However, the recording space for answers is usually adequate. No country example of an income question using the landscape format was located, however.

**4. Suggestions for enhancing effectiveness of questions**

448. The *Principles and Recommendations, Revision 2* recommend household income measures as well as individual measures. Household cash income is a major general welfare measure, and is needed for studies on poverty as well as other welfare related issues. Depending on the question asked at the individual level, the household estimates may be derived by summing the individual measures; or as suggested earlier, a separate, broader question may be asked at the household level, if there is an opportunity to add in other income that is not included in the individual measures.

**Example J.5 (Untested, hypothetical question)**

<p><b>J.5.1:</b> <i>Please estimate the total household <u>cash</u> income for the last twelve months from all sources including any income from employment mentioned earlier for each person but add on any other income from pensions, rent or other investments</i></p> <p><b>Interviewer:</b> <b>Work with respondents to add individual cash incomes plus any cash income from any type of pension/ welfare payments or insurance plus rental income plus any other investment income.</b></p>	<p style="text-align: center;">TOTAL HOUSEHOLD INCOME \$ .....</p>
---	--

449. The example is a hypothetical untested question that may however assist in the development of a suitable household question in some countries. It must be stressed that the decision to include an income question should not be made lightly. Pre-tests on this topic should be extensive, particularly when the topic is asked for the first time in a census. If possible pre-tests should be followed by extensive re-interviews using household survey methodology to establish the probable limited accuracy of the census measures and thus to advise users. Early pre-tests should establish the prompts necessary for interviewers and their problems with agricultural and informal sector income in particular. As interviewers are required in some cases to make calculations, a check on their accuracy is also essential. Users of the resulting statistics should be consulted during testing to ensure the question(s) are eliciting the type of information that they require and with a degree of accuracy that they regard as acceptable.

**5. Issues of measurement**

450. It must be noted that even with the restriction to cash income only, the income information in a census is usually very approximate. Regular wage earners may know their regularly received pay, but the self-employed and particularly the small informal sector operators are very unlikely to be able to give good estimates of their income. This also applies to agricultural income, as it is also often seasonal. Measurement of income received for casual, temporary and intermittent work will also very often be approximate. A major problem that often is encountered is that one member of the household gives the census replies for all household members. This person does not often know other household members income and only very rough estimates, if any, may be obtainable. Due to these accuracy problems some countries have restricted their income measures to

paid employees only, even in labour force surveys. Others have excluded agricultural income and only measured this component in specialized agricultural surveys.



**PART 4: DATA PROCESSING, TABULATION,  
ANALYSIS AND DISSEMINATION**





## **Chapter 9. Data Processing, tabulation and other considerations**

### **A Introduction**

451. Data processing relates to the activities in converting census responses into a computer readable data file of records that have data that are as free of errors as possible and that are ready for tabulation and further analysis.

452. Unless advanced technologies are used for data collection (for example on-line Internet completion and return of electronic questionnaires, automated telephone interviewing, hand held devices, etc), most countries will have paper questionnaires that must be converted by key entry and/or optical recognition into computer data files for analysis and tabulation. This data entry process also applies to other census topics and is covered in *Principles and Recommendations for Population and Housing Censuses, Rev 2*. Consequently, they are not mentioned further in this chapter.

453. When precoded responses are not given for a particular question, clerical, computer-assisted, or automated coding may also be required. This especially relates to occupation and industry and is covered in Chapters 10 and 11.

454. Because of the size and complexity of census operations, it is likely that errors of one kind or another may arise at any stage of the census. This chapter outlines specific points to identify and correct as many of these errors as possible (data editing and amendment), the derivation of selected data from census responses (for example, whether a person is employed in the informal sector), and the output phase (tabulation and dissemination of census results).

455. Every national census organization should establish a system of quality assurance and improvement as an integral part of its census programme. The data processing to detect and correct errors is only part of that programme. There is no single standard quality control and improvement system that can be applied to all censuses or even to all steps within a census. Census designers and administrators must keep in mind that no matter how much effort is expended; complete coverage and accuracy in the census data are unattainable goals. However, efforts to first detect and then to control errors should be at a level that is sufficient to produce data of a reasonable quality within the constraints of the budget and time allotted.

456. Automated data capture, repair and coding systems (see later in this chapter) both increase greatly and introduce a different set of risks to data quality compared with traditional census processing approaches. If not properly monitored and managed, data quality problems can remain undetected until late in the process when cost and timing constraints limit the options for any corrective activity.

457. Some methods of measuring data quality from data capture processes, such as substitution rates or measures of key entry errors, are inadequate as these forms of monitoring simply measure the overall incidence of errors but not the significance of the

errors. Indeed this approach could lead to considerable extra expenditure for the correction of trivial errors that lead to no appreciable gain in quality. For this reason, data quality should be measured at the topic response level rather than at the individual character/numeral level. This should be done in two ways: independently processing a sample of records using manual processes and comparing the results for each of the records with those obtained through the automated systems; and in aggregate by comparing the overall data for an area with the expected results based on other information for that area (e.g. from the previous census or other data sources). Where manual or computer assisted processes are used, it is useful to establish a quality control system in which a sample of records are coded independently by a second operator. Any discrepancies in the results can then be resolved by an expert or experienced operator. This will allow estimates of error rates to be calculated as well as the identification of problem areas and cases where re-training may be necessary.

458. This process should be undertaken continuously during processing with a focus on early detection of quality problems and an understanding of any systems or processes that have contributed to these. The amount of error that is acceptable and the degree of intervention and systems or process changes undertaken will depend on the assessment by the census agency of the overall fitness of purpose of the output and the overall cost and timeliness impacts. This will vary from topic to topic.

459. Chapter 2 (concerning planning and design for a population census) has already mentioned the need to involve data processing staff at an early stage of census planning, and the importance of appropriate form design (to facilitate data entry and with sufficient space for codes to be entered). Chapter 2 also mentioned the possibility of processing a sample to produce advance results. This chapter covers more detailed aspects of data processing for particular topics and the tabulation of census results.

## **B Data processing aspects for particular topics**

### ***1. Current and usual economic activity status***

460. The usual activity reference period and the current activity reference period do not necessarily overlap, e.g., if the period for usual activity covers to the end of the previous month and the period for current activity is defined as the previous seven days from the enumeration day. In this gap the person may have changed activity to, e.g., become employed or unemployed or retired. More generally, a person's economic activity status during the longer reference period for usual status may be different from that person's current status over the shorter period used for determining their current status. Thus, for those countries collecting both current and usual economic activity status, it is not recommended to force any consistency between these data items during data processing.

461. Also, most of the not economically active categories can change status from the usual measure to the current measure and consistency here should not be forced either.

For example, a person could be a student all months of the usual measure but be engaged solely in housework, or have started work in the seven days of the current measure.

462. If retired/aged is given as a category for the not economically active, it is strongly recommended that a lower age limit is put on this category (e.g., 45 years) for the editing procedures. The limit should not be set too high as some people are retiring or being retrenched from wage employment earlier in recent years. It would be particularly wise to remove persons in the younger age groups from this category as their inclusion may make the results look silly, e.g., having persons ‘aged or retired’ who are 20 years old.

463. Cross-tabulations of usual by current activity are a useful method to check for large numbers of unusual cases before publication. It may be necessary to go to small area data to find explanations or possible errors of interpretation, e.g., a new large factory or a natural disaster may explain odd looking differences between usual and current activities in particular areas. These are worth explaining in the presentation of the results.

## 2. *Employment*

464. The determination of whether a person is currently employed may depend on the answer to more than one question (especially if there is a separate question to identify those who are temporarily absent).

465. The determination of whether a person should be classified as employed or not is almost always as a derived variable. All possible combinations of answers to the questions on employment should be allowed for. The derivation should be automatic (by computer program) and will also check that the proper sequence of responses has been followed. Inconsistencies between responses to different questions (including questions being answered when they should be skipped over) must be resolved.

466. It is useful to do an age distribution of the categories of the employed. Extreme occurrences, e.g., employed persons aged over 80 years, and any unusual groupings at unusual ages in general should be investigated.

467. Tabulations of the characteristics (age/occupation etc.) of the temporarily absent should be done as part of evaluation of this category.

## 3. *Unemployment*

468. Given the wide variations in the reference periods used, especially when two criteria are used in the questions to derive the unemployed, and in the way either one or two of the criteria may be applied, the definition and more importantly the question(s) used need to be presented when the results are disseminated. Also, when questions on both “seeking” and “available for work” are used it would be useful and would facilitate comparisons across countries if some of the tables show the two components, as well as the resulting estimates when they are combined to identify the “unemployed”.

#### **4. *Status in employment***

469. It may be necessary to derive the standard status in employment categories from some of the responses given. This is relatively simple with modern tabulation software and most statisticians and data processing staff are now accustomed to much more complex derivations from survey data.

470. There are no viable consistency checks for the International Classification of Status of Employment categories using age, sex or education level because the categories can be found among employed persons of all ages, educational levels and sex.

#### **5. *Occupation***

471. As noted in Chapter 11, it is strongly recommended that countries code their occupational data in a way that makes it possible to produce statistics in accordance with the International Standard Classification of Occupations. ILO has provided considerable assistance in the use of occupational classifications, with the major documentation being contained in Chapters 10 and 11 below, Hoffmann and others (1995), and Hussmanns and others (1990).

472. Data consistency edits can be made between occupation groups and education levels: e.g., most persons classified to a 'professionals' category will have educational attainment which is higher than 'no education' or 'primary school' level. Care should be taken in the use of such edits, however, so as to avoid the risk of biasing the relationship between variables such as level of educational attainment and occupation. Analysts may be particularly interested in identifying the number of persons who do not have the level of education normally expected for particular occupational groups. Checks are also possible between occupation and sector of employment: e.g., 'government tax and excise officials' should be in the government sector. (Note, however, that in the 'government' sector there will be a large number of jobs given occupation codes other than that for 'government administrative worker'.) Occupational information should be coded to as much detail as made possible by the recorded responses (see Chapter 10 for further details).

473. As indicated earlier, any unusual groupings in general should be investigated.

#### **6. *Place of work***

474. Checks of the consistency between the type of place of work and other selected variables (such as occupation, institutional sector) may identify possible errors to be investigated. (For example, doctors or public sector employees operating from a market stall).

475. Similarly the geographic location of place of work would normally be located within reasonable commuting distance of the place of residence, and large discrepancies

would need to be explained when disseminating results. An exception concerns those commuting workers who come home only for the weekend but who work far from home.

476. In respect of tabulations, the main categories of type of place of work are useful information in themselves for labour market analysts. The cross-classification of these location categories by status in employment and industry and occupation will give very valuable information on the structure of employment. This will be the case particularly for the informal sector, and also for the identification of other categories that often are of special interest, such as 'paid employee home-workers'. These target groups should be clearly identified early during the census preparations and the processing steps necessary to extract information about them should be tested, to ensure the target groups are being correctly identified.

## **7. *Industry***

477. Industry coding in population censuses and surveys is discussed in some detail in Chapters 10 and 11. There are useful suggestions also in Hussmanns and others (1990), as well as in other ILO papers and documents. Although this is an established topic, there may be problems with industry data collected in censuses and the points discussed below should be particularly noted.

478. The industry description and code should refer to the establishment where the person actually works, and not the legal unit to which the establishment may belong. For example, a major company in a country, 'XYZ Ltd', may be well known but engage in a variety of activities at different locations around the country. Generally industry should be coded for each separate establishment (usually at separate locations), not only to the main activity of XYZ Ltd. With access to a reasonably up-to-date register of establishments, where industry is coded for each unit, the industry question should aim to collect the detailed name and address of the person's place of work. Then coding can be done by reference to the register and be compatible with employer surveys based on the same register. This is discussed further in Chapter 10.

479. Unfortunately few countries have a comprehensive updated register of employers with separate identification of all establishment locations. Thus, in most countries the detailed name and address of the establishment can only be used as a source for the correct industry code for some of the responses, and a description of the activity at the establishment has to be used as the basis for assigning codes for the rest.

480. However, even if a register of establishments (or of employers) is very poor, it is often useful to have a limited list of major companies/enterprises with their industry codes when creating the coding indexes described in Chapter 10. For example, XYZ Ltd may be dominantly an iron and steel mill but may also have XYZ Coal Mines and XYZ Iron Ore Mines as separate establishments. Entries in the coding index for industry reflecting this structure will help to ensure that respondents working at these separate establishments are given their relevant industry codes and not simply coded to the code given to XYZ Ltd. A similar listing of major government departments and organizations

is also very useful in coding respondents working for the government to their correct industry codes: e.g., to code those working with the country's health administration separate from those working in government hospitals.

481. There are a limited number of consistency checks of industry codes with the results from other questions. These include checks that persons coded to government administration are in the government sector (but note that many government sector workers will work in other industry groups), and that only certain occupations are coded to the industry category for 'private households with employed persons'. Other edits are generally not definite and depend on the situation in that country. Particular attention should be paid to any codes identifying separately 'production for home consumption', including subsistence agriculture.

482. Informal sector derivations should be carefully checked to ensure that only relevant industries are shown. Some limited checks are possible of industry against status in employment: e.g., self-employment; cannot occur for 'public administration', 'private households with employed persons' or for 'international organizations'. A full range of edits should be prepared before data processing commences but cross-tabulations with other topics should be done early and progressively to locate inexplicable occurrences which will need explanation or possibly further edits. It is recommended that this is not left to the final tabulation phase because the resolution of these queries can delay the results considerably.

#### **8. *Institutional sector***

483. If the question on institutional sector is combined with the status in employment question(s) as presented in some of the examples shown earlier, it will not be necessary to do a consistency check between status in employment and institutional sector (e.g., government sector workers being classified as self employed or unpaid helpers). There are no possible checks against sex, age or education level. Checks against the industry and occupation classifications are also possible as discussed earlier.

#### **9. *Informal sector***

484. The design of a procedure to measure employment in the informal sector should be clearly thought out and well tested. Some computer derivation may be necessary and will need very careful consideration and prior testing.

485. Careful checks and cross-tabulations of status in employment categories by industry and occupation groups will be essential to edit any informal sector data and re-coding is likely to be necessary to remove absurdities.

486. If the derivation procedure is properly designed then consistency checks with institutional sector will not be needed, otherwise consistency checks will be required to ensure that there is no public sector or company employment in the informal sector.

487. Such preparations will prove to have been well worth the extra effort when the demands for small area and regional estimates of informal sector employment are received.

#### **10. *Informal employment***

488. The procedure to identify those in informal employment should also be carefully designed. There are no viable consistency checks since those in informal employment may be in any occupation, industry or institutional sector. If information is collected to derive whether a person is in the informal sector, then it would be useful to check that all who are engaged in the informal sector are also classified as informally employed.

#### **11. *Working time***

489. The following consistency checks should be made for this topic: If a person is recorded as temporarily absent from their main job then the actual hours recorded for that main job should be zero. If a person has more than one job reported then working time should be recorded for all jobs for both hours actually worked and usual hours. If hours actually worked are recorded separately for all jobs then it would be reasonable to expect that usual hours would also be recorded for these jobs as well.

490. Consideration should be given to compiling statistics of average hours actually worked as well as distributions of the employed in five-hour groupings of hours actually worked. In presenting the distributions of employment by working time, care should be taken to ensure that the distribution shows both low hours as well as excessive hours.

#### **12. *Income***

491. When income is collected for the same reference period as the measurement of economic activity status, then only persons working in paid jobs or in self-employment, according to the status in employment variable, should have income from employment. Unemployed and inactive persons may receive other types of income, but not income from employment. Persons with secondary paid or self-employment jobs should have income also from this second activity.

492. In cases where the reference period for the measurement of income differs from the reference period for the measurement of economic activity status, then no viable consistency checks are suggested.

493. Total household income should always be equal to or more than the sum of the individual job-related incomes in the household. Codes for 'not stated' must be allowed for this question in particular, with a clearly defined strategy on how to deal with such cases in the processing of the question, e.g., by imputation or by exclusion from averages.

494. Extremely low and high incomes should be checked since these can distort

averages. These checks can be done by status in employment and by occupation.

## **C Dissemination**

495. A census is not complete until the information collected is made available to users in a form, and to a timetable, suited to their needs. Thus in disseminating the results of the Census much emphasis should be put on responsiveness to users and on high standards of quality in the production of statistics. One major strength of a census is that it supplies statistics for small areas and small population groups, and the tabulation and dissemination systems should make full use of this by generating detailed tables of results.

496. Annex E lists the tables relating to economic characteristics that are recommended for in the *Principles and Recommendations for Population and Housing Censuses, Rev 2*. (The annex also includes two tables, one on the economic activity of persons with disabilities and one relating to economically active migrants.)

497. The tables in Annex E have generally been recommended for production for national, regional and provincial levels with urban/rural dissection, and even for production at locality level. Some are in considerable detail (for example, those cross-classified by occupation and industry) and therefore it would be more appropriate to release these more detailed tables electronically (see below) or in publications and abstracts that relate only to those small geographic areas.

498. As indicated in *Principles and Recommendations for Population and Housing Censuses, Rev 2* (paragraphs 1.208 to 1.209), there are several ways of making the results of a census available to the user:

As printed reports containing standard and pre-agreed tabulations, usually at the national, regional or local district area level, that may be obtained from government agencies or directly from booksellers;

- (a) As unpublished reports (often referred to as abstracts) comprising standard tables but produced for either smaller geographies or population sub-groups not otherwise included in the published reports – these may often be requested by users who may have to contribute towards a proportion of the marginal costs of their production;
- (b) As commissioned output produced from a database, comprising customized cross-tabulations of variables not otherwise available from standard reports or abstracts; and
- (c) As micro data, usual available in restricted format only and supplied under strictly controlled conditions.

499. A range of products and output media should be available to meet the changing requirements of users. Major users should be consulted in advance to determine these needs. Users will also need supplementary meta-data covering definitions,



classifications, and coverage and quality assessments.

500. Each of these different forms of output are covered in the following discussion.

**1. *Printed publications***

501. Due to their ever increasing production costs, printed publications may become less the preferred choice for the dissemination of the main census results, though paper still provides a media that does not readily deteriorate and does not require the user to have any necessary hardware, software or technical skills.

502. It is suggested that only the basic and recommended set of tables should appear in the printed publications that are prepared for broad user and community use, and that even then these tables do not need to include statistics for small geographic areas. National, regional and provincial statistics would generally suffice.

503. These printed publications (and output via the Internet CD-ROM, etc) should include graphs, maps, and textual analysis as appropriate.

**2. *Electronic dissemination***

504. The release of some outputs (especially of detailed cross-classified results and results for small population groups, and some of the optimum set of tables) may be possible only by distribution through the use of high capacity electronic media. However, when data are provided in electronic form, special attention should be given to providing users with easy means of data retrieval.

505. The outputs and relevant meta-data should be accessible in standard formats as well as in common database and spreadsheet format for easy retrieval and manipulation.

**3. *Micro data***

506. National statistical agencies should also consider the possibility of releasing data files of unidentifiable unit records so that advanced users can undertake more detailed analysis. Such releases should be subject to careful controls on data security and confidentiality.

**4. *Confidentiality and data security***

507. The computer systems handling census data should have strict safeguards to prevent unauthorized access to the information. Care should be exercised to avoid the inadvertent disclosure of information about identifiable individuals through the statistical results of the Census. Special precautions may apply particularly to statistical output for small areas.

508. Measures to ensure disclosure control may include some, or all, of the following procedures:

- (a) swapping of some records between households of similar demographic characteristics within the same higher geographical area;
- (b) restricting the number of output categories into which a variable may be classified, such as aggregated age groups rather than single years of age;
- (c) where the number of people or households in an area falls below a minimum threshold, suppressing statistical output – except, perhaps, for basic head counts – or amalgamating with that for a sufficiently large enough neighbouring area;
- (d) randomly modifying or rounding data before the statistics are released; and/or
- (e) in the case of micro data or public use samples, removing all information from databases relating to name, address and any unique characteristics that might permit the identification of individual respondents.

#### **5. *Graphical and related output***

509. Products should be developed that allow statistical and geographical information to be delivered together with geographical information systems to meet as widespread an interest, and with as much flexibility, as possible commensurate with assurances on confidentiality. By having associated graphing and mapping capabilities, databases will greatly increase their usefulness.

510. Ideally users should, themselves, be able to generate graphs and/or maps easily, and then to print or plot them or make the images available for other uses. Several census organizations have produced this kind of product, sometimes in co-operation with a commercial company.

511. As already mentioned, printed and electronic output should include summary graphs and maps.

#### **D Meta Data**

512. A meta-data system contains especially definitions of terms, classifications, and nomenclatures. Classifications and nomenclatures are used for unified classification of data. For indicators for which international standard classifications had been created, these classifications may be used. For indicators that cannot be classified by international standards, new nomenclatures must be created.

513. A meta-data system provides supplementary information on characteristics of surveyed and published data. Such information is essential for users to understand the strengths and limitations of the statistics and to use the data properly in more detailed analysis.

514. The population census for the 2010 round of censuses has to ensure comparability with the data from the previous population censuses, to the extent possible. On the other hand, it should include new elements relevant for development that has taken place during the time since the previous census. Thus, the meta-data for the population census around 2010 should follow the meta-data system from the previous population census with an update in line with the needs resulting from the development since the previous population census. With the rapid changes in information technology that are taking place, the design of the newer meta-data system may also need to change.

## **E Emerging technologies for data processing and output**

515. Technology has been used to assist in all phases of population censuses for many years, and particularly in data processing, analysis and dissemination. Well established approaches and technologies might well be the most viable option for many countries and adoption of new technology or approach should only be considered where there is a sound understanding of the new approaches and technologies and the developments can be managed. There should also be a clear understanding of both the risks and the benefits.

516. This section summarizes only some of the current technologies (for more information, see Principles and Recommendations for Population and Housing Censuses United Nations, New York 1998 and Handbook on Census Management for Population and Housing Censuses, United Nations, New York 2001). Further technological development is certain and will affect future census taking.

## **F Data capture and coding methods**

517. Several data capture technologies have traditionally been used such as key entry and optical mark recognition (OMR). The last decade has seen significant improvements in intelligent character recognition (ICR), data repair, imaging and automated coding technologies that have reduced the cost of census processing and the time taken to do it. They have also improved data quality. These trends are likely to continue.

518. For most countries, the most cost-effective option is likely to be a combination of digital imaging, ICR, repair and automated coding whereby the census forms are first processed through scanners to produce an image. Recognition software is used to identify tick box responses and translate handwritten responses into textual values. Confidence levels are set to determine which responses require further repair or validation. Automated repair typically involves the use of dictionary look-up tables and contextual editing. Operator repair can be undertaken on images not recognised. This is only cost-effective for those questions where there is a high probability that the repaired data can then be automatically coded.

519. Automatic coding uses computerised algorithms to match captured responses against indexes. Those responses that cannot be matched are then passed to a computer assisted coding process. This methodology eliminates inconsistencies caused by human

error inherent in manual and computer assisted coding. The quality of the automated capture and coding needs to be carefully monitored during processing to ensure that the system is functioning as specified. However, the process does offer the opportunity to improve the quality of the data and has the potential to significantly reduce processing costs.

520. In computer assisted coding, a coding clerk typically key enters a truncation of selected words from responses written on the form or captured in some other way. Matching index entries are displayed on the computer screen and, when the index entry most closely matching most closely matching the response is selected, the appropriate classification code may be automatically written to a data file or database.

521. Census agencies also need to consider how the data are going to be held through the processing stream. Traditionally, census processing has been conducted using a flat file that gets progressively updated with the earlier version of the file retained for backup and recovery processes. Typically this has been allied with batch processing where a discrete group of forms (typically for an enumeration area) are processed together. Thus the forms will be data entered, edited and coded as a group. This allows a high degree of workload control. Databases allow information to be held and processed at the individual field level. This provides a greater degree of flexibility, as once census data is electronically captured, it is easily organised to maximise both processing efficiency and quality as similar responses can be readily grouped and coded together. However, holding the census data in a database requires more complex systems to manage and deliver work. Consideration also needs to be given to backup and recovery mechanisms.

522. These systems typically require far more extensive systems development and testing than traditional census processing system. There are a number of factors that need to be considered through the systems design and integrated into the systems development such as the work organization of the remaining clerical processes.

## **G Output**

523. Traditionally, census output has been conceived in terms of generating tabulations, usually for sophisticated clients well acquainted with the census data, its structure and content and other meta-data. Less sophisticated users traditionally relied on static products such as publications that generally contained a limited range of outputs.

524. Internet dissemination allows for a much improved service to users by permitting the design of appropriate products to meet the needs of different types of census data users from novice users to sophisticated users, the cost-effective dissemination of a much wider range of census data, and an improved usability of the census data.

525. As already mentioned, the main strength of a census in a developed statistical system is to complement the information provided by other data collection methods such as surveys with a focus on small domain statistics – that is information for small geographic areas and for small population groups (both social and economic). Internet

dissemination can support both types of use of the data. For small geographic areas, GIS technology can be used as means for both defining areas of interest in searching for data and for mapping of the outputs of the search. There is a range of packages that can be used to hone in on populations of interest from large pre-defined matrix tables.

526. The Internet dissemination system should provide freedom for clients to specify the form of the output – whether as hard copy or a data file that can be exported into a range of commonly available statistical analysis, tabulation or mapping packages.

527. Some countries may wish to consider providing access to clients to submit tabulations directly on-line to be run against the census unit record file. As already indicated above, protecting the confidentiality of the census data is a prime consideration in such systems. As well as implementing confidentiality procedures (such as random rounding), there may be a requirement to limit the size of tabulations that can be submitted through this method.



# **PART 5 CODING OF OCCUPATION AND INDUSTRY**

## **Chapter 10: Preparations For Coding Occupation And Industry**

528. This chapter discusses the different strategies and preparations that are necessary to ensure effective and reliable coding of the industry and occupation responses, as well as the costs and advantages of the different alternatives. It outlines the objectives and main strategic choices to be made, and presents the main organizational factors determining the effectiveness and success of the coding and processing tasks. The chapter describes the development of the major tools, the coding indexes, and how to use them effectively.

529. Among the basic tools available are international classifications, i.e., International Standard Classification of Occupations (ISCO), and International Standard Industrial Classification (ISIC)<sup>89</sup>, which are available on the Internet. Given that requirements will vary from country to country, the chapter covers common issues that must be addressed in any census operation when questions related to occupation and industries are included in the census questionnaire.

530. At the time of release of this manual the most recent versions of these classifications are ISCO-08, to be published in 2008 and ISIC, Revision 4, endorsed by the UN Statistics Commission at its 37<sup>th</sup> Session in 2006. In general, it is preferable to adapt these international classifications for national or regional use than to use them directly in national statistical collections. In many countries such adaptations may already exist (or be under development) for use in economic and labour statistics. Wherever possible such adaptations should also be used for the Census. Some countries may still be using national classifications in their economic and labour statistics based on earlier versions of the international standards (E.g. ISCO-88 or ISIC Rev 3.1) and will need to make a decision whether or not to update their national classifications in time for use in the 2010 round of Censuses.

### **A Objectives**

531. The main aim of the coding and processing of the information given by the respondents about their place and type of work, i.e., about the industry and occupation in which they work, is to determine and record correctly to which of the groups in the respective classifications the jobs belong, at the most detailed level of the classification possible on the basis of the information provided in the response. This task has to be completed within an overall processing plan for the census, to a pre-specified timetable and either within pre-specified cost limits or in a fashion which will minimize cost, given the specified data requirements.

---

<sup>89</sup> <http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=2&Lg=1>.



532. The development of a census processing strategy with these aims needs to consider many aspects and requirements of the processing task, such as:

- (a) the existing data processing capacity and infrastructure;
- (b) the type and format of the information to be processed;
- (c) the volume of data to be processed and the throughput rates required;
- (d) how the processing of the information about industry and occupation is embedded within the total data processing task for the census.
- (e) the level of detail required to satisfy important user needs in the national context, as well as for international reporting.

533. The precise impacts of these aspects will depend on the choices made with respect to some strategic aspects of the coding and processing of census forms, discussed in the following section.

## **B Strategic coding and processing options**

### ***1. Process all cases or a sample only***

534. Coding of industry and occupation is typically one of the most expensive and time-consuming operations in the processing of a census. In order to reduce costs, to make the management and quality control of coding easier and to enable results to be produced earlier, consideration should be given to obtaining and/or coding the information for a sub-sample only.

535. A policy to code a sub-sample only may be implemented at the data collection stage by fielding longer and shorter versions of the census form, so that only a sample of the population answers the questions designed to capture the information needed to code industry and occupation. To avoid error-prone field sampling procedures and administration of different forms or schedules one may alternatively collect the information from everyone, but process it for a sample only. Either form of sampling will significantly, almost proportionately, reduce the operational costs of coding, but not the costs of preparing for the coding operation. However, sampling means the introduction of sampling related errors (i.e., imprecision in the resulting estimates), and this will be an important concern when estimates are to be produced for small population groups or small geographic areas. As the provision of small area and small group data on a nationally consistent basis is seen as one of the major functions of the census in most countries, many users may see the processing of industry and occupational information for a sample only as being completely against the objective of the census unless the sample size is sufficient to meet these requirements. This will especially be the case if there is a regularly conducted, annual or quarterly labour force survey that already provides sample-based national statistics for employment and unemployment, classified by occupations and industry.

536. Other considerations relevant to the coding of a sample only (or a sub-sample) are the following:

- (a) The use of a sample for the coding of industry and occupation will mean that sampling imprecision will make inter-group and over time comparisons more difficult or uncertain variables.
- (b) Sampling may involve a serious loss of descriptive power, particularly for small areas, because it may hinder the production of summary socio-economic indices for small areas or groups, as these are based on the industry and occupation variables.

537. If the decision is made to collect industry and occupation information only for a sample of the population, then attention must be given to the design of the sample, balancing the requirements of statistical precision against those of operational simplicity and robustness. The sampling fraction should in principle be determined by balancing the precision of estimates required for the smallest aggregates of the population for which separate figures are to be produced, against the saving of cost and time resulting from reducing the processing load. For operational convenience this has often been done by selecting all households in a whole enumeration area to be included in the sample, thus sending whole bundles of questionnaire into one processing stream or the other. This may, of course, result in very high clustering to particular occupations and industries, and will give valid statistics only at high geographic levels, thus contributing to further loss of precision.

538. If the sampling within enumeration areas is done manually it is also important that the sampling procedure be simple (e.g., by selecting every  $n$ -th case, with  $n = 5$  for 20 per cent sample, 10 for 10 per cent, 20 for 5 per cent sample, as appropriate). The sample selection should be done on a probability basis. In practice this can conveniently be approximated by taking every  $n$ -th census household in each data collection area (enumeration district) and include all members of those households in the processing sample. In this case, due to variable household size and to the clustering of individuals within households, there will be a departure from a true simple random sample of individuals (the units of analysis for occupational statistics). An effort should be made to estimate the sample design effect, but guidance on how this should be done is beyond the scope of this publication.<sup>90</sup>

## 2. *Field or office coding?*

539. In censuses there is a choice between:

- (a) the respondent coding himself/herself to a predefined group;

---

<sup>90</sup> For details on design effects, see for example, Kish, Leslie (1995). *Survey Sampling*. United States of America: John Wiley and Sons, Inc. Wiley Classics Library Edition Published 1995.

- (b) the enumerator coding in the field, either during the interview or before the questionnaire is forwarded for further processing;
- (c) coding by specially trained coders in connection with consistency checks of the questionnaire and data entry.

540. The choice is a strategic one with respect to balance between costs, quality of coding and of the resulting statistics, and control of the coding process.

### *2.1. Coding by the respondent*

541. In practice, coding by the respondent means that respondents are requested to place their job in one of a set of predefined groups presented to them in written form on a questionnaire, or on a card read or handed over by the enumerator.

542. The main advantage of this approach is that it is the least expensive of the possible coding procedures. It can easily adapt to OMR (optical mark recognition) resulting in even greater cost savings. The main disadvantage is that it results in lower quality data, in terms of both reliability and validity.

543. The lower level of reliability is due to the difficulty in assuring consistency in how the respondents relate the pre-defined groups to the job they are asked to give information about. The content of each pre-defined group has to be described with a limited number of words, normally in the form of a group title, and it may be very difficult to give respondents the intended understanding of how their jobs relate to the different groups. The fact that the number of possible groups to choose from has to be severely limited will of course also limit the number of possible mistakes to be made. This is of little comfort, however, when many users of industry and occupation statistics need to make much more detailed distinctions and to work with much more homogeneous groups than those obtainable when using this approach.

544. However, the cost-savings are such that some national statistical agencies have been using this strategy in their census, thus foregoing the census' main advantage and *raison d'être*: the possibility of providing statistics for (relatively) small groups in a consistent manner for the whole country. The data quality of self coded occupation is so restrictive it would probably be better to augment the sample size of existing sample surveys as an alternative method for obtaining more detailed occupational data.

### *2.2. Coding by the enumerators*

545. Coding by enumerators can take two forms. First, an application of the above strategy, i.e., it is the enumerator who assigns the response to a pre-coded alternative during the interview, based on the information received from the respondent to the standard questions. Second, the enumerator writes down key words of the respondent's answer(s) and then codes the response after the interview, but before the questionnaire is forwarded to the processing centre.

546. The cost advantages of the first possibility are almost the same as for respondent coding. The main difference is that the interview is likely to take a little longer because of the need for the enumerator to understand and 'translate' the information received to the appropriate group. This procedure should be used if the respondents themselves are not expected to read the questions and write the answers. The consequence of this procedure for the validity and usefulness of the resulting statistics is as described above. The reliability of the coding may, however, improve relative to coding done by the respondent if the enumerator has been well trained and has received more detailed instructions on what types of jobs the different pre-coded groups are supposed to cover and where there are ambiguities which will require probing. On the other hand the enumerator may misunderstand what the respondent reports and therefore not select the correct group.

547. The advantage of the second approach, with the enumerator writing down the response that is to be coded later, over the use of recoded categories by respondents or enumerators, is the possibility this gives for much more detailed coding and thus much more valid and useful results. An enumerator coding outside the immediate interview situation may be given complete coding indexes to assist the coding process, as well as other coding aids - including the possibility to forward queries to supervisors. Another advantage is that, as the enumerators gain experience with coding, they will become more aware of the type of information that is required to code correctly and improve the recording of these characteristics with interviews. The main disadvantage is that because they are more numerous and geographically scattered, the enumerators cannot be given the same amount of training, supervision and support as specialized coders would be given, to assure coding reliability. Narrow casting of satellite transmissions has been successfully used to provide interactive training of census field staff in remote locations and should be considered as a solution to the training of geographically scattered enumerators.

548. A 'statistical' consideration with this approach is that all coders have irreducible idiosyncrasies that produce biases in the distribution of codes, which they allocate, relative to the mean of all coders. It can be shown that, if biases of similar magnitude are imparted by all coders, whether in the field or in the office, there will be less effect on the overall variance of the results if each of a large number of field workers codes a relatively small number of cases, than if each of a relatively small number of office coders code a large number of cases. Depending on the relative sizes of the work quotas of field and office workers, the total variance of estimates with field coding may still be lower even if field workers code in a more variable (less consistent) way than office coders. Note, however, that this reasoning is mainly valid for national aggregates. For smaller areas the 'biases' of individual field coders may not cancel each other out and it may be possible to distribute the workload between central coders to prevent clustering of their biases. The best is, however, to provide the coders with training and tools, which can minimize systematic errors and biases.

549. It may be possible to save time and reduce costs and operational complications by using field coding in circumstances where industry and occupation are the only items requiring office examination and coding and the field-coded questionnaires could be

passed straight on to computer data entry. On the other hand, field workers may be thought unsuitable in terms of background and training to act as coders. Also, field coding sacrifices to a large extent the important advantages of a controlled and supervised coding environment that can provide a direct feedback on coding quality and an incentive for correct coding. Such direct feedback can be provided to office coders in a census, through the quality control procedures, but will be much more difficult to provide to the enumerators, as field operations are likely to have been finished by the time serious problems are discovered, and the field supervisors will normally not be equipped to give quality control of coding results high priority.

550. The arguments in favour of using field coding, as compared to specialized coders detached from the data collection process, are closely linked to having a permanent field staff, which can be trained cost effectively in coding and which can accumulate experience. This means that although field coding may be a realistic option for continuous labour force and similar surveys, as well as for local administrative offices, the *ad hoc* and much larger scale operations like population censuses should normally include the coding of 'industry' and 'occupation' in the central processing operation.

### 2.3. *Office coding*

551. Specialized coders, located in a few processing centres, may be entirely specialized on the coding of one variable, or specialized on coding in general and/or they may carry out the coding as one part of an integrated data entry, coding and data control operation for the census. The exact context will depend on the whole organization of the processing operation, but the following concerns should be given serious consideration.

552. With a large coding operation, such as for the census, the coding staff will become thoroughly accustomed to the practical routines of the task, the rates of work required, and so on. Specialized coding units will tend to become strongly 'production minded' and cut off from the rest of the organization, with little incentive to assess their product in terms of external validation criteria. It is therefore necessary to establish rigorous measures of quality control in these operations. Also, individual coders tend to find 'short-cut' methods which reduce the laboriousness of their task but which may also incorporate errors or unjustified assumptions, if not actual violations of the coding instructions. These unofficial departures from, or additions to, the specified coding procedures tend to become institutionalized to the point where no distinction is perceived between them and coding rules derived logically from and designed to support the system of classification in use. Supervision and routine systematic audits of the coders must be employed to ensure compliance with the established coding rules and procedures. More generally, issues of external validity and internal consistency of coding tend to fall into abeyance, unless the routine procedures of the unit include specific and properly designed checks on the levels of coding validity and reliability achieved.

553. Coding may be one element in a larger processing task for each operator. This may make the coders tasks more interesting and can be a useful method of using employee resources efficiently. An additional argument for this solution is that fewer

persons will be handling the census forms and that simplifies the control task. (Note, however, that to simplify the discussion the larger processing context is mostly ignored in the rest of this chapter.)

554. The large-scale coding exercises mounted for national censuses of population generally rely on special recruitment and training of inexperienced coding staff. The performance of staff that is inexperienced may limit or delay the acquisition of the adequate coding habits described. Nevertheless, the coding results are likely to be poor unless recruitment, training and coding procedures are well planned, executed and supervised. A particular problem is that, because of resource limitations or practical difficulties at a time of heavy stress in the preparation for the census processing, there may be insufficient time to establish rules and routines and generally 'run in' the coding procedures before production coding has to start. If this happens the organization may initially be temporarily overwhelmed by the sheer volume of documents and data to be handled, with a consequent loss of control and severe fall in standards, which may prove difficult to re-capture and reverse later.

555. In general, existing coding procedures used in regularly conducted labour force surveys should be adapted for use in the Census. This avoids the need to 'reinvent the wheel' and helps ensure that coding is consistent among collections.

#### 2.4. *Level of coding*

556. The purpose of the coding process is to determine and record, from the responses obtained from the respondent, the code for the group in the classification to which the job of the respondent belongs. The coding can be seen as a translation process where the coder 'translates' the industry and occupation responses into the correct groups, because only in a minority of cases will the words of the responses be the same as those used to designate groups in the classifications. The raw materials for this process are the responses, and the tools the coders can use are the coding index, the coding instructions and the persons responsible for answering queries.

557. The coding process should therefore be designed to find and record the most detailed codes supported by the responses. The more information retained after the coding, the more valid and therefore valuable the resulting statistics can be for the users. A compromise option, often overlooked, is coding at different levels within different parts of the structure. However, traditionally, the most common procedure has been to decide that coding should be done at a particular level of the classification structure, e.g., the 3 digit level, no matter what information has been provided in the response. The arguments for this have commonly been that:

- (a) it would be too costly to code to a larger number of groups, both in terms of coding errors and in terms of staff hours required;
- (b) the responses would not support coding to more detailed groups; and
- (c) (when coding only a sample) it would not be possible to publish results for the more detailed groups because of lack of observations.

558. However, closer examination of these arguments in the light of the experience gained by statistical agencies has shown that:

- (a) The marginal costs of coding to a larger number of groups in the classification, i.e., to a lower level of aggregation, are rather small in terms of increased error rate as well as in terms of work hours needed for coding. The error rate for aggregate groups does not seem to increase. The most significant cost associated with coding to a lower level in the aggregation structure is the associated increase in the required sample size.
- (b) Experience clearly shows that the industry and occupation responses are very uneven in the level of detail they will provide. Many responses will support detailed coding, especially if the questions are formulated along the lines outlined in Chapter Three. At the same time a significant number of responses will, however, not support the level conventionally chosen. By insisting on a predefined level, the coding process may therefore both lead to unnecessary loss of information for a large part of the returns and to misrepresentation of the data quality for other parts.
- (c) The similarity criteria used to define the groups in an industry or occupation classification are mostly defined with reference to the nature of the production process and the type of work performed, without much regard for the number of employed persons in the resulting groups. This means that the number of jobs that can be found in groups defined at the same level in the classification will differ greatly, except for the few cases where 'statistical balance' has been adopted as an important consideration when constructing the classification structure. The number of jobs in a group defined at an aggregate level may therefore be smaller than that of a group defined at a lower level in the structure of the classification, but within another high-level group. In addition, the tabulation of industry and occupation statistics also typically involves both the merging of groups and cross-classification with other variables such as age, sex, and region. Even in cases where the data cannot be published, the micro data can be extremely useful for internal analysis. Consequently one should not restrict tabulation possibilities during the coding process.

#### 2.5. *Coding of vague and difficult responses*

559. Most industry and occupation classifications specify residual groups of "type x industry/occupations not elsewhere classified (nec)". These "nec" groups are designed to take care of activities and jobs that belong to the more aggregate group but that are not similar enough to the specified sub-groups within it to belong in any of them, and which in themselves include too few cases to warrant separately specified groups. Responses should only be coded to an "nec" group when they match index items of the "nec" group. The "nec" groups should not be used to code those responses which the coders cannot assign to any of the specified groups. Such responses can be either:

- (a) too vague and imprecise to allow the coder to determine to which group the job belongs;
- (b) indicate that the establishment (or job) in question produces a combination of goods or services (or the job involves tasks and duties) which cut across the distinctions made in the industry (occupation) classification; or
- (c) represent a type of production or work not covered by the classification.

560. The proper way of handling such responses will depend on the type of case:

- (a) Vague and imprecise responses should be coded to the level in the aggregation structure supported by the information contained in them - they should not be forced into any particular detailed group where it is likely that only a small proportion of these jobs would fall if one had an adequate response. For example, in the Australian 1986 Census 15 percent of the jobs coded to the major group 'clerks' could not be coded to any of the more detailed groups within this major group. It would obviously represent a significant distortion of the results if they had all been placed in one particular more detailed group together with those jobs that properly belonged to that group. A common method of dealing with this type of response is to provide entries in the coding index for commonly occurring vague responses. These entries are assigned the code for the relevant higher group, followed by trailing zeroes. The Canadian Census coded to the level supported by the responses and then allocated these responses proportionally to the more detailed groups in a transparent manner.
- (b) The classification of establishments with an uncommon mix of activities or jobs with an uncommon mixture of tasks or duties should, as far as possible, be made on the basis of the general priority rules of the classification. Such responses should preferably be left to expert coders or raised as queries for the classification experts. Disruptions of the coding process can be minimised if these responses are given a special code and the questionnaires put aside for later examination by the experts. This treatment should also be given to responses that seem to represent establishments with activities or functions not covered by the industry classification and jobs with tasks or duties not covered by the occupational classification. The reporting of difficult cases of these types is an important input to the process of updating, maintaining and possibly expanding and revising the relevant classification and coding index.

561. The information to be used as a basis for the coding of industry or occupation is first and foremost that contained in the responses to the respective industry and occupation questions. However, in a significant number of cases where this information is not sufficient for the coder to choose between possible alternatives, ancillary information provided by the responses given to other questions may provide the basis for making the choice. (For 'occupation' the most important type of such information is often that pertaining to the 'industry' of the workplace.) However, it must be emphasized that



the use of such ancillary information should be specific and strictly controlled, in order to avoid an undermining of the descriptive and analytical use of these variables, in particular when they are being used together or for cross-classification. This means that rules for proper use of such ancillary information, i.e., when and how to use it, must be incorporated into the coding indexes and the coding instructions.

## **C Planning and organizing coding operations**

562. As earlier recommended, the discussion which follows is based on the assumption that coding is to be carried out by specialist coding staff in the context of the final processing of the census. This is because this is the solution most commonly chosen for census operations by national statistical agencies, and because it is, in most cases, easy to deduce from these the guidelines that should apply to field staff coding, or to other decentralized arrangements.

563. Assembling the right resources for processing of the industry and occupation information in the right places at the right time and managing those resources efficiently are fairly complicated tasks. They require anticipation and co-operation between different parts of the census organization and must be coordinated or integrated with the other processing tasks. Large volumes of documents and data have to be handled and, because of the interdependence of different stages within the overall processing plan, the penalties for failures of operational or quality controls, in terms of delays and cost increase, can be heavy. The main outlines of the processing plan for a population census, particularly as they affect requirements for finance, staff, equipment and premises, may need to be worked out long before the start of the actual processing. Managerial staff who will be involved in the planning and supervision of the coding operation and professional staff responsible for designing the classification and coding procedures, the training of coders, the updating of classifications and coding indexes, and interpretation of results need to collaborate closely at the planning stage.

### **1. Finance and resources**

564. Substantial amounts of money are required to support the processing of a census and these will have to be estimated and provided for under appropriate budgets. Estimates for each part of the processing task often need to be made several years in advance and to be fed into the financial planning and procurement procedures of the responsible agency to ensure adequate provision. This demands early decisions about resource requirements and these may in turn precipitate strategic processing decisions which have resource implications (e.g., staff numbers and pay rates, number of processing offices, and whether to use computer-assisted techniques). It is important to ensure that financial, resource and operational planning are coordinated, so that the technical assessment of requirements determines bids for resources, rather than vice versa.

## 2. *Expertise, experience and rehearsal*

565. The coding of industry and occupation demands special expertise in those who plan, manage and supervise the operation. Processing of each census relies heavily on technical information, expertise and experience gathered at the last similar exercise. Information and experience from that exercise may be documented in detail, but practical expertise is likely to reside very largely in the heads of a small number of experienced staff. Staffing continuity in key positions is therefore extremely desirable. However, circumstances change and it is not possible to rely entirely on documented or undocumented experience from the last census. The staff working with the Labour Force Survey or other, related surveys where information about industry and occupation is collected on a regular basis will often have very relevant experience and tools, and they should be consulted. If and when outside advisers are used it is also extremely important to verify the validity of their experience for the local context and to have the opportunity to modify their estimates of costs and time requirements, based on concrete experience. A processing rehearsal is very important as an aid to planning and estimation, and the results of such a rehearsal need to be available early and at a stage when it is still possible to adjust plans for the main operation in the light of them.

## 3. *Estimating coding rates*

566. Some key planning parameters, e.g., coder work rates and effective coding throughput rates, can only be estimated reliably from well-documented previous experience or well-planned tests. Certain problems arising from the scale of the full operation, e.g., its effect on problems of recruiting, maintaining and controlling staff, may be hard to test in advance. Experience also shows that performance rates vary significantly over time during the processing period. Coding rates are much lower, and query rates much higher early in the process than later. There is, also, the danger that relaxing controls can cause a fall in standards towards the end of the process, beyond what is warranted by the improvements in the coding operation, resulting from the early improvements in the coding tools used and on-the-job learning by the coders, supervisors and classification experts.

## 4. *Coding staff*

567. It is important to make good estimates both of the number of coders required and of the numbers of first-line supervisors needed to control the coding process, as well as of the number of specially trained staff needed to resolve queries.

568. In a census operation the large volume of work to be processed within a limited period will require special staff recruitment and training, both for initial recruitment and for anticipated staff turnover in the course of the task. Thorough enquiries and consultations should be undertaken about likely sources of suitable staff recruits, since financial constraints are likely to prevent actual recruitment until the last moment. There may be external pressures to employ particular groups of persons, even when their

suitability cannot be guaranteed. Common criteria should be defined and applied in the selection of all staff. The terms, on which staff are to be employed, including the minimum acceptable level of education, pay rates, grading, disciplinary and hiring/firing rules, need to be carefully defined. It is important to provide adequate time and resources for staff training at both the coder and supervisor levels, and to recognize that the specialists, who are to advise the supervisors and to resolve the more difficult queries, normally cannot, and should not, be recruited and trained for a particular census operation, but be part of the permanent competence of the organization.

569. Persons with the following characteristics best perform the tasks of the coder.

- (a) Literate and reasonably intelligent, but not over-eager to display independence of judgement as this may lead them to find the task demeaning or frustrating.
- (b) Willing and able to speak up when a problem is detected.
- (c) Clerically accurate and careful.
- (d) Willing and able to follow detailed instructions conscientiously, without attempting to alter or improve upon them on their own accord and without consultation and clearance with supervisors, but prepared to raise queries in cases of genuine doubt.
- (e) Honest and trustworthy and thus not likely to falsify or omit procedures in order to reduce the amount of work to be done per case, or for other reasons.
- (f) Persistent and willing to work steadily for long periods.
- (g) Able to work reasonably rapidly and to maintain a steady level of productivity.

570. Those responsible for recruiting and selecting coding staff should have these characteristics in mind. Several of them, e.g., (b) (c) and (d), are best assessed through an objective-screening test (which may also be applicable to other types of routine clerical task). Mistakes at the initial recruiting stage are likely to lead to high staff turnover, and the need to recruit and train replacements while production coding is in progress.

571. Proper training of coders is very important: Coding of industry and occupation is best learned through practical instruction in specific procedures (e.g., document handling routines, use of the coding indexes, etc.), interspersed with supervised practice on appropriate, specially designed exercises. Relative slowness in learning need not be a disadvantage if accompanied by good retention of what is learned and the desirable temperamental characteristics. The training period can also be used to identify persons who are unwilling or unable to follow the coding instructions precisely. It is important to remember that staff recruited to replace coding staff leaving before the end of the coding operation will also need to be trained.

##### **5. *Coding teams and supervisors***

572. Production coding on large jobs is best organized by allocating coders to teams,

each under a first line supervisor. The supervisor's role and work tasks need to be carefully specified and are likely to include: controlling work flows; monitoring and maintaining work rates; enforcing work discipline; resolving and recording coding queries, and applying quality control procedures, etc. First line supervisors need to be trustworthy persons with the necessary intelligence and force of personality to master their duties and to control and motivate coders. The need for them to have previous experience in coding operations depends on their role in query resolution. In principle one may organize the coding operation in a way that gives the operational supervisors a very limited role in query resolution. Then it is not essential or necessarily desirable that they should have had prior experience of industry and occupation coding. However, in most cases it will be preferable to give supervisors the responsibility for first line query resolution because of their close contact with the coders and the shorter response time and higher capacity relative to classification experts. Supervisors with responsibility for query resolution should be given a good understanding of and training in the classifications and coding systems.

573. The number of coders allocated to each supervisor is important. Typical ratios lie between 6 to 1 and 12 to 1, but the appropriate ratio needs to be assessed in each case, taking account of the flow of work with which supervisors will be required to cope. Overloading of supervisors is likely to cause not only bottlenecks and poor staff morale, but also under-reporting of problems and queries and a reduction in the reliability of coding. Because coding is a fairly monotonous task it is important to ensure that work morale and discipline is maintained and that productivity rates do not fall. Particular problems may arise where coders expect that it will be difficult to find a new job after the end of the coding operation and who therefore may want to make the work last as long as possible. Special bonuses may be effective in counteracting such tendencies. It may also be possible to retain to the end of the process only staff who have a (possibility of) more permanent and long term employment with the statistical service. Their experience will be valuable both for other surveys and for the documentation and explanation of the census procedures to the users of census results as well as to those who in time will prepare for the next census.

## **6. *Coding tools***

574. It will be necessary to provide appropriate documentation, procedures and training materials not only to coders, to guide the coding process itself, but also to supervisors. The basic tools required by coders will include:

- (a) Coding instructions: These should cover all operations that the coder is required to carry out. The procedures and instructions for handling all relevant items and operations should be integrated. The instructions relating to the coding of industry and occupation will need to be particularly clear and specific on: (i) the order in which checking, coding and editing tasks are to be carried out; (ii) the procedure for analysing verbatim material for significant terms; (iii) the use of the coding index or other coding documents; (iv) the circumstances and procedure for using ancillary data; and (v) when to refer a 'difficult to code' response to a supervisor for query resolution;
- (b) Coding index: This is the key coding document through which verbatim terms incorporated in job titles, descriptions of tasks, etc., are translated into codes. Coders should not be encouraged to interpret verbatim responses in terms of their own conception of the purpose or criteria of classification, but rather to follow in a conscientious way the instructions laid down for consulting the index. For these reasons it is essential that the index be clearly set out, explicit and easy for coders to use. The use of the coding index, instructions and procedures needs to allow for updating in the light of decisions made in resolving queries and problems which arise and are dealt with in the course of coding.
- (c) Query resolution procedures: A query occurs when coding clerks cannot assign a code using the specified procedures and indexes. There should be clear instructions on when and how the coders should raise queries, how to record and report them, and their resolution. Queries are the most useful inputs to both immediate and future work to up-date the coding index and the classification itself, and it may prove necessary to carry out such updating frequently early in the census processing, if the coding indexes, or even the classifications, prove to be out of date or inadequate for other reasons. Updating of coding indexes should be done by, or in close collaboration with, those responsible for the standard classifications so that consistency between collections can be maintained and any problems with the classification itself can be identified.
- (d) Legal and administrative forms: A legally binding undertaking to maintain the confidentiality of census data should be signed by the coders. Other documents used by both coders and supervisors are likely to include forms for recording queries and their resolution; for controlling the flow of work and reporting progress; for quality monitoring, etc. To ensure that target throughput rates are achieved, the productivity of coders and coding teams needs to be monitored and progress charts maintained. Special measures for motivating coders should be used, for example, the posting of productivity and error rates for coding teams (but not for individual coders, except perhaps the best ones).

## **7. *Coding problems and queries***

575. A coding query occurs when a coder is unable to assign a classification code to a census or survey response using the standard coding procedures and tools. No matter how carefully coding instructions and the coding indexes have been prepared, it can be guaranteed that large numbers of detailed queries will be raised in the course of a major coding operation. This happens if the indexes prove to be out of date or incomplete in some respects. Another reason is that actual responses will be more varied than anticipated by the index constructors, even with the most carefully analysed pre-census tests and well-designed questions and instructions to enumerators. Any revision of the structure of the classifications since the last census or survey may also lead to a new crop of problems in the treatment of vague and inadequate responses at the borderlines between categories.

576. Evidence of shortcomings of the documentation thrown up in the course of production coding, will need to be rapidly and consistently processed and fed back to the coders and their supervisors in the form of amendments to their tools. Appropriate procedures for reporting and recording queries need to be laid down in advance, and decisions made in resolving them and for incorporating any consequent amendments in the coding documentation and procedures. The roles of supervisors in processing queries and amendments need to be defined, e.g., how and when they should communicate with the coding experts, and how new versions of the tools should be distributed to the coders. Particular care is needed in the co-ordination of query reporting and document amendment where coding is being carried out in several different locations, if for example different provincial or local offices have responsibility for processing the census returns. All coding sites and teams need to be informed and given updated tools as quickly as possible.

## **8. *Quality assessment and quality control***

577. Casual observation by supervisors and ad hoc checking of coded output do not provide adequate information on quality of output. Explicit allowances for the resource and time costs of formal quality control need to be built into the processing plan. These will cover the establishment and staffing of a quality control unit responsible for acceptance testing of coding during the start up of the coding operation, and for assessment of the reliability and consistency of the operation as a whole. A procedure for (sub-) sampling the work of the coders for quality control purposes needs to be defined and the quality control unit needs to be staffed at a level which will enable it to keep pace with the main coding operation. Coding schedules must make allowance for corrective action (e.g., 100 per cent checking) in the case of batches that fail a quality control test.

578. In-built quality control procedures will also be required. It is necessary to design separate procedures to handle: (i) on-line acceptance testing of coders' work; and (ii) overall monitoring and assessment of performance.

579. The aim of acceptance testing is to identify rapidly coders whose performance does not meet criterion levels of accuracy in following coding instructions, so that corrective measures can be taken. The aim of overall monitoring is to estimate average levels of coding accuracy and inter-coder consistency for the entire coding exercise. Note that estimates of coding reliability need to be supplemented by estimates of the validity of coding if a balanced overall assessment of the quality of the statistical output is to be made. Estimates of validity may be obtained from a post enumeration study in which the whole data collection, coding and editing process is repeated for a sample of census cases. On the basis of such quality assessments it may be possible to separate the contributions made to total variance by source of variance, i.e., due either to data collection or to coding.

#### **9. *Premises, infrastructure and equipment***

580. The large volume of work and of questionnaires entailed in census processing requires suitable office space and the entire necessary infrastructure for properly supervised clerical operations, as well as for easy movement, storage and retrieval of forms. A special requirement is for security of documents bearing personal details. Proper attention must also be paid to the functionality and capacity of the desks, chairs, shelves, filing cabinets, lighting, heating and ventilation as well as the adequacy of paper, pencils and other stationery. Neglect of such factors can easily influence the morale of the staff and result in higher than anticipated staff turnover and inadequate attention to work quality and speed. Suitable premises need to be specified, identified, costed, approved and booked well in advance. If coding staffs are to operate with special equipment, such as computer terminals or optical readers, special arrangements may be needed to estimate the requirement, identify suitable and reliable equipment and carry out tests, estimate and provide for capital expenditure and depreciation, provide for replacement in case of breakdowns, go through procurement procedures, etc.

#### **10. *Process in one location or in several***

581. Census processing creates a substantial but temporary demand for suitable staff and premises and the extra staff recruitment and other management problems may be greater if all processing is carried out in a single location. There may also be other cost and logistic arguments for carrying out processing at one or several locations other than, or in addition to, the central census processing office. On the other hand, it needs to be borne in mind that for relatively complex tasks such as census coding and processing, it is both difficult and important to maintain consistency in coding between coders and coding teams. One reason for this is that census coding inevitably generates large numbers of queries. (In the 1981 Census of the United Kingdom, more than 30,000 coding queries were processed and it is thought that many more were dealt with informally). These in turn lead in some cases to amplifications or changes to coding indexes and instructions, which then need to be applied in a consistent fashion (This is much easier if all documents are in electronic format). This and the maintenance of consistent production

and quality control standards more generally, are more difficult to achieve when coding is carried out in several locations than if they are centralized to one place. It will be important to establish possibilities for communication by telephone, fax and/or electronic mail, especially if the operations are not centralized.

### **11. *Handling of documents***

582. The coding of occupation and industry will normally be an integrated part of the total processing of the information on the census questionnaires. In that case the main concern will be, (i) how to receive the forms, (ii) how to store them and (iii) how to allocate them to staff so that one can control and ascertain that forms have been processed. It should be easy to find individual forms that for some reason need to be rechecked. If each form has to be handled by more than one person, for example because the coding of different variables are carried out by different persons or because data entry is done by special operators, then the flow of documents must be planned to avoid bottlenecks and loss of forms. All movements of questionnaires from one location, i.e., workstation or office, to another and to storage, should be carefully recorded.

583. Capturing the responses by OCR to create electronic records is one option for documents which has the potential to significantly reduce processing costs and problems caused by flow of forms from one process to another. A less expensive and technically simpler option is to scan the questionnaires and to work with image files of the documents. When considering the cost of either OCR or scanning it should be noted that an archive is created as part of the process.

### **12. *Use of automatic or computer assisted coding***

584. Until quite recently most census operations reserved computer usage for “downstream” applications. That is, after data had been collected, manually transposed to data collection sheets and key punched on to cards or paper tape or entered on magnetic tapes or disks, the machine-readable data would be fed into a computer (usually a large main-frame computer) and verified through a series of computer edit-checks. This process would involve passing the data through a computer program to ensure that any out-of-range codes or non-allowable combinations of codes were detected and reported for further investigation. These types of controls are still valid and important, but computer edit-checking cannot detect invalid coding within the range of valid codes if errors in data collection and coding do not generate non-allowable combinations of data. Apart from edit-checking, data processing by computers was usually reserved for the final stage of the whole operation, the production of statistical tables.

585. More recently, data processing has benefited from the development of database management systems as well as the development of a wide range of applications for the vastly enhanced capacity of “small” personal computers, ‘standing alone’ or linked through internal networks. For large-scale data processing applications there are both main-frame and PC versions of these systems which allow the non-specialist to design



screens and forms and to develop a system for the processing of data at all stages from data entry, through edit-checking, management of the processing facilities to the presentation of statistics for publication. Previously, the manipulations of census information or large-scale social survey information via such systems were prohibitively expensive in terms of programming and the hardware requirement for on-line data storage. Technical advances have now made large-scale database management a feasible proposition for such applications.

586. It is also only relatively recently that real progress has been made in bringing computers to bear on the tasks carried out, in an inherently slow and laborious way, by census coders. However, the situation is changing rapidly and systems for automatic coding (AC) or computer-assisted coding (CAC) of industry and occupations are now used in a number of countries. Automated coding and Computer Assisted Coding improve the consistency of coding and reduce the time and cost of coding. Both types of system read captured responses. The responses can be captured by OCR or manually inputted. Manual inputting requires operators with appropriate keyboard skills that may not be available in some countries. The success of OCR varies by language, alphabets or characters. Successful OCR operations may require the use of special quality paper and ink, as well as handling procedures for the questionnaires before, during and after the data collection that will protect them from humidity, the sun and other spoiling influences. This may prove difficult to ensure in many countries.

587. So far no workable system has been developed that can fully automate the human decision-making task in coding occupation and industry. Some AC systems allocate codes automatically to more than 70 per cent of cases, but the development costs have often been high and there have been problems in making the systems sufficiently 'intelligent' to simulate reliably the performance of trained human coders. The reported error rates for those responses coded by many of these systems are of the same order of magnitude as those of human coders. Moreover, the residual need for human intervention in the coding process in a substantial proportion of cases reduces the benefits that such automation could have on simplifying, speeding up and reducing the cost of data processing. Both manual and automated systems will have difficulty with vague and contradictory responses. These limitations do not, however, eliminate the gains that can be achieved using an AC system, in particular if this is integrated into a data entry and processing system, which starts with optical reading of the questionnaires.

588. If it is intended to introduce an AC or CAC system, trials of the hardware and software and of the machine/operator interface need to be conducted well in advance. If there are other surveys (perhaps a labour force survey) that collect occupation and industry variables in a similar manner it would be advantageous to develop and test an AC or CAC system with those surveys. Until the feasibility and operational robustness of the machine-based system have been established, it is prudent to make parallel plans for reversion to a manual/clerical system as a fallback position.

589. Computer assisted coding (CAC) uses the computer to assist the coder to find and

assign the correct code. Typically the coder inputs a keyword or a portion of the keyword. The structured index items appropriate to the inputted keyword appears on the screen with the appropriate coding instructions. The process uses few keystrokes as typically the coder can scroll through presented options on the screen. The computer assisted systems maintain the advantages associated with manual coding but at the same time can enhance the quality consistency and speed of the coding.

590. Although computer systems have the potential to yield dramatic improvements in the quality of data, to control the environment within which data are collected and to reduce the amount of time which elapses between data collection and data dissemination, due regard must be given to the true costs of implementing the system. The costs must:

- (a) include a realistic estimate of the rate of depreciation of hardware although these may frequently be put to good use for statistical surveys after the completion of the census processing; and
- (b) take account of the dependence on specialist programming and systems analysis skills for the implementation of the required software.

591. To minimize the risks of development, and probably the cost, one should seek to acquire the rights to operate a system already tried and tested. In selecting a system, ease of operation and adaptation to national circumstances (e.g., the language used by the operators, the classification and the coding index) should be given priority. The operational advantage of an AC or CAC system is likely to depend more on the type and cost of data registration and the suitability of the coding index than on any particular feature of the search and decision making algorithms of the system. However, fast response times and an easy-to-understand interface between the computer and the operator should also be important considerations when selecting a CAC system.

### **13. *The problem of different languages***

592. In the discussion above no reference has been made to the problems encountered in countries where there is more than one language that the population uses in their daily life, although this issue was discussed under multi-language questionnaires in Chapter Two, (see para. 65). On the assumption that enumerators will know the language of the respondent and therefore can write down correctly in that language the answers to the industry and occupation questions, the best solution will be to make sure that the coding indexes can reflect those answers:

- (a) Separate coding indexes for each major language are the best solution where this is operationally feasible.

- (b) Otherwise it may be possible to create “multi-language” coding indexes, to allow the coder, whether this is the enumerator or a specialized coder, to find an index entry that corresponds to what was written down as the response of the respondent. A multi-language coding index will be larger than a single language one, but not necessarily dramatically larger, because in many countries the terminology reflecting modern sector activities and jobs will be common to many languages, and it will mainly be the terminology reflecting traditional activities and jobs which will differ. Such activities and jobs are normally less varied than those in the modern sectors, and there may therefore in each language be fewer terms describing them.

593. If the coding index can only be constructed in a single language, then this will require that someone, usually the enumerator, translate from the response to the language of the index. The problem with this is that the correct translation of occupational terms will not only require good general knowledge of the two languages in question, but also knowledge of the particular area of work in order to understand precisely how particular terms for activities, products, services and jobs are used in the local context. Very few persons will normally be able to satisfy this requirement over the whole range of work situations covered by a population census or labour force survey.

## **Chapter 11. The Development And Use Of Coding Indexes**

594. This Chapter is based mainly on experience from English speaking industrialized countries, because the limited documentation readily available on the development and use of coding indexes and coding procedures have mainly originated in such countries. It is difficult to assess the extent to which the documented experience is transferable to other languages and cultures. While this should be kept in mind when reading the following text, the experience from English-speaking countries may provide a starting point for work and experiments in other languages.

### **A What is a coding index?**

595. The process of coding industry and occupation information involves the task of matching responses to questions in a census or survey with index entries, to find the appropriate classification codes. The coding index is the key instrument for this matching process. The index can take the physical form of a durable printed publication, a loose-leaf binder, a computer print-out or a machine-readable file within a computer system. The matching can be carried out by a person, i.e., the coder, by a computer, or through interaction between the coder and a computer.

596. Detailed industry and occupation coding has traditionally been carried out using clerical (i.e. manual) procedures. Relevant information is recorded verbatim in the field by the respondent or by enumerators and the resulting raw data are brought to one or more central offices. Here clerical staff scrutinize each case, decide to which industry or occupation group to allocate it and record an appropriate code on to a document, or directly on to a computer-readable medium, for further processing. It should be noted however that AC and CAC systems are rapidly emerging. Users of manual coding should consider the eventual adoption of such systems in the long term and begin preparing for these systems far in advance of their implementation.

597. The coding index is the principal instrument for linking the words used in the various parts of the response to the numerical code that represents the allocation of that response to the corresponding group of the classification. The coding index guides the coder by listing information, e.g., key words, which can be found in the responses, and indicate how different responses are allocated to the detailed or more aggregate groups of the classification, depending on the nature of the information in the response and on the instructions for the coding process. Thus the census coding index must be a reflection of the type of responses which one will find written on the census form by the respondent, or by the enumerator on the basis of information received from the respondent. It should reflect the type of words and expressions that the respondents will use when asked to give the information about their place of work and their job in the census enumeration.

598. It is important to recognize that a coding index is different from the index of groups specified in the classification, which is just a list of the titles given to those groups that are separately defined in the classification. The titles chosen for these groups are

designed to be as descriptive as possible for the group content, given that only a few words may be used. Only a few of these titles will correspond to the terms used by individuals when asked about the activities of their place of work or about tasks and duties of their job. The coding index is also different from a list of titles or terms that may have been chosen to be descriptive of the content of the groups, i.e., to illustrate their content. This list may also contain entries that may never be used as a title by any person describing their job or place of work. When they exist, the latter type of list may, however, serve as a useful starting point for the construction of a coding index.

599. Since it has to be in place before the census coding operations start, the coding index has to be constructed in anticipation of what the census responses will look like. The basis therefore will have to be actual responses to similar questions in the last census and in household surveys carried out after that time, and in census pre-tests. Terms and expressions concerning types of economic activities and jobs which people have and which can be found in advertisements for products and services (for the industry index) and job vacancies (for the occupation index), and registrations of vacancies and job seekers in employment offices.

600. The collection and coding of the elements to be included in the coding indexes should be carried out by experts on the respective classifications to ensure that they are done correctly. The work will be painstaking and time consuming, but the investment involved in the collection and coding of up to 5, 10, 20 or even 30 thousand index entries ahead of the census will prove well worth the effort in terms of the speed and reliability with which hundreds of thousands or millions of census forms, (depending on the size of the work force, the complexity of the economy and the coding strategy chosen) can be coded during census operations.<sup>91</sup>

601. At the start of the census operations the coding index must be considered as incomplete, and provisions must be made to update it during the whole period of the census coding operation, but probably much more frequently and with more new items early in the process than later. The updating process should be an extension of the query resolution process in the sense that the nature and outcome of resolved queries should be made available to all coders as soon as possible, in case they encounter the same situation. The best way to achieve this is to issue a new version of the coding index. It is better to issue a complete, new version of the coding index than to issuing additions to an existing index, because the new entries will belong in different places in that index and the coders will have to transfer the information from the note on additions into the main

---

<sup>91</sup> It is in many ways correct to regard the coding index as the ultimate manifestation or embodiment of the classification. However, it should be kept as a working tool and not be given the status of being an official part of the classification, for two reasons: (i) It will be necessary to update the coding index during the census operations, and when it is used later in other coding operations. This flexibility may be difficult to achieve if it is a formal part of the classification. (ii) In order to reflect actual responses it may need to include terms, e.g., brand names, which are commonly used to describe places of work or types of jobs, but which may be trademarks protected by copyright or be difficult to include for other reasons in an official publication. Such problems will usually not arise when the coding index is used only as an internal working document, and made available to other organizations only for use as such.

coding index, with the danger that they will make mistakes. New, complete versions of the coding indexes issued frequently in the first weeks of the coding operation will also reduce the danger that individual coders will keep their own notes on the coding of particular responses. Such notes can easily be the source for systematic differences between coders in the coding of responses not reflected in the initial version of the coding index.

602. During the census operations, the physical form of the coding indexes must reflect the temporary nature of the version currently in use at any point in time. Each issue should be precisely dated and, when choosing the form of the paper versions, the main consideration should be the speed of reproduction. Both paper and electronic versions should be issued with orders to destroy earlier versions<sup>92</sup>. If using paper, a binder is very adaptable if space is initially left on each page for several additional entries. It is then possible to reprint and replace one page without requiring changes to other pages. If using an electronic index it is best to keep the index on a server. Changes to an index on a server are immediately reflected at all workstations.

603. Only when it is clear that further significant additions to the coding index will not be forthcoming during the remainder of the census coding operation should it be issued in a form more suited for dissemination to other users, such as various government departments, survey agencies and academic users. Then a well-bound quality publication may be considered, but a format more compatible with its status as a working tool, and with regular, although less frequent, updating may be preferred, e.g., a ring-file or computer print-out may be the appropriate format.

604. Up to this point the presentation of the coding operation and coding tools has been relevant to both industry and occupation coding, as the issues are the same for the two variables, and when both are present in the census these issues and the preparations should be addressed jointly. However, as the information sought for the two variables is different the responses obtained will also be different, and therefore the content and construction of the two coding indexes will be different. They will therefore be presented separately in the remainder of this chapter. The next two sections concern the occupation coding index and its use, while the two final remaining sections concern the industry coding index.

## **B Developing and updating the occupation coding index**

### ***1. Sources of information for construction of the index***

605. The process of updating a coding index for occupations should be one part of the general process required to maintain the national standard classification of occupations. As new ways of organizing work between or within enterprises or new technologies are

---

<sup>92</sup> Copies of each version of the index need to be retained by management and by the team responsible for the classification, however, to assist in analysis of data quality and to allow for recovery in the event that an error introduced to the index may result in systematic errors in the coded data.

introduced, new jobs will appear with new combinations of tasks or new types of tasks associated with them. These new jobs may be given new job titles by their incumbents or their employers, or may be referred to under existing job titles. At the same time existing jobs may be given a new title without their tasks and duties being changed in any significant manner, e.g., as a result of for example, reorganization of the enterprise or because their placement in a wage hierarchy has been changed. Thus, there is a need to keep track of job titles and the associated job descriptions, monitoring the relationship between this information, index entries and the associated occupational codes. However, those responsible for census preparations will frequently find that this work has not been undertaken by the custodian of the national standard classification of occupations, nor by anyone else, and that in fact not only the updating but also the creation of the occupation coding index must be undertaken from scratch for the census.

606. A full scale job content monitoring exercise cannot be used for the census preparations as that would be too time-consuming and costly. The most realistic alternatives would be to carry out (i) post-coding reviews of recent survey operations; (ii) job vacancy reviews; (iii) consultations with job placement services; and (iv) a job monitoring exercise for the industries known to be either restructuring or undergoing technological changes.

607. The first procedure makes use of the tools and members of the coding team(s) used for recent survey(s). Coders are a good source of information on the adequacy of the index that they have used. Their suggestions for improving the index and for additional or revised entries must be recorded and investigated. Ideally preparations for the use of 'post-coding review procedures' for the development of index material should have been designed into the data processing operation for the respective surveys, as it is essential that information which may be useful in updating the classification and index be carefully collated and retained. Such information typically consists of records of problems encountered, queries raised and decisions and amendments to the working instructions adopted in the course of coding. A sample of the actual coded responses captured in machine readable format can be very useful in classification development, index development, development of AC or CAC systems and coder training. Such procedures should be made part of the normal routine for any continuous or regular surveys, e.g., Labour Force Surveys, as well as for the registration of occupations, which takes place in the local offices of the employment services.

608. When available from job advertisements in newspapers, journals, bulletin boards and/or registrations at a local office of the employment services, job vacancy descriptions may provide a useful source for constructing or up-dating information on job titles and detailed job descriptions, and therefore also a coding index, especially where these have been coded to the occupational classification as part of the job vacancy recording process in an employment service. From such announcements of job vacancies it should be possible to find a reflection of the impact of technical and organizational change on the allocation of tasks to jobs, and to develop proposals for new entries for the index (and the classification). The advantage of this approach is that it does not require expensive initial

search for contacts, as follow-up inquiries to a vacancy of interest can use the name, address and contact person of the employer found in the vacancy notice. The main disadvantage is that job vacancies that have been advertised in newspapers or on bulletin boards, or notified to and recorded at an employment agency, normally only cover job vacancies for a limited range of occupations and industries.

609. In cases where no national index of occupational titles exists, the index developed for ISCO may be a useful *starting point* for the development of a national index. It should be stressed, however, that even for English speaking countries using a national classification based on ISCO, the ISCO English index would need to be modified for use in national circumstances, as the words in the index need to reflect national usage and there is great variation between countries. Direct translation of the English, French or Spanish ISCO indexes into other languages will not yield good results. These indexes are useful sources of information concerning the content of categories in ISCO and on the correct placement of some internationally used occupation titles.

610. In some countries the employment agencies have established standard procedures for the collection of job vacancy materials. For example, when employers contact an employment agency, the agency may create a computerized record of information containing the job title and a brief description of the main duties or tasks associated with that job title. These records may be used within a word-processing system or may form part of a database of vacancy information. In some cases they may have been linked, or coded, by the national employment service to the national standard classification of occupations.

## **2. *Organization and structure of the index***

611. When organizing the material for the coding index of occupations the first issue concerns the structure of the index itself. Basically the choice is between two different approaches, which can be described as *all inclusive* and *structured*.

612. When the *all-inclusive* approach is adopted, every distinct type of response found in the process of coding should, in theory, have an entry in the index, although allowance may be made for misspellings and inversions of words which are without consequence for the meaning of the response. An advantage of this approach is that it may be possible for coders, when faced with an obscure job title and/or task descriptions, to find that title and/or those tasks listed in the index. A significant disadvantage is that an all inclusive index may become very large and its sheer size may slow down the process of searching for the "right" entry in the index, and thereby slow down the coding, whether the coding is done manually or is computer-assisted.

613. Also, large verbatim indexes may create the impression that coding is a simple task, involving a straightforward matching between a response and an index entry. However, no matter how large the index (and some occupation indexes have over 30,000 entries) it will always be the case that a significant proportion of responses fail to match the index entries exactly. In these cases the coder had to be given rules and/or use



judgement to make the 'best' match. For these reasons, in a number of statistical agencies, the approach has been to develop a structured index.

614. A *structured index* does not try to reflect every possible response because it is accompanied with instructions to the coder on how to break down the available response into key words and qualifying nouns or adjectives. The primary entries in the index are the key words. If a key word in itself is not sufficient to uniquely identify the group, an appropriate qualifying word (or phrase) must be added to distinguish between the possible alternatives having the same key word. If this is not sufficient to resolve all ambiguities, a second or higher order qualifying word should be used. The following examples may illustrate the system for transforming an occupational response into an entry in a structured coding index according to the following format:

Response: ⇒key word/1st qualifying word/2nd qualifying word

Cost accountant:⇒ accountant/ cost

Drilling machine operator:⇒ operator/ machine/ drilling

Aircraft instrument maker:⇒ maker/ instrument/ aircraft

Room maid:⇒ maid/ room

Marine biologist:⇒ biologist/ marine

Capstan lathe setter-operator:⇒ setter-operator/ capstan lathe

615. The following examples, based on the coding index used for coding occupation in the 1986 Population Census in Australia, will illustrate the use of the qualifying words as well as the way instructions about the use of the index can be incorporated with the index entries (the codes given are those of the Australian Standard Classification of Occupations, first edition - ASCO):

5999 Researcher/market/interviewing

2909 Researcher/market/statistician

2907 Researcher/market (except above)

2701 Researcher/accountancy

2107 Researcher/agricultural

2907 Researcher/anthropology

2999 Researcher/assistant to parliamentarian

2107 Researcher/biological sciences (except medical)

2101 Researcher/chemistry (except medical)

2109 Researcher/medical

3103 Researcher/toxicology

2000 Researcher (no additional information about type of research)

8919 Restaurateur/assisting in kitchen

4705 Restaurateur/cooking

1503 Restaurateur/supervising staff and administration

6505 Restaurateur/waiting on tables

1503 Restaurateur (no additional information about specific tasks)

3999 Retoucher/photographic  
 4503 Retoucher/printing  
 1311 Secretary/assistant/senior government officer/computing division  
 1307 Secretary/assistant/senior government officer/distribution division  
 1313 Secretary/assistant/senior government officer (except above)  
 6503 Secretary/club/tending bar  
 1599 Secretary/club (except above)  
 1201 Secretary/trade union  
 5601 Secretary/receptionist  
 5101 Secretary (no additional information about specific tasks)  
 5101 Secretary (except above)  
 4405 Signwriter<sup>93</sup>  
 4921 Silverer/glass  
 4923 Silversmith  
 2815 Singer

616. The much smaller number of index entries in a structured coding index than in a complete listing index is a result both of the restriction of the index to key words when possible and the use of '(except above)' instructions, which allows the exclusion from the index of a large number of responses where the qualifying words are immaterial for the selection of the correct group.

617. In a structured index for coding occupation, the key word is the single word in the relevant response that alone can serve as an occupational title, however imprecise. The qualifying words usually will indicate some form of specialization of tasks. Sometimes the key word may be precise and in itself suffice as an index entry, such as, 'Silversmith' in the examples above. However, the key word may also be very ambiguous, as in the 'Researcher' and 'Secretary' examples above. Note that the qualifying words in some of the 'Secretary' examples do not serve to distinguish between specializations, but between occupations that are very different in nature.

618. The construction of a structured coding index must reflect and support the coding rules to be used for assigning occupation codes on the basis of the responses to the relevant questions, using, when necessary, permissible ancillary information given in other responses and indicated in the index as qualifying words. This means that one should organize the index alphabetically first in terms of key words, then in terms of the first qualifying word with those entries which also have a second qualifying word listed before those which do not, and the '(except above)' instruction should follow the entries with qualifying words. The key words listed in the index must reflect those which can be selected from the permissible parts of the responses, and the qualifying words must reflect those which can permissibly be selected, and in the order in which they should be selected.

---

<sup>93</sup> Note that although the term 'signwriter' may not appear in dictionaries as a single word, it needs to be included in the coding index if it is given as a single word in census responses. An entry 'Writer, sign' would also need to be included in the index.

619. In English it will be normal to use as key words those that can be found in either (as first priority) the title component of the occupational response or (as second priority) the task component. First priority for qualifying words will be those normally found in the title or task components of the response. Second priority, and based on the rules for using industry information in the coding of occupation, should be given to words found in the response for industry or name/type of employer.

620. The advantages of having a structured coding index are threefold. First, it causes the coder to search for index entries in a way that is consistent with the coding rules. Second, it speeds up the task of coding by restricting the coder's search through the index because of the smaller number of entries. Third, when the index is searched either by computer or human, it reduces the risk of finding matches with words that are irrelevant for the coding decision and therefore reduces error.

621. Some words may be found to be in common usage in job titles, but can be ignored for the purpose of creating a coding index. For example, words such as 'senior', 'junior' and 'executive', when given as qualifying words, may not carry information about the tasks which constitute a particular job. It is possible to exclude such words from the index and to instruct coders to ignore them.

## **C Using the occupation coding index**

### ***1. Using the occupational response***

622. Coding can be seen as a process in which the task of the coder is to 'translate' the information provided by the recorded responses to the appropriate code in the occupational classification structure. The main tools for this 'translation' are the coding index and the coding instructions - including instructions on when a response should be treated as a query to be resolved by supervisors or expert staff. The instructions should specify:

- (a) how this translation process should be carried out;
- (b) what items to look for in the occupational response and in what order;
- (c) what type of ancillary information to use from other responses;
- (d) when such ancillary information can permissibly be used and how to use it.

Ideally the coding index should have been constructed to reflect and support the use of these instructions.

623. The starting point should always be the response given to the occupational question(s), i.e., question(s) on the type of work in the person's own job and the usual or main tasks and duties carried out in the job. This question should normally result in a job title and a few words on main tasks. When using an "all inclusive" coding index the coder should start by marking those words that are relevant for the search in the index. When using a structured index the coder should identify the "key word", normally part of the

response pertaining to the job title, and look for this in the index. The tasks part of the response should then be used where necessary to supplement or modify the information provided by the title, or may need to be transformed to a title, e.g., 'baking bread' to 'Baker, bread', 'cleaned school' to 'Cleaner, school'. Transformation of a task response to a title should be performed when there is no proper title response or when there is no index entry corresponding to the title given, as may be the case for 'civil servant', 'helper' and other non-informational titles. If the title is inconsistent with the tasks or if the occupational responses are not sufficient to determine a detailed occupational group then the coder should choose one of three alternatives:

- (a) Look at the form for recorded ancillary information of a specified type for further clarification.
- (b) Use an appropriate code for inadequately specified responses.
- (c) Refer the case to supervisors as a query.

The coders should be given clear instructions and training on the circumstances that would allow any of these alternatives to be chosen. Coders should not be expected to make judgements that require an understanding of the classification structure.

## **2. *Using ancillary information on industry or name and type of employer***

624. Most modern occupational classifications, including ISCO-88 and ISCO-08, are designed on the principle that 'occupation', meaning a particular pattern of work tasks and skills which constitute an individual's job, should be kept conceptually separate from 'industry', meaning the sector of the economy in which the job is performed. Thus, an 'electrical maintenance fitter' may work in any of a range of different industries and this person's occupation cannot be validly deduced from knowledge of the industrial category of the employing organization. Without breaching this principle, it must nevertheless be recognized that certain occupations or occupational titles are to be found solely or predominantly in particular industrial sectors. In such cases, knowledge of the industry may clarify an occupational title or description that is inadequate for coding purposes. For example a 'face worker' working in the coal mining industry may be deduced to be a miner engaged in coal cutting.

625. In other cases the descriptions of work activities used to identify an occupation are best formulated in terms of, for example, the nature of the material worked with (e.g., wood, rubber, leather, etc.). This information may be deducible from knowledge of the industrial sector in which the job is located and again help to clarify a vague occupational description. For example, the occupational term 'coil winder', used on its own, is ambiguous because coding depends on whether what is wound is some form of metal wire, some form of textile product, etc. Knowledge that the job is located in a textile-manufacturing establishment may be sufficient to resolve the ambiguity with a reasonable degree of certainty.

626. Thus, because some interrelationships between occupation and industry are

inherent in the industrial structure of the economy, they can be made use of to improve the accuracy of occupational coding. However, there are costs as well as benefits in this practice. In the first place, there is always some danger that inferences from industry to occupation will be based on incorrect or out-of-date assumptions about the distribution of occupations across industries. In the second place, when coding is being done on a large scale, coding work rates and inter-coder consistency are important considerations. Coding rates are likely to be slowed if the coder routinely will consider extra sources of information in order to arrive at a code, particularly if the extra information is itself hard to interpret. In such circumstances there is also a danger of increasing inter-coder variability, since different coders will tend to interpret the information in different ways. These latter two problems are minimized if:

- (a) industry is coded in advance of, or at the same time as, occupation so that no further interpretation of verbatim responses is required for the occupation code;
- (b) coders are allowed to use data on industry only where the responses to the specific questions on job title and activities are inadequate to determine an occupational code; and
- (c) if the choices that an occupational coder can make on the basis of the industry code are exhaustively specified through index-referenced instructions.

627. A simplified example may clarify this: A coder encountering the job title 'coil winder' would be instructed in the coding index under 'Winder, coil' to look first at the description of job activities for information on the type of material wound. In cases where no indication of this was found the coder would look next at the pre-allocated industry code. If this was code 'x', standing for 'textile manufacturing', the occupational category would be determined as 'textile yarn winder'; if the industry code was 'y', standing for 'electrical machinery manufacturing', and the occupational category would be determined as 'wire winder'. If the industry code were other than 'x' or 'y' the occupation would be placed in an appropriate 'inadequately described' category by the coding index.

### **3. *Use of other ancillary information for occupation coding***

628. Some countries allow information about the educational and vocational qualifications of respondents to be used ancillary information that it is permissible for coders to use to determine the appropriate occupational code. Again, this should be based on detailed knowledge of the relationships between training and qualifications, on the one hand, and the corresponding occupations, on the other. In all countries this relationship varies between occupations, and in most countries the relationship is close only for a limited number of occupations. Even when the relationship is close it must be recognized that the fact that a person has a particular qualification does not mean that his/her job will include the corresponding tasks. (A person with a medical degree who is working in a hospital may not have tasks corresponding to his/her degree. This may be because s/he has been promoted to a job which consists mainly of management tasks, or it may be

because s/he could not get a job corresponding to the type and level of technical training, for example, because s/he lacks necessary language skills.)

629. For these reasons, education is most useful in an exclusionary function. A person working in a hospital without a medical degree can reasonably be excluded from being a medical doctor. The exclusionary function also can be employed as an edit to resolve problems associated with distinguishing between levels of the same discipline. The use of information on qualifications or educational attainment as ancillary information should therefore be very carefully controlled and probably restricted to query resolution by expert coders or to computerized edits of coded responses.

630. The use of ancillary information of any kind in an occupation coding process may have a tendency to bias the relationship between occupation and other Census variables collected, as well as impacting negatively on the efficiency of the coding process. Coders should not, therefore, be allowed to use responses to questions on age, income, hours worked or other socio-economic variables to determine an occupation code.

#### **4. *Inadequate occupation responses and queries***

631. Some responses simply cannot be coded to a detailed occupational group. This will normally be for one of the following reasons:

- (a) The response may be vague, i.e., it does not contain enough information to be coded according to the coding index and coding rules; or
- (b) The response may be precise, but may use a title and/or indicate types of tasks or combinations of tasks that do not correspond to any of the index entries. Supplementary material should be used by experts to code such precise responses and then make appropriate changes to the index.

632. Unfortunately, the number of cases of type (a) is likely to be quite substantial, even with well-formulated occupational questions and well-trained enumerators. In order to keep to manageable proportions the number of queries which the supervisors and expert coders must handle, the coding index and the coding instructions should be designed to guide the coders with respect to the most common of such cases.

633. The simplest solution will be to specify that the response should be coded to a 'default' group, as in the examples of 'Researcher', 'Restaurateur' and 'Secretary' in the above Australian example. This default group may in some cases be a specific detailed group because this reflects the dominant usage of the terms found in the response, cf. '1503 Restaurateur' and '5101 Secretary' as default groups. However, the default group will often have to be one of the aggregate groups in the classification, because it is not possible to identify one particular detailed group as dominant within the aggregate group indicated. In the Australian example above '2000 Researcher' indicates that a response giving only "researcher" as information could only be coded validly to Major group 2. Similarly a response like 'Manager, running a business' would normally have to be coded to the aggregate group for 'Managers', unless the industry response gave very clear

information.

634. There is a real danger that coders may use 'default' groups as 'dump groups' for difficult to code responses before they have tried to find a precise code. Some countries have therefore tried to keep the first line coders ignorant of the possibility of using such codes and only allowed them to be used by better-trained supervisors. This strategy may create a morale problem among the first line coders and a very large query burden on the supervisors. It may therefore be better to allow and monitor carefully the use of 'default' codes by first line coders.

635. The fully specified responses which are not adequately covered by the coding index should always be handled by expert coders and recorded carefully, both to ensure consistent treatment of similar cases and because these cases represent an important source of information for the updating of the coding index as well as the classification itself. During the coding operation, these cases can either be handled using priority rules specified for the classification or by assigning them to one or several groups (or supplementary codes) for occupations not adequately covered by the classification.

636. Priority rules can be applied to some of the responses that indicate task combinations that cut across the groups defined in the classification, e.g., 'Baker, baking/selling/managing shop'. Most classifications based on ISCO-88 or on ISCO-08 will specify priority rules in terms of tasks performed for the allocation of such jobs to occupational groups. In ISCO-88 it is specified that priority should first be given to the tasks that require the highest skill level, and secondly that production-oriented tasks should be given priority over managerial or administrative tasks. 'Main tasks', in terms of time spent, for example, are not to be given priority unless they completely dominate, both because an employer is likely to be concerned that a worker can carry out the most skilled tasks required, even if they are only seldom activated, as in emergencies, and because information about time allocation of tasks is normally not available. Thus in the above example the code to be specified in the coding index should be the code for "baker" according to the priority rules of ISCO-88, but this may change in ISCO-08.

637. Precise responses which cannot be resolved by priority rules should be assigned special "inadequately described" or "not defined" codes placed in groups created for coding purposes within the aggregate groups to which the jobs clearly belong. Steps should also be taken to ensure that these cases can be closely examined outside the coding operation to determine what, if any, contribution such cases can make to the updating of the coding index and to the classification. Note that these groups are not the same as the 'not elsewhere classified' groups of the classification. Great care must be taken to avoid confusion between these two types of group.

#### **D Developing and updating the industry coding index**

638. Most census coding operations will find it useful to have two "coding indexes" for the coding of industry.

- (a) A list of as many as possible of the establishments which, at the time of the census, are/were operational in the geographic region covered by the coding operation, where each establishment has been given the correct industry code by those who are specialists in establishment surveys and the coding of establishments' activity. In practice such lists (or business registers) may often cover only large, formal sector establishments as they have been created from lists kept in tax offices, licensing offices, and/or chambers of industry and commerce. They may nevertheless cover significant proportions of the work force, and their use for census coding will eliminate one possible source of inconsistency in employment statistics between the census results and the results of establishment surveys.
- (b) A list of significant word combinations reflecting the answers given in response to industry questions, i.e., an index of the same type as that created for the coding of occupation.

639. The industry coding process will therefore usually involve firstly an attempt to match the name and address of the person's employer with those in the list or register of establishments. If a match cannot be made using the register of establishments, then an attempt is made to match the description of the industry with the index of type (b) above.

640. The process of updating the coding indexes for industry responses should be viewed as part of the general process required to maintain the industry classification. As new ways of organizing work between or within enterprises or new technologies are introduced, new functions, end products and services with separately identifiable units, i.e., separate establishments, will appear. For example with the introduction of movies on video tape, units of production providing rentals of these tapes were created. New industries will be created and some old industries may disappear. However, frequently those responsible for census preparations will find that while work to maintain establishment lists of type (a) above may have been carried out by those responsible for an establishment register or for establishment surveys, this work has not been undertaken with respect to lists of type (b), neither by the custodian of the national standard industry classification, nor by anyone else, and that in fact not only the updating but also the creation of this type of industry coding index must be undertaken from scratch for the census.

641. It is important to note that indexes of type (b) for use in population censuses and other household or person based data collection activities, are quite different from those used in economic surveys of establishments. The words used in responses provided by individuals to questions about the kind of industry of their employers, are likely to be quite different from those provided in establishment based collections. Where information on industry is regularly collected in household surveys such as Labour Force Surveys, it is possible that such an index may have been maintained for this purpose.

642. A full-scale establishment monitoring exercise cannot be used for the census preparations as that would be too time-consuming and costly. The most realistic alternatives would be to carry out (i) post-coding reviews of recent household survey



operations; (ii) reviews of advertisements and notices in newspapers and other media; and (iii) consultations with tax authorities, chambers of commerce, etc. In addition as much help and information as possible need to be extracted from those responsible for establishment surveys.

643. The first procedure makes use of the tools and members of the coding team(s) used for recent survey(s). Coders are a good source of information on the adequacy of the index and other tools that they used. Their suggestions for improving the index and for additional or revised entries must be recorded and investigated. Ideally preparations for the use of 'post-coding review procedures' for the development of index material should have been designed into the data processing operation, as it is essential that information which may be useful in updating a classification and index be carefully collated and retained. Such information typically consists of records of problems encountered, queries raised and decisions and amendments to the working instructions adopted in the course of coding. Such procedures should be made part of the normal routine for continuous or regular surveys, e.g., Labour Force Surveys, as well as for the registration of establishments and activities which takes place in the local tax offices, licensing offices or chambers of commerce, and by those responsible for establishment surveys.

644. Notices and advertisements in newspapers, journals and/or bulletin boards about their products and services, creation and expansion, or about vacancies, will be used by many formal sector businesses, and they may provide a useful source for constructing or up-dating information on establishments and their activities. Telephone directories, trade directories and similar publications are most useful. They often classify industries (yellow pages) and in some cases provide short descriptions for some entries. The main disadvantage is that small, informal sector establishments are not normally well covered by such reference materials.

645. The list of establishments should for each unit give both a name and the physical location, indicated by a street address if possible, otherwise by naming the (smallest) district in which the unit is located or is operating as one unit. If alternative forms of the name, such as abbreviations, initials, old names, etc., are used or have recently been in use, they should also be included in the list as separate entries, because of the possibility that persons working for them may use these variants in their answers. The entries in the list should be organized alphabetically, with clear rules for where to find entries consisting of initials and abbreviations.

646. When organizing the material for the coding index of activities (industries) the first issue concerns the structure of the index itself. Basically the choice is between two different approaches. In some statistical agencies the approach has been that the index should be all-inclusive, that is, every distinct type of response found in the process of coding should, in theory, have an entry in the index, although allowance may be made for misspellings and inversions of words that are without consequence for the meaning of the response. An advantage of this approach is that it may be possible for coders, when faced with an obscure type of activity or product, to find those terms listed in the index. The

main disadvantage is that the index may become very large and its sheer size may slow down the process of searching for the "right" entry in the index, and thereby slow down the coding, whether the coding is done manually or is computer-assisted.

647. Large verbatim indexes also create the impression that coding is a simple task, involving a straightforward matching between a response and an index entry. However, no matter how large the index, it will always be the case that a significant proportion of responses fail to match the index entries exactly, and one has to use rules and/or judgement to make the 'best' match. Once an exact match is not found, the chances that an incorrect entry will be selected will tend to increase with the number of entries available. For these reasons, an alternative approach may be to develop a structured index.

648. A *structured index* does not try to reflect every possible response because it is accompanied by instructions to the coder on how to break down the available response into key words and qualifying nouns or adjectives. The primary entries in the index are the key words. If a key word in itself is not sufficient to uniquely identify the group, an appropriate qualifying word (or phrase) must be added to distinguish between the possible alternatives having the same key word. If this is not sufficient to resolve all ambiguities, second or higher order qualifying words should be used. The following examples may illustrate the system for transforming an industry (type of activity) response into an entry in a structured coding index according to the following format:

1. Response:
2.     ⇒Key word/1st qualifying word/2nd-qualifying word:

Examples:

Sheep farm:⇒sheep/farm  
Car rental agency:⇒car/rental  
Youth club:⇒club/youth  
Tax assessment office:⇒office/tax/assessment  
Cleaning service:⇒cleaning/services  
Cleaning products production:⇒cleaning/products/production

649. The following examples from the coding index used for coding industry to the United Kingdom (UK) Standard Industrial Classification of Industrial Activities 1992.

15.11/1 Abattoir  
74.40 Advertising/agency  
74.40 Advertising/agent  
74.40 Advertising/campaigns/creation  
74.40 Advertising/campaigns/realization  
74.40 Advertising/consultant  
74.40 Advertising/contractor  
92.11 Advertising/film/production

31.50	Advertising/lights/manufacturing
74.40	Advertising/material/design
22.22	Advertising/material/printing
22.22	Advertising/newspaper/printing
22.12	Advertising/newspaper/publishing
74.40	Advertising (no further information)

650. In a structured index for coding industry, the key word is the word in the relevant response that alone can serve as a designation of a service, a product or a function, however imprecise. The qualifying words usually will indicate some special form or variety and/or the type of activity associated with the product or service. This sequence has been chosen because the number of different designations for activities is much smaller than the number of designations of different products, services and functions. Sometimes the key word may be precise and in itself suffice as an index entry, such as 'Abattoir' in the examples above. However, the key word may also be very ambiguous, as in the Advertising examples above.

651. If coding of industry and occupation are to be done by the same staff, the construction of the structured coding index for industry must not be in conflict with coding rules to be used for assigning the occupation codes on the basis of the responses to the relevant questions, using, when necessary, permissible ancillary information given in other responses and indicated in the index as qualifying words. This means that one should organize the index alphabetically first in terms of key words, then in terms of the first qualifying word with those entries which also have a second qualifying word listed before those which do not, and the 'except above' instruction should follow the entries with qualifying words. The key words listed in the index must reflect those which can be selected from the permissible parts of the responses, and the qualifying words must reflect those which can permissibly be selected, in the order in which they should be selected. Rules for identifying the key/functional words for industry may need to be quite different from those for occupation, however.

652. The advantages of having a structured coding index are threefold. First, it causes the coder to search for index entries in a way that is consistent with the coding rules. Second, it speeds up the task of coding by restricting the coder's search through the index because of the smaller number of entries. Third, when the index is searched either by computer or human it reduces the risk of finding matches with words that are irrelevant for the coding decision and therefore reduces error.

## **E Using the industry coding indexes**

### ***1. Using the industry response***

653. Coding can be seen as a process in which the task of the coder is to 'translate' the information provided by the recorded responses to the appropriate code in the occupational classification structure. The main tools for this 'translation' are the coding

indexes and the coding instructions - including instructions on when a response should be treated as a query to be resolved by supervisors or expert staff. The instructions should specify:

- (a) how this translation process should be carried out;
- (b) what items to look for in the industry response and in what order;
- (c) what type of ancillary information to use from other responses;
- (d) when such ancillary information can permissibly be used and how to use it.

Ideally the coding indexes should have been constructed to reflect and support the use of these instructions.

654. The industry response is the information given in response to a question or questions about the name and address of the person's employer in a particular job, and the main type of industrial activity undertaken by the employer at that location. The starting point should always be the response given to the first part of industry question(s), i.e., question(s) on the name and geographical location, e.g., street address, of the place of work. If the name list provides an exact match on both name and location, then the industry code given to this unit in the name list can be given to the response. If there is no exact match then the coder should make use of the regular coding list for industry, by selecting a word from the response which provides information about the type of products, services or function the unit produces or provides. If this is not sufficient to determine a code, as with "advertising" in the example above, then the coder should identify supplementary words, qualifiers, which may give more precise information about the product, and/or the type of process involved. If the industry response does not contain information sufficient to determine a detailed industry group then the coder should choose one of three alternatives:

- (a) look at the form for recorded ancillary information of a specified type for further clarification;
- (b) use an appropriate code for inadequately specified responses;
- (c) refer the case to supervisors as a query.

The coders should be given clear instructions on the proper alternative to choose.

## **2. *Using ancillary information on occupation***

655. Most modern industry classifications, including the International Standard Industrial Classification (ISIC), Revision 3, are designed on the principle that 'industry', meaning a particular set of productive activities resulting in one or more products or services produced by an economic unit, should be kept conceptually separate from 'occupation', meaning the type of work performed by a person working in the establishment, and since many different occupations may be represented in the same establishment, one cannot normally, draw valid conclusions about the industry of the workplace from one person's occupation, even if this happens to be an occupation which

tends to cluster in a particular industry, e.g., bus drivers. However, there are a few exceptions to this, e.g., “university teachers” are only found in the “education” industry and ‘taxi drivers’ only in transportation, and for some “own-account” workers there will be a direct link to a particular industry, e.g., own account plumbers should logically only work in the construction industry. Such cases can be identified and incorporated into the industry-coding index.

656. Thus, because some interrelationships between occupation and industry are inherent in the industrial structure of the economy, they can be made use of to improve the industry coding. However, there are costs as well as benefits in this practice. In the first place, there is always some danger that inferences from occupation to industry will be based on incorrect or out-of-date assumptions about the uniqueness of an occupation to a particular industry. In the second place, when coding is being done on a large scale, coding work rates and inter-coder consistency are important considerations. Coding rates are likely to be slowed if the coder will routinely consider extra sources of information in order to arrive at a code, particularly if the extra information is itself hard to interpret. In such circumstances there is also a danger of increasing inter-coder variability, since different coders will tend to interpret the information in different ways. These latter two problems are minimized if:

- (a) occupation is coded in advance of or at the same time as industry, so that no further interpretation of verbatim responses is required for the industry code;
- (b) coders are allowed to use information on occupation only where the responses to the specific questions on job title and activities are inadequate to determine an industry code; and
- (c) the choices that an industry coder can make on the basis of the occupation information are exhaustively specified through index-referenced instructions.

657. A simplified example may clarify this. A coder encountering the establishment name 'Institute for marketing studies' might be instructed in the coding index under 'Institute/marketing' to choose between the code for “marketing” and the code for “education”. If the occupation is given as “manager”, “secretary” or “janitor” there would be no basis in the occupation information to make the choice, as all three types of occupations may exist in either industry. However, if the occupation is given as “account executive” then it is likely that the establishment is a marketing firm (with a fancy name), and if the occupation is given as “professor” or “lecturer” then it is likely that the establishment is a training institution.

### **3. *Use of other ancillary information for industry coding***

658. Among the information included in the census questionnaire only occupation seems relevant as ancillary information for industry coding. The only exception is that one could imagine that in one-company locations it may be possible to extend the range

of occupations which will only be found in that company's establishments, thus reducing or eliminating the possibility that a bus driver living in this location, for example, could be employed by any other type of establishment than the local bus company, as this company owns and runs all busses in the area.

#### **4. *Inadequate industry responses and queries***

659. Some responses simply cannot be coded to a detailed industry group. This will normally be for one of the following reasons:

- (a) the response may be vague, i.e., not contain enough information to be coded according to the coding index and coding rules; or
- (b) the response may be precise, but may use a title and/or indicate types or combinations of products, services or functions which do not correspond to any of the index entries.

660. Unfortunately, the number of cases of type (a) is likely to be quite substantial, even with well-formulated industry questions and well-trained enumerators. In order to keep to manageable proportions the number of queries which the supervisors and expert coders must handle, the coding index and the coding instructions should be designed to guide the coders with respect to the most common of such cases. The simplest solution will be to specify that the response should be coded to a 'default' group. This default group may in some cases be a specific detailed group because this reflects the dominant usage of the terms found in the response, such as '74.40 Advertising' as default group in the example above from the United Kingdom. However, the default group will often have to be one of the aggregate groups in the classification, because it is not possible to identify one particular detailed group as dominant within the aggregate group indicated.

661. As stated above, there is a real danger that coders may use 'default' groups as 'dump groups' for difficult to code responses before they have tried to find a precise code. Some countries have therefore tried to keep the first line coders ignorant of the possibility of using such codes and only allowed them to be used by better-trained supervisors. This strategy may create a morale problem among the first line coders and a very large query burden on the supervisors.

662. The fully specified responses, that are not adequately covered by the coding index, should always be handled by expert coders and recorded carefully, both to ensure consistent treatment of equal cases and because these cases represent an important source of information for the updating of the coding index as well as the classification itself. During the coding operation these cases can either be handled using the priority rules specified for the classification or by assigning them to one or several groups for occupations 'not adequately covered' by the classification.

663. Priority rules are difficult to apply to the responses that indicate activity combinations that cut across the groups defined in the classification, e.g., 'repairing cars & selling petrol', because most industry classifications based on ISIC, Revision 3 and 4

will specify priority rules in terms of contribution to value added of the enterprise, or number of persons employed; and this information will not be available to the coders or their supervisors. One solution may be to refer these cases to the classification specialists and they may be able to identify the establishment and on that basis determine a code.

664. Precise responses which cannot be resolved by priority rules should be placed in special 'not adequately covered' groups created for coding purposes within the aggregate groups to which the activity clearly belong. Steps should also be taken to ensure that these cases can be closely examined outside the coding operation itself to determine what, if any, contribution such cases can make to the updating of the coding index and of the classification itself. Note that these groups are not the same as the 'not elsewhere classified' groups of the classification, which include clearly defined activities which are not important enough in scale to be given a separate identity within the larger group to which they belong. Great care must be taken not to confuse the two types of groups.

**PART 6: USING POPULATION CENSUSES  
TO IMPROVE LABOUR FORCE AND RELATED  
STATISTICS**





## **Chapter 12. Types of Data Collection on the Labour Force and Other Economic Characteristics**

### **A Organization of the Part**

665. The previous Parts have reviewed how a population census may collect information on the labour force and other economic characteristics, either through a complete coverage or from a sample. As indicated in the earlier Parts, there also exist other sources of such information for national level statistics, in particular household surveys, establishment surveys, economic and agricultural censuses, and administrative sources. This Part is concerned with how the results from the population census may be used to develop, evaluate, improve and better utilise these other sources, specifically labour force and other household and establishment sample surveys. An important underlying issue is how the population census design and procedures may be chosen so as to maximise its usefulness for this purpose.

666. This Part describes the various types of sample surveys providing information on economic characteristics of the population and presents aspects of the structure and arrangements for the labour force survey, which in many countries is the main other source of such information. It discusses ways in which the population census may be utilised to develop, evaluate and improve such surveys, and aims to clarify the diverse manners in which sampling may be used in conjunction with the population census. These include, for the purpose of census operations, evaluation and tabulation, and for the purpose of additional data collection through surveys attached to or based on the census. The Part explains the concept of sampling frame and master samples, two of the main uses of the population census, and describes how the population census provides the basis for constructing them. The other uses of census data covered in the Part are weighting and preparing estimates from survey results, consistency between and combined uses of census and survey data, and the use of census information for the evaluation of content and coverage of post-censal sample surveys, and for small area estimation.

### **B Diversity of sources**

667. The primary objective of a census is to obtain information on the size and a few basic characteristics of the whole population, and to provide the maximum possible detail for local areas and small domains or for particular groups of the population. Usually, only selected topics on economic characteristics are investigated in a population census, and Chapters 6 to 8 has given a detailed review of how to implement the topics to be included in a census in accordance with the *Principles and Recommendations, Revision 2*.

668. As already stressed the number of questions used in a population census to investigate any of these items has to be limited, often only to a single question per item, (i.e., activity status, occupation, industry, status in employment, working time, income, sector of employment and place of work), even though actual practices vary significantly

between countries. In contrast, household sample surveys, by virtue of their smaller size, can be designed to obtain a wide variety of data for different kinds of analyses. This chapter provides a description of the diverse types of sample surveys that provide information on economic characteristics of the population. This is followed in the next chapter by a description of the main type of source on economic characteristics and activity of the population, namely the labour force survey. Subsequent chapters describe ways in which the population census may be utilised to develop, evaluate and improve such surveys.

669. Apart from the population census, data on the labour force and other economic characteristics come from a variety of sources:

- (a) The main source for this purpose is, of course, the labour force survey (LFS). The LFS is often the largest, official survey conducted continuously or repeated at regular intervals. It is normally based on large probability samples, with a wide coverage of the general population, and often forms the basis of other household surveys or there may be modules on special subjects attached to the core LFS.
- (b) There are diverse other population-based surveys providing information on economic characteristics and activity of the population, e.g., surveys focused on special populations, or particular economic categories, or particular areas of the country; surveys collecting information on labour force and other economic characteristics as background variables or as part of a larger set of social variables; longitudinal surveys, etc.
- (c) There can also be large-scale one-time census-like operations, covering variables that cannot and/or need not be collected on a census basis. Such surveys are normally closely linked to the population census in terms of objectives and operations, such as intercensal demographic surveys (sometimes referred to as sample censuses). These differ from 'normal' sample surveys in that the infrastructure and impetus developed during the census permits exceptionally large sample sizes in the period closely following the census.
- (d) Information on economic characteristics is also obtained from surveys with establishments (rather than households) as the reporting units. These economic surveys tend to have special sampling requirements, which arise because of the nature of the units involved. The units involved are often predominantly small-scale and numerous, scattered but unevenly distributed in the population, heterogeneous in type and size, etc.
- (e) The above also includes economic and agricultural 'censuses', which, despite the name, are often conducted on a sample basis – albeit on samples of large size.

670. As noted by the International Conference of Labour Statisticians<sup>94</sup> “Population censuses and sample surveys of households or individuals generally constitute a comprehensive means of collection of data on the economically active population which can be linked with data on other related topics. Establishment surveys and administrative records may also serve as sources for obtaining in some cases more precise, more frequent and more detailed statistics on particular components of the economically active population. The different sources of information should be regarded as complementary and may be used in combination for deriving, where necessary, integrated sets of statistics ...” Such integrated sets of statistics on employment are in some countries prepared as supplements to the national accounts.<sup>95</sup>

### **C Household-based inquiries**

671. Population censuses and household surveys cover in principle the same population and employ the same type of measurement units (households and individuals). Differences between the uses to which population censuses and household surveys may be put arise primarily from the differences in the scale of the operations involved (complete enumeration versus sampling), which lead to differences in methodology, practical conditions of implementation, timing and complexity of the data collected. As noted earlier, the number of questions used in a census to investigate any of the issues covered has to be limited, often to a single question per item. In contrast, a household labour force sample survey, by virtue of its smaller size, can be designed to obtain a wide variety of data for different kinds of analyses. It can be tailored more flexibly to fit a variety of users' needs and methods of data collection. Household surveys, though by no means inexpensive, are obviously less costly than complete censuses. They can be repeated more frequently and thus provide information on changes over a period of time. Because of their smaller size, sample surveys also permit better control of response and other non-sampling errors, and the results can be produced more speedily.

672. Due to the limited size of the samples, the major limitation of household surveys is their inability to provide sufficient detail for small areas or subgroups in the population. In addition, samples of moderate size, while often capable of providing good estimates of proportions, rates, ratios, etc., tend to be less satisfactory for estimating population aggregates (such as the total number of unemployed persons in a particular group) and changes in such aggregates, i.e., for items that may be of particular interest to the user. To obtain reliable estimates of population aggregates, it is usually necessary to supplement or adjust data from samples by using information from other sources, often from the population census or population registers, where available.

---

<sup>94</sup> International Labour Office (1983). Thirteenth International Conference of Labour Statisticians (ICLS), Resolution I, para. 3.

<sup>95</sup> Some of the methodological issues that have to be tackled and some relevant national experiences are reviewed in E. Hoffmann: “Developing labour account estimates: Issues and approaches” in *Household Accounting: Experiences and Concepts in Concepts and Compilation*. Vol. 2: *Household Satellite Extensions*. Studies in Methods. Handbook of National Accounting. Series F, No. 75 (Vol. 2). United Nations, New York, 2000.

673. Mutually supportive roles and combined uses of population censuses and sample surveys have been described in many sources.<sup>96</sup> The sampling method can profitably be used to facilitate the planning, testing, controlling, evaluating, processing and supplementing of census data collection. In return, the census experience provides the infrastructure, sampling frame, benchmark statistics, etc., that are needed to conduct household sample surveys, together with a general impetus to the development of statistical capability. These contributions are particularly important for large-scale sample surveys with wide coverage, such as national labour force surveys.

674. Combinations of the two also exist. As discussed earlier the population census has been designed in many countries to have two components: (a) an enumeration of the population and its basic demographic and related characteristics on a 100 per cent basis, supplemented by (b) a large sample, attached to the census, covering a broader range of items. Such a design can considerably enhance the role of the census as a source of statistics on the economically active population and related topics.

675. It should further be noted that data from population censuses and labour force surveys are increasingly being used in conjunction with suitable statistical techniques to yield post-census and current estimates for local areas and small domains.

676. Labour force and other household surveys allow for the joint measurement of the employed, unemployed and economically inactive. They can be designed to cover virtually the entire population of a country, all branches of economic activity, all sectors of the economy and all categories of workers, including own-account workers, contributing family workers, and persons engaged in casual work or marginal economic activity. This gives such surveys a unique advantage for obtaining information on the total labour force and its structure. As concepts, definitions and subject details can easily be adapted to particular data requirements, different degrees of labour force attachment among various groups of the population can be measured. There is also considerable flexibility as regards the data items that can be covered. Since in household surveys households or individuals are reached directly, relevant supplementary information on demographic and socio-economic characteristics of individuals and households can be obtained at relatively low additional cost along with information on labour force characteristics.<sup>97</sup> This offers many possibilities for data analysis. With appropriate design and rules of association, household surveys can also provide a means to collect information on household-based and other small-scale establishments.

#### **D Establishment-based inquiries from information on the population's economic characteristics and activities**

677. While household surveys constitute a primary source of information on the economically active population, some of the data can also be obtained from other sources.

---

<sup>96</sup> See for example United Nations, (1984); paras. 1.7-1.11 and Kish and Verma, (1986).

<sup>97</sup> United Nations (1984). *Handbook of Household Surveys*. Series F, No. 31, paras. 11.10-11.11.

A basic decision in statistical planning concerns the choice of an appropriate combination of methods so as best to meet the various data needs in a given situation. In addition to the population census and labour force and other household surveys, main sources of labour force data include: (1) censuses and sample surveys of establishments; and (2) administrative records of different types. The various sources differ in coverage, scope, units of measurement and methods of data collection. Each source has advantages and limitations in terms of the cost, quality and type of information yielded. Generally, one approach tends to be stronger where another is weaker, and vice versa. The various sources tend, therefore, to be complementary rather than competitive or mutually exclusive. Their results can be combined to some extent, depending on the degree to which concepts, definitions, coverage, reference periods, classifications, etc., agree.

678. Hence, depending on specific requirements, statistics on employment and related topics may also be obtained from censuses and sample surveys of establishments, as distinct from population censuses and household sample surveys. The differences between these two systems concern the population covered, the reporting units and the scope of the information collected.

679. Compared with household-based surveys, censuses and sample surveys that use establishments as reporting units are generally more focused, hence more precise and economical. They can also be more limited in coverage and content, however. In relation to coverage, two types of establishments may be distinguished: (i) large establishments belonging to the more organised sector of the economy, possibly registered in some formal way and employing more than a certain number of persons; and (ii) small establishments in the relatively unorganised sector, which may be either non-household establishments or purely household-based operations run by households on a proprietary or partnership basis. Reasons of cost, logistical difficulties or specific objectives confine many establishment surveys to larger establishments. The employment statistics obtained from such surveys tend therefore to be restricted to paid employees in the more organised sector of the economy. Often, such surveys cover only particular branches of economic activity, such as mining and quarrying, manufacturing and construction. They are normally based on existing list frames, and tend to suffer from the more or less serious deficiencies of coverage common to such frames.

680. Furthermore, rules of association between different types of survey units can be rather complex. The sampling units used for selection, the responding units that provide the information, and the units of enumeration and analysis on which information is sought in the survey may be difficult to identify and associate with one another consistently. The practical difficulties of distinguishing consistently between establishments and enterprises are well known. An illustration of another kind of difficulty would show how an establishment survey system enumerating occupied jobs on the basis of persons listed in the payrolls, would exclude persons with a job but temporarily away without pay, but would count multiple jobholders more than once. Such problems do not arise in household surveys.

681. The strength of establishment surveys lies in their greater *specificity*, both in terms

of coverage and content. When the interest is in specific industries, establishment surveys, given an adequate sampling frame, can achieve more efficient sample design, i.e., a larger coverage of the employed population, and procedures than household surveys covering the whole population. More reliable and more detailed information on certain topics, such as labour turnover, earnings, social security and pension benefits, etc., can be obtained in establishment surveys, especially where the respondents can draw upon payrolls and other available records. This type of survey can provide an opportunity to collect information on many other economic variables such as output, costs, investment, technological and organisational factors, which can then be directly related to information on employment, wages and productivity and can form a much more comprehensive basis for the analysis of economic activity.

682. Establishment surveys may also be more economical and timely than household surveys. This is because there tend to be fewer respondents in the former, especially if only larger units are covered, and they tend to be congregated together and easier to contact. Moreover, cheaper methods of enumeration such as mailed questionnaires or telephone interviews can sometimes be used in place of the more expensive face-to-face interviewing. Another cost-reducing factor is that a single or a few respondents can provide summary information on all the persons employed in a large establishment.

683. Turning to small establishments, the distinction between establishment and household surveys is less clear-cut. List frames are generally not available for small establishments, which are characterised by high fluctuation and often lack recognisable features, and therefore the only feasible approach is the usual household survey, based on multi-stage sampling of areas with special listings of units at the last stage. However, unlike households, even small establishments tend to be rather unevenly distributed in the population, often in pockets of considerable concentration by type of economic activity. Information on the pattern of distribution from censuses or other sources is helpful and often necessary to improve economy and efficiency of survey design and implementation.<sup>98</sup>

## **E Administrative sources**

684. Note should also be taken of non-survey/census based sources of information on economic characteristics and activity of the population, though this is not a primary concern in the present Part.

685. Statistics based on administrative records are by-products of administrative processes. Administrative records can thus be a very economical source of statistical information. They are often based on continuous operations, and the statistics can therefore in principle be produced at almost any frequency or for any reference period. Other administrative records such as the payrolls and files of civil service organisations, government enterprises and other public institutions, may also be usefully exploited to

---

<sup>98</sup> Murthy M.N, and Roy, A.S. (1970), pp. 123-135.

obtain information on parts of the employment in the public sector. Where these sources are available and tabulated at frequent and regular intervals, they can be used to good account in particular analyses, if careful attention is paid to their coverage and other limitations. Administrative sources can also be used in compiling and updating sampling frames for employment and related surveys based on samples of establishments, at least in respect of the larger establishments in the more organised sectors of the economy.

686. However, administrative sources can also suffer from various shortcomings such as limited coverage and content, inflexible concepts and definitions, incompleteness, inconsistencies, and restricted access due to legal or administrative constraints. As noted in the *Handbook of Household Surveys*,<sup>99</sup> in developing countries with unorganised labour markets, administrative sources such as unemployment insurance and employment exchange records often do not exist at all or are limited to certain narrowly defined categories of workers. A more comprehensive discussion of the possibilities of using administrative records can be found in ILO/EASMAT (1997)<sup>100</sup> and Pember (1998)<sup>101</sup>

---

<sup>99</sup> United Nations, (1984); paras. 1.13, 1.14.

<sup>100</sup> ILO/EASMAT (1997): *Labour Statistics Based on Administrative Records: Guidelines on Compilation and Presentation*. ILO Regional Office for Asia and the Pacific, Bangkok.

<sup>101</sup> Pember, R J: *Compilation and presentation of labour statistics based on administrative records*, in ILO's Bulletin of Labour Statistics, 1998-1.

See also <http://www.ilo.org/public/english/bureau/stat/papers/listart.htm>





## **Chapter 13. Labour Force Survey Structure And Arrangements**

### **A Purpose of the Labour Force Survey**

687. In order to appreciate the relationship between the population census and sample surveys on economic characteristics of the population, and how the census may be designed and used to aid the latter, it is useful to describe main features of the structure and arrangements of labour force surveys, the objective of this chapter.

688. The labour force survey is meant to be a comprehensive source of information on economic activity of the entire population. According to the International Conference of Labour Statisticians, the programme of statistics of the economically active population should cover all branches of economic activity, all sectors of the economy and all categories of workers, and should be developed to the fullest extent possible in harmony with other economic and social statistics: “The programme [of statistics of the economically active population] should specifically provide for both short-term and long-term needs, i.e., statistics for current purposes compiled frequently on a recurrent basis and statistics compiled at longer intervals for structural in-depth analysis and as benchmark data.

- (a) The current statistics programme should encompass statistics of the currently active population and its components in such a way that trends and seasonal variations can be adequately monitored.
- (b) The non-current statistics programme, which may include censuses and surveys should provide:
  - (i.) comprehensive data on the economically active population;
  - (ii.) in-depth statistics on the activity pattern of the economically active population over the year and the relationships between employment, income and other social and economic characteristics; and
  - (iii.) data on other particular topics (e.g., children and youth, women households) as determined by the long-term continuing needs”.<sup>102</sup>

### **B Common characteristics of surveys on the economically active population**

689. National practices in conducting labour force surveys or, more generally, surveys of the economically active population vary greatly, depending on specific data requirements and survey conditions and facilities. However, a number of commonly encountered features can be identified on the basis of available empirical information.

---

<sup>102</sup> International Labour Office (1993). Thirteenth ICLS, 1982, Resolution I, para. 2. Minimum requirements concerning the frequency, item coverage and classifications of the statistics are set forth in the Labour Statistics Recommendation (No. 170), adopted by the International Labour Conference in 1985 to supplement the Labour Statistics Convention (No. 160).

- (a) Labour force surveys tend to be relatively large-scale surveys of the whole population; they are often national in scope and have some sort of official status. Especially in developing countries, only the national statistical office or some other major public agency engaged in statistical work can generally undertake such major operations. In many instances, the primary users of the information are also major public agencies.
- (b) Many countries undertake labour force surveys on a continuing basis with the objective of measuring current levels of employment and unemployment for important groups of the population and the corresponding changes. Monthly or quarterly releases of survey results of course constitute the predominant pattern, particularly in developed countries, while in many others, such surveys conduct them at quarterly or less frequent intervals. A number of countries also undertake surveys that aim at providing more detailed information on the structure of the labour market, of longer-term interest, because of the value of such surveys in providing essential information for planning and policy formulation at the national level. There is however not enough emphasis being put on the production of structural information of longer-term interest, as compared to the attention given to the production of statistics on current levels.<sup>103</sup>
- (c) In most countries, information pertaining to households and individuals in labour force surveys is collected through face-to-face interviewing by field staff visiting survey respondents in private households. Other methods of data collection, such as mail or telephone interviewing, are largely unfeasible in many countries, though telephone interviewing is used increasingly at least in some, mostly developed, countries for the second and subsequent interviews with the same household.

690. The common characteristics of labour force surveys sketched above have a number of consequences for survey design and execution. Firstly, because of their official status and national scope, labour force surveys are likely to be subject to rather stringent requirements of timing, data accuracy and internal consistency, especially consistency of the time-series generated by the continuing type of survey. These requirements can only be met by probability samples of fairly large size, drawn from a good frame covering the whole population and representative of it, not only geographically but also over a period of time because of seasonal and other variations and the need to provide good estimates of changes. The survey estimates have to be as consistent as possible with external data from other official sources. Many users require estimates of aggregates or population totals (as distinct from, and in addition to, estimates of means, proportions and rates), and estimates of changes in such aggregates. These generally require the use of ratio estimates inflated appropriately on the basis of control totals obtained from outside the survey (see Chapter 16).

---

103 For a documentation of Labour Force Surveys from about 90 countries see ILO (2002b): Sources and Methods: Labour Statistics. Vol. 3, 3<sup>rd</sup> edition: Economically active population, employment, unemployment and hours of work. International Labour Office, Geneva.

691. Secondly, in many situations, the executing agency of a labour force survey is also involved in various other surveys and statistical operations. This increases the importance of co-ordination and integration in survey planning, design and execution.

692. Thirdly, where personal interviewing is involved – which is mostly the case in labour force surveys in developing countries, and at least for the first enumeration in developed country surveys - the time and cost of travel for field work are often a major component of the total survey cost. The sample has therefore to be clustered through multi-stage area-based designs.

693. Fourthly, many operations have to be decentralised, resulting in a need for thorough interviewer training, supervision and quality control procedures and measures.

694. At the same time, however, there can be a tendency to avoid experimentation and innovation because of the necessary size, regularity and repetitive nature of the survey operations. Proper attention therefore needs to be paid to the continuous evaluation and periodic redesign of survey methods and procedures. This is by no means an easy task. The problem can be serious when, as is often the case, the available resources are limited and there is pressure to increase the quantity (variety, volume) of statistics routinely generated, at the expense of evaluating and improving the quality of the statistics. Rigid timetables for data collection can further accentuate the quality control problem.

## **C Diversity of labour force survey structure and designs**

695. Among household surveys of the economically active population a variety of designs and arrangements are possible. The primary determining factors are the substantive objectives of the survey, i.e., the content, complexity, timeliness and periodicity of the information sought. The survey may be designed to obtain regular time-series for statistics on current levels and trends. Alternatively, or in addition, it may focus on less frequent information of a more structural nature and longer-term interest. In contrast, the survey may be confined to a few basic characteristics of the labour force, such as the levels of employment and unemployment. These substantive considerations will also determine the appropriate timing, frequency, reference period, sampling arrangements and other aspects of the survey structure. The requirement of appropriate linkages with other surveys, both in terms of subject matter and field operations, can be another important factor determining the survey structure and arrangements.<sup>104</sup>

696. The following sections consider several dimensions of this diversity. Firstly, a major distinction by type of survey is between continuing versus occasional labour force surveys. Secondly, the LFS may be linked in different ways to other household surveys in the country, from a separate survey to a system of closely integrated surveys. There can also be other types of surveys, different from the ‘typical’ LFS, providing information on

---

<sup>104</sup> For further discussion, see Hussmanns, R., Mehran, F., and V. Verma (1990). *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods*. Geneva: International Labour Office.

economic characteristics and activity of the population.

## **D Survey frequency**

### **1. *Continuing surveys for current data***

697. Continuing surveys are primarily conducted to generate a time-series of statistics on current levels and trends. With appropriate design and data of sufficiently high quality, the survey may also provide estimates of gross changes and flows of individuals between different activity statuses and types of economic activity, but such statistics are yet to be published on a regular basis, even when the requisite information have been collected. Typically, such a survey consists of an ongoing series of survey "rounds", each round being designed to produce separate estimates covering a specified time period. Continuing surveys are used to monitor the performance of the economy; to obtain indicators of changes in current rates of labour force participation, employment, unemployment and underemployment; and to measure trend, cyclic and seasonal variations in these rates.

#### *1.1. Surveys conducted on a continuous basis*

698. Among continuing surveys, two types of arrangement are commonly found. The first arrangement is to conduct the survey on a continuous basis, i.e., the fieldwork is carried out uninterrupted. Typically, the information is obtained by using a moving reference period, i.e., a reference period relating to a specified duration immediately prior to the interview, which varies among respondents depending on when the interview is held. If the survey round is of a sufficiently long duration, it may be divided into "sub-rounds", each covering a representative sample during a part of the whole period for the survey round. Division into sub-rounds permits better, in the sense of more representative, coverage of the sample over shorter time segments during the round (such as over months or quarters during the year). Results from the sub-rounds can therefore be used to study seasonal and other variations during the round. The system also permits more frequent and timely release of the results.

699. The distinction between rounds and sub-rounds is mainly one of sample size. In the case of each round, with a larger sample size obtained by the accumulation of sub-rounds, it may be possible to tabulate and analyse the data in fuller detail; whereas in the case of individual sub-rounds the sample size may only be large enough to provide main estimates with adequate precision, with less geographical and other disaggregation. With the field work divided into representative sub-samples for each time segment (sub-round), the aggregated results for the whole round are themselves improved, since seasonal and other temporal variations during the round are covered or averaged-out in a more balanced way. Division into sub-rounds can also have important operational advantages: the fieldwork can be better controlled and distributed more evenly over a period of time. The cost of the system is of course increased travel in order to cover well-dispersed samples separately during each sub-round.

## 1.2. *Surveys conducted periodically*

700. The second arrangement is that of periodic surveys with intermittent field work concentrated over relatively short intervals. There can be certain advantages in concentrating field work:

- (a) It can make it easier to control and implement field operations.
- (b) It becomes easier to obtain information with a "fixed" reference period, i.e., with the same reference period for all respondents in terms of fixed calendar dates.

701. While it may be considered necessary or even preferable for certain purposes to use such a fixed reference period, there can be some disadvantages in concentrating fieldwork.

- (a) The average conditions over a period such as a year may not be as well represented in periodic surveys as in continuous surveys with fieldwork evenly distributed throughout.
- (b) Periodic surveys do not provide field enumerators with a continuous and evenly distributed workload.

702. A choice has thus to be made between (a) employing permanent field workers and letting them remain idle for part of the time; (b) resorting to the use of temporary staff during the periods of field work; and (c) having a permanent field staff and using it for other survey operations during the slack period, such as sample updating, editing, coding, or for other statistical work.

703. The first option, (a), is obviously wasteful, though apparently it has been seriously considered in some situations where the use of permanent, well-trained enumerators is considered essential to ensure data quality, while personnel field costs are relatively low because of low wages. Option (b), i.e., the use of temporary staff, can give more flexibility, but can result in lower quality and hinder the development of permanent survey capability of the organisation. Option (c), combining field operations across surveys, is in common use, particularly in countries with relatively well-established statistical infrastructures, where permanent field staff stationed in different parts of the country can handle a variety of data collection tasks, including household survey interviewing. In the intervals between periods of fieldwork for the labour force survey, the field staff can continue with data collection for other statistics. A distinct alternative would be to use the staff between rounds of the survey not for other field work and data collection, but for editing, coding, summarising, and evaluating the labour force survey data collected during the preceding round of field work. This is becoming more feasible with increasing decentralisation of the data-processing operations with the development of communication and computer technology. It should also be mentioned that in some countries the national statistical office has sub-contracted the actual field operations and processing for the labour force survey to commercial survey organizations, subject to appropriate quality control and confidentiality mechanisms.

704. Given that the main purpose of continuing surveys, whether through continuous or periodic field work, is to generate a regular sequence of data, it is imperative that, in such surveys, the results are released regularly and opportunely, and that the volume and complexity of the information collected do not overwhelm the organisation's data processing and reporting capacity. Otherwise, unprocessed data will pile up in increasing quantities, and the whole objective of providing current statistics will be defeated.

### *1.3. Occasional surveys for more structural information*

705. Comprehensive surveys of the economically active population may be conducted less frequently to obtain benchmark data and detailed structural information at the national level. This may include, for example, detailed information on the economically active population by industry, occupation, status in employment, on activity patterns over the year, work experience, multiple job-holding, education and training, hours worked, income from employment and so on. Similarly, the population not economically active may be classified by type and various socio-economic and demographic characteristics. In national development planning such surveys are needed for analyses of the employment conditions at the beginning of the plan period, and for fixing targets and goals.

706. While surveys of this type are not designed to yield a continuous flow of current statistics or information on changes over short periods, they are well suited to provide less frequently needed information on essential structural characteristics and on longer-term changes, mostly at the national level. As these characteristics do not change rapidly, it is not necessary to undertake such surveys more than once every few years. For example, they may be conducted every five years as post-censal or inter-censal surveys in countries with decennial population censuses. In any case, it is often simply not feasible to take such detailed surveys more frequently because of resource constraints.

707. The timing of the survey has to be determined carefully. The results should be available when structural and benchmark data are most needed, as for example to provide a basis for the formulation of development plans. To be of long-term value, the survey timing should not coincide with periods of abnormal or transient employment conditions (unless the measurement of such abnormality itself constitutes a main objective of undertaking a special survey). The survey period should also take into account seasonal and other short-term variations. For these reasons, it may be useful to spread out field work over a whole year covering all seasons; this may be done on a continuous basis with uninterrupted field work throughout the year, or in the form of more concentrated periods of field work spaced out over the year.

## **E Linkages of the labour force survey with other surveys**

708. There is an increasing use in many countries of household surveys as a source of a wide variety of statistical information. As a consequence, it is often necessary to undertake surveys, whether on the economically active population or on other topics, as part of a common survey system or programme, or at least to ensure that individual

surveys are closely co-ordinated. These requirements can be particularly important in the case of surveys on the economically active population, which often tend to be comprehensive in coverage, national in scope, and relatively large in size. The need for linkages becomes even greater when a survey involves regular operations repeated periodically or continuously, for these can greatly affect and be affected by other operations in which the statistical organisation is simultaneously involved.

709. Linkages between surveys involve two broad aspects:

- (a) co-ordination at the design and operational level, where common procedures, arrangements and facilities are used to increase the economy and flexibility of the operations; and
- (b) integration at the substantive or subject-matter level, where a number of topics are covered in conjunction with each other, to permit the production of inter-related statistics which can be analysed jointly.

710. *Co-ordination* implies that individual surveys are designed and undertaken in proper operational relationships to one another, utilising common procedures and infrastructures, including organisational arrangements, sampling frames and other materials, technical and supervisory staff, field and office personnel, as well as transport, data processing, printing and other facilities. The degree of co-ordination and sharing of facilities may vary, depending on the type of organisation involved, the nature of its operations, special requirements, funding and other arrangements for the surveys, and so on. While smaller one-time surveys can often be carried out on the basis of more or less special or ad hoc arrangements, this is much less likely to be the case for major undertakings such as national labour force surveys. Often, labour force surveys are carried out as part of the regular operations of a national statistical agency, and their planning requires careful consideration of operational links with other undertakings, an evaluation of the possible constraints and problems (such as increased pressure on available data-processing facilities and technical staff), as well as of the opportunities and flexibility which these linkages can offer.

711. *Integration* at the substantive level implies the use of a common study population; common concepts and definitions; a common system of classifications, and possibly standard survey questions for frequently used classifiers such as age, sex, ethnic group, education and economic activity status; and common or overlapping samples of respondents. Sometimes the term "complete integration" is used to indicate coverage of multiple topics in a single survey, over a common sample, and possibly during a single interview with the respondent. By contrast, the term "partial integration" implies a situation where the various topics are covered using the same sample of areas but with different samples of households within each area. Only complete integration permits data linkages at the micro-level.

712. In practice, various patterns can be found in the manner and degree of linkages of labour force surveys with surveys on other topics:



- (a) a labour force survey may be organised as an operation more or less separate from other surveys;
- (b) a more comprehensive labour force survey may serve as a vehicle for covering other related topics as well;
- (c) a labour force survey with limited content may be incorporated into some other ongoing survey as a "module";
- (d) a labour force survey may form part of a multipurpose survey covering a range of topics, or of surveys specially designed to enumerate population groups of special interest;
- (e) a labour force survey undertaken as one round of an ongoing survey system which focuses on different topics in different rounds.

### ***1. Separate labour force surveys***

713. A number of developing and developed countries carry out surveys, which are primarily or exclusively concerned with labour force topics. In the present context, these may best be described as "separate" labour force surveys. Their single-subject focus does not, of course, preclude operational co-ordination and the use of common facilities and arrangements with other surveys, or the use of common coverage, concepts, definitions and classifications. What is implied by "separate" is their single-subject focus and a considerable degree of separation in design and execution. Such separation can sometimes be helpful in providing better control and supervision and greater flexibility in the design and operation of the survey.

### ***2. Labour force survey as vehicle for other data***

714. The establishment of a continuing labour force survey can be a major and relatively expensive undertaking. Once in place, the survey can be usefully exploited as a vehicle for covering additional topics and for supporting various household surveys in other areas. Indeed, diverse household surveys may be developed to use the labour force survey capacity and become integrated with it in design and operations to varying degrees, and conducted as supplements to the LFS. As an alternative, or in addition, we may have more independent surveys using the same sample areas as the LFS, but different sets of households and different survey periods.

### ***3. Labour force "module" attached to other surveys***

715. Conversely to the above, a labour force survey with limited content may itself be attached as a "module" to surveys focused on other topics. It is most effective when the survey serving as the base for the LFS is a continuing one, so that it can be used for generating a regular time series of labour force statistics. Inclusion of basic items in other ongoing surveys, where they are useful as background variables for the main items of the surveys, can be an extremely economical way of obtaining some essential information on the labour force and its characteristics. This can be particularly useful when a full-fledged

labour force survey cannot be undertaken because of limited resources or competing priorities. A module on labour force items may be included in other surveys on an occasional basis, or on a more sustained basis, depending on the requirements.

716. However, it is also necessary to recognise some limitations of the approach. Firstly, there are limits to the number and detail of labour force items that can reasonably be inserted into operations concerned primarily with other topics. In a population census, for example, each of the few labour force items that may be included has to be confined generally to a single simple question, or to a simple sequence similar to that used in the last population census. Care must be taken to ensure that such additions do not adversely affect the overall quality of the information obtained in a census or survey due to excessive respondent burden, delays in data processing, or other consequences of the increased size and complexity of the operation.

717. Secondly, to ensure data quality and usefulness of the results, it is necessary that the various topics included in the same survey are compatible in terms of concepts, definitions, survey methods, reference periods, coverage and other design requirements. It is not always possible or easy to achieve such compatibility. At the same time, the requirement of compatibility with other topics covered in the same survey may itself limit the usefulness of the resulting labour force data. Furthermore, the surveys may be confined to specific groups in the population, so that the labour force data can only be used as explanatory variables for other topics.

#### **4. *Omnibus multi-purpose and specialised surveys***

718. Integration can also take a more extreme form in which a large number of detailed topics are combined in a single omnibus undertaking. In principle, the main advantages of comprehensive multi-purpose surveys are the possible economies of scale, and the potential for cross-checking and for combined analysis of detailed data on various topics. Multi-purpose surveys can thus yield a wealth of information on labour force variables in conjunction with other topics, provided that the samples are sufficiently large. It should be noted, however, that complex multi-subject surveys can, and often do, suffer from serious disadvantages, especially in the more difficult circumstances of some developing countries. Such disadvantages include the increased length and complexity of the interview, increased respondent burden, possible increase in non-sampling errors, reduced efficiency of design for any particular topic due to compromises needed to accommodate diverse requirements, and the danger of delays and failures at the data-processing stage because of the increased volume and complexity of the data collected.<sup>105</sup> Some of these problems may be reduced in scope, for example, by using different sub-samples for different sets of topics in addition to certain core topics obtained from the overall sample (in which case, not all the topics can be linked at the micro-level), or by organising data processing separately for different sets of topics. Nevertheless, caution is needed to avoid making any survey system too complex or over-burdened.

---

<sup>105</sup> United Nations (1984). *Handbook of Household Surveys*, paras. 1.20-1.28,

719. Occasionally, more specialised surveys may be undertaken to investigate in depth certain relationships, special phenomena, problems and issues, or population groups of special interest. Possible examples are relationships between labour input, training and experience on the one hand, and income from employment, family income, welfare, etc., on the other. However, such surveys tend to be quite complex in content and involve special arrangements, specialised staff, and relatively heavy cost and effort. Consequently, they are usually undertaken on a one-time or infrequent basis, often with less than national coverage and with smaller and less well-dispersed samples. In so far as a specialised survey is "research oriented", (i.e., aims at providing information of longer-term interest to enhance the general understanding of issues and problems, rather than at meeting some more immediate data needs), there may be considerable flexibility in its timing. For the same reason, comprehensiveness of content and high quality of data may be more important considerations than extensive coverage and quick release of data.

720. It may also be noted that sometimes survey objectives require the collection of additional information on particular population groups of special interest such as the handicapped, migrants, female household heads, unemployed young persons or underemployed workers. Where such groups are small, special arrangements such as multi-phase sampling with screening may be required to include sufficient numbers of respondents. In continuing surveys, there is also the possibility of accumulating such cases from several rounds. If the additional information required on groups of special interest is too detailed or complex, it may be necessary to organise its collection as an operation separate from the main labour force survey (employing for instance special questionnaires and/or special interviewers), though co-ordinated with it in an appropriate manner.

721. Population censuses, demographic surveys, household income and expenditure surveys and other surveys often collect basic information on the economic characteristics and activities of the population. Often the objective is to provide "explanatory" variables which are useful for cross-classification and analysis of those characteristics which are the primary focus of the survey. Although in this case the survey is not aimed at providing estimates for labour force variables per se, it can be useful in enhancing the understanding of relationships between labour force and other characteristics such as fertility, child care, health, income, consumption behaviour, and so on. Indeed, some topics are so closely related to labour force characteristics that any survey on the former may require a fairly comprehensive coverage of the latter as well. For example, a survey on migration may need to include items such as activity and employment status, occupation, industry, sector of employment and income, in order to help understand the causes and consequences of migration.

##### **5. *LFS as periodic round of a survey system with varying focus***

722. Another pattern of integration, which may be particularly suited to conditions and requirements in some developing countries, is to establish an ongoing survey system with a varying substantive focus from round to round. Each round may cover a specified

period such as a year and a separate, representative sample. The survey system can use common organisation, personnel and other facilities, but the subject-matter changes from round to round; with the possible exception of some core items common to all rounds. Among the common elements are likely to be some on the employment situation of the respondents, in addition to the basic demographic variables. Such a system can include periodically comprehensive surveys of the economically active population as the main focus of the rounds. The comprehensive survey can provide structural or in-depth information of longer-term interest. Such an arrangement has many potential advantages for survey work in developing countries<sup>106</sup>, and is a model that was promoted through the United Nations National Household Survey Capability Programme.<sup>107</sup>

---

<sup>106</sup> The Indian National Sample Survey provides a well-known example of such a survey system; see for example, Murthy and Roy (1970) and Murthy, M.N. (1975).

<sup>107</sup> United Nations (1980a). The National Household Survey Capability Programme: Prospectus. DP/UN/INT-79-020/1. New York.



## Chapter 14 Sampling And Related Uses Of Census Information

### A Sampling in conjunction with the population census

723. In most countries today the decennial census of population is the primary source of geographically detailed information on the basic demographic, economic and related characteristics of the population. Many official statisticians agree that a population census need not gather all demographic and housing information on a 100 per cent basis. The issue of the use of sampling in conjunction with the population census is pertinent in the context of the development of a system of labour force surveys (LFS). The LFS is often the largest and most important of population-based survey undertaken regularly in many countries.

724. According to the United Nations, "The use of sampling actually saves a good deal of time and money; and, furthermore, in certain circumstances, only the sampling approach ensures data of acceptable accuracy."<sup>108</sup> The practice of governments is increasingly to collect information on their populations' size, age and sex composition, geographic distribution, and certain other basic demographic and socio-economic characteristics on the basis of a complete (100 per cent) enumeration, and to supplement this basic information by collecting information on a larger range of variables on a sample basis. The additional information, when it is gathered as part of the census operation, may be related to employment, income, migration, mortality, fertility, and health.<sup>109</sup> **(PEMBER COMMENT: IF THIS FOOTNOTE CANNOT BE UPDATED, THEN WE SHOULD DELETE IT.)**

725. Sampling may be used in conjunction with the population census diverse ways and for different purposes, such as in census operations (planning, design, testing, evaluating, tabulating), and for additional data collection. Additional data may be obtained in the census in a number of ways, such as: (i) from enumeration of more detailed items on a sample basis as part of the census; (ii) through sample surveys attached to the census; through large-scale surveys closely following the census; and more generally; (iii) through developing a system of sample surveys based on the census frame for sample design and selection; and (iv) using the census data in the production of improved estimates from the surveys. By combining the census with a sample survey it is

---

<sup>108</sup> United Nations (1971). Handbook of Population and Housing Census Methods, Part VI, Sampling in Connection with Population and Housing Censuses. Studies in Methods, Series F, No. 16; page 2. See also, United Nations (2007). Principles and Recommendations for Population and Housing Censuses, Revision 2. (Sales No. E.07.XVII.8), paras 1.408-1.441.

<sup>109</sup> For example, among the 19 countries of the Asian and Pacific region (including the United States) for which information is available about the 1980 census round, as many as 12 employed complete and sample enumerations in combination, whereas in the remaining seven only complete enumeration was used. See: Cho, Lee-Jay, and Robert L. Hearn, (eds.) (1984). Censuses of Asia and the Pacific: 1980 Round. Honolulu: East-West Population Institute, East-West Center.

possible to exploit the advantages of each.

726. Combining data collection on a 100 per cent and a sample basis during the census is just one way (albeit the most important) that sampling is used in conjunction with the census. Complete census and sampling can be used in combination or in related ways with the objective of capturing the advantages of each.<sup>110</sup> These applications include:

- (a) using sampling in the design and control of census operations, such as in planning, testing, controlling, and evaluating the census;
- (b) using sample enumeration to supplement the items covered in the complete census. In certain circumstances it is also possible to consider substituting complete enumeration with one or a series of sample enumerations;
- (c) sampling the census results for processing, with the objective of making the results available more quickly and at lower cost; extracting samples of micro-data files of detailed census data so as to facilitate dissemination of primary data more widely for analysis by other users;
- (d) using the census as a basis for sample surveys by enhancing its statistical capability and resources and by providing baseline data, population controls for estimation, and sampling frames for surveys in the post-censal period; and
- (e) using the census and post-censal sample survey data in combination to provide estimates for local areas and small domains.

## **B Census as source of sampling frames**

727. In most countries, sampling frames for household surveys are based on the most recent available census of population, though the census information is often supplemented or updated from other sources for this purpose. It is much less common to have alternative sources entirely independent of the population census.

728. While the census of population is obviously the primary source of sampling frame for household surveys, with foresight and careful planning it can also serve to create area frames for surveys and censuses of establishments, particularly small-scale establishments, in the agricultural and non-agricultural sectors.

### **1. The survey population**

---

<sup>110</sup> See for instance: Kish, L., and V. Verma. 1986. Complete censuses and samples. *Journal of Official Statistics* 2(4):381-395. Murthy, M. N. 1980. Use of Sampling in Population Census with Reference to the Experience in Some Countries of Asia and Pacific. Report of the Training Seminar on Use of Sampling in Population Census. New Delhi: Office of the Registrar General of India. Verma, V (1989). "The Use of Sampling in Conjunction with Population Censuses". *Asian and Pacific Population Forum* Vol. 3 Nos. 1-2, pp. 11-20, 35-40.

729. The definition of the population to which the sample results are to be generalised is a fundamental aspect of survey planning and design. While basic decisions about the nature and scope of the population to be covered are taken early in the survey planning process, the content and extent of the population has to be specified more precisely at the stage of technical design. This specification is in terms of:

- (a) population content, i.e., definition of the type and characteristics of the elementary units comprising it;
- (b) population extent in space, i.e., the boundaries of its geographical coverage;
- (c) and its extent in time, i.e., the time period to which it refers.

730. For example, in the labour force and other household surveys, the elementary or analysis units are generally households and persons. Many surveys are aimed at covering the general population, but may nevertheless exclude (for substantive or practical reasons) certain types of household or persons: such as collective, institutional, foreign, diplomatic, or non-civilian households; homeless persons or families; persons temporarily abroad; or persons above or below certain age limits. Various considerations may also result in the exclusion of certain areas from geographical coverage of the survey; examples are remote and sparsely populated areas in many countries. Units and their characteristics change with time, and the population covered in any survey is also bounded in time, even though these boundaries may not be as sharply defined as geographical boundaries.

731. Delimitation of the population content and extent is often defined more narrowly in surveys and censuses in the non-household sector. For example, many surveys of establishments cover only those units that, in relation to a certain specified period in the past, existed and satisfied certain criteria in terms of the size and type of activity (sector/industry) of their operation. In agricultural censuses and surveys as well, it is common to limit the coverage to holdings meeting certain minimum size criteria, which may differ according to the type of activity.

732. Three important practical points should be noted. First, in any survey, rules of population inclusion and exclusion must be defined in clear operational terms. Otherwise confusion and errors result at the implementation stage. Secondly, the limitations in the population covered must be kept in view in drawing inferences from the survey results, and in comparing results from different sources. It is important to document the extent of exclusions, since they define the limits within which inferences can be scientifically drawn from the survey. The effect of exclusions depends on the type of statistics produced. For instance, sometimes, the exclusions do not greatly affect the rates and ratios estimated from the survey, but affect estimates of population aggregates more seriously. Rates and ratios are affected to the extent exclusions are selective with respect to the characteristics measured. Thirdly, one should periodically reassess the assumptions behind excluding certain parts of the population which should ideally be covered, and seek clear justification for the continued exclusion in each case. Sometimes it may be possible to adopt special arrangements or take reduced samples from the more difficult and expensive parts of the target population, rather than to exclude them



altogether.

733. Apart from deliberate and explicit exclusions, surveys also suffer from coverage errors, which are less easily identified and measured. Painstaking work is usually required to control these errors and assess their effect on the survey results. Their magnitude depends on the quality of the sampling frame and sample implementation.

## 2. *The sampling frame*

734. The population to be surveyed has to be represented in a *physical form* from which samples of the required type can be selected. A *sampling frame* is such a representation. In the simplest case, the frame is simply an explicit list of all units in the population, from which a sample of the units concerned can be selected directly. With more complex designs, the representation in the frame may be partly implicit, but there is still need to account for all the units.

735. In practice, the required frame is defined in relation to the required structure of the sample and the procedure for selecting it. Hence in order to understand the concept and requirements of the sampling frame, it is necessary to understand some basic aspects of structure of the samples as typically encountered in large-scale household surveys. The simplest, but generally quite impractical, design is to select a simple random sample of the elements of interest such as households or individual persons. In such a design, each element and each combination of elements in the population has the same constant probability of appearing in the sample. The most common aspects in which samples in practice depart from simple random sampling are: multi-stage sampling; stratification; and in some situations, selection of the elements with unequal probabilities.

736. A basic distinction to be made is between list frames from which the units of interest can be directly selected; and the more general multi-stage frames, which are usually area-based. In the latter, the frame for the first stage of sampling (called the primary sampling frame) has to cover the entire population. Following the first stage of selection, the list of units at any lower stage is required only within the larger units selected at the preceding stage.

737. Several variations and combinations are possible. A frame may be constructed from a single source, or may have to be compiled by combining information from a number of sources. Different types and/or sources of frames may be used for different parts of the population. It is also possible to use more than one frame in combination to represent the same population more adequately.

### 2.1. *Area-based frames*

738. In the context of household surveys, especially in developing countries, the frame generally consists of one or more stages of area units, followed by lists of households or dwellings within the selected ultimate area units. It is useful to distinguish three

components of such a multi-stage frame:

- (a) The primary sampling frame (PSF), which is a frame of the first or primary sampling units (PSUs) and must cover the entire population exhaustively and without overlaps. Following the first stage of selection, the list of units at any lower stage is required only within the larger units selected at the preceding stage.
- (b) Possibly, a hierarchy of secondary area-based frames - consisting at each stage of units covering but confined to units selected at the preceding stage, till a frame of the lowest or ultimate area units (UAUs) is obtained. Below the UAUs, the sampling process moves from areas to the listing and selection of individual dwellings, households or persons.

Explicit lists of the ultimate sampling units (USUs) such as dwellings or households within the selected UAUs. The elements for data collection and analysis in the survey may be the USUs themselves, or may be other units uniquely identifiable from the USUs through definite rules of association. For example, persons (elements) may be associated with selected households (UAUs) on the basis of "population present" or "usual resident population" coverage definition.

739. The three components have major differences in terms of the cost and durability of the frame. The durability of the frame declines as we move down the hierarchy of the units. The lists of ultimate units require frequent updating. In most surveys, it is necessary to prepare fresh lists of USU's shortly before the survey enumeration. It is a major advantage in a survey if it can utilise lists prepared for some other recent survey or census. Lists of structural units such as dwellings are usually more durable than lists of social units such as households; pre-existing lists of individual persons are hardly ever useful.

740. By contrast, the primary sampling frame (and to a lesser extent, frames of intermediate level units) usually represents a major investment for long-term use. A most important decision in the development of the PSF concerns the choice of the type of units to be used as the PSUs. Several practical considerations in this choice may be noted. Larger areas generally provide more stable and clearly demarcated units, but that choice can also increase the segmentation, listing and other work required to complete lower stages of the frame construction and sample selection. The main source of the PSF is the most recent census of population.

741. Similar considerations apply also to secondary frames of area units, but to a much lesser extent. For one thing, secondary frames are required only for the PSUs selected in the first stage, and therefore it is not impossible to consider their construction for use over a limited duration, or even for the purpose of a single survey. In any case, many surveys are based on a design involving a single area stage (e.g., a sample of census EAs as the PSUs, followed by direct listing and selection of households within selected EAs), and no secondary area frames are involved except on a selective basis where particularly large original PSUs need to be segmented.

742. Area-based frames are also used for surveys of small-scale agricultural and non-agricultural establishments, which are too numerous and dispersed to be covered by list frames, and for which no registers or lists are available in most cases. Special procedures are required in constructing such frames because the units to be covered are often very unevenly distributed in the population. Compared with households in the general population, they also tend to be more heterogeneous in size and characteristics, and are divided into distinct types each requiring separate coverage.

### 2.2. *Census enumeration areas (EA's) as the PSU's*

743. Careful consideration must be given during the planning and execution of the census to its function as a source of area sampling frame for diverse surveys. The EAs of the census have multiple functions to:

- partition the population into geographic areas with clear, stable, and identifiable boundaries, which can be mapped and described;
- facilitate complete and unique coverage of the units in the population; to create reasonably equitable and feasible workloads;
- facilitate the organisation and control of census operations; to provide a flexible basis for the production of areal statistics at various levels and types of aggregation; and to provide a basis for scientific and efficient sample selection for subsequent surveys.

Although these requirements cannot all be satisfied simultaneously to the same degree, they indicate desirable characteristics of census EAs.

744. EAs should be small (containing a few hundred people on the average) as well as reasonably uniform in population size. However, the requirement of clear boundaries is more important than uniformity in size. EAs should be real units covering the entire country exhaustively and should be mapped and described for clear identification of boundaries; they should not usually cut across administrative subdivisions or natural boundaries. They should be geographically ordered with proper identification systems, to facilitate the production of results at different levels of aggregation and for different types of geographic units, e.g., administrative, commuting, water catchments. Information on their size and other basic characteristics should be collected, coded, and tabulated for individual areas.

745. The census generally provides sampling frames or master samples only of areal units for subsequent surveys. Lists of housing units, households, and individuals from the census are usually not allowed outside the organization responsible for the census, because of confidentiality concerns. Such lists are also too difficult to arrange, and individual households are too mobile to be used for subsequent surveys, except perhaps in surveys attached to or conducted soon after the census.

### 2.3. *List frames*

746. Completeness of the frame is a most critical requirement (and perhaps also the

most common problem) of list frames. It is also important that the list contains pertinent and accurate information on the size and other characteristics of individual units so as to permit efficient stratification and control of the selection process.

747. When the units to be surveyed are relatively large in size and limited in number, they are best selected in a single stage from a list frame. Often good lists exist or can be compiled with reasonable effort for large units. However coverage and other problems generally increase as we move to smaller and more numerous units, especially when there is no system for registration of the units or for updating existing lists. Several situations may be distinguished for separate treatment, depending on the size, number and distribution of the units involved, for instance:

- (a) large and few units for which good lists may be available;
- (b) medium-sized units, which also can be covered only by a list frame, but for which such frames are more difficult to construct and maintain;
- (c) medium to small units which may require a combination of list and area-based frames;
- (d) and small or very small units which can only be covered with area-based frames.

748. The last category may itself require distinction into several types. For instance:

- (a) Numerous and relatively well-dispersed units may be adequately covered by an area-based frame of the general population.
- (b) Relatively numerous but unevenly distributed units would require specially constructed area frames, taking account of the patterns of concentration.
- (c) Some special populations (such as fishing households, or others engaged in very specialised activity) are confined to pockets of concentration, which may be numerous but vary greatly in size. They are not adequately captured by the general frame; yet it can be extremely difficult to identify and list all the 'pockets'. The best that can be done in some circumstances is to concentrate the effort on covering the larger and more important concentrations, with an attempt to improve the coverage over time, as knowledge on the patterns of distribution of the units accumulates.
- (d) Rare but dispersed populations also require special methods and arrangements, as do mobile populations.

#### 2.4. *Multiple frames*

749. When more than one overlapping frames are used simultaneously, it is important to ensure that the units' probabilities of selection remain definite and known. Various approaches are possible in selecting samples from multiple frames. Perhaps the simplest option is to make the frames non-overlapping, if possible. This may be done by eliminating duplicated listings of any unit from all but one frame. Another way is to impose an order of priority among the frames such that any unit can be selected from only one of the frames. For instance, if a list and an area frames are used in combination,

any unit selected from the latter may be automatically excluded if it is also present (whether or not selected) in the list; in other words, any unit present in the list frame is taken as eliminated from (treated as a 'blank' in) the area frame.

### 3. *Common problems with sampling frames*

#### 3.1. *Problems with area frames*

750. Area frames have the advantage over list frames in that areas as units are larger, more easily identified and more stable than dwellings, households, establishments or persons which appear as units in typical lists. Nevertheless, area-based frames also suffer from coverage and related errors. These usually arise from failures to define and identify physical boundaries of the area units correctly, and poor quality of the lists of the ultimate units.

751. For population-based surveys, it is usually possible to assess whether large under-coverage has occurred (which appears to be an extremely common problem) by comparing estimates of population aggregates obtained directly from the survey (the so-called 'simple unbiased estimates') with some more reliable external information on population size. For other types of surveys, however, such external information is often not available.

752. Common imperfections of area frames include the following.

- (a) The failure to cover the population of interest exhaustively. In a number of countries with inadequate cartographic work, the available frames are actually composed of lists of localities rather than of proper aerial units; scattered populations outside the listed localities may not be covered in such frames.
- (b) Under-coverage due to outdated frame of area units. The problem of under-coverage can become much more serious as the frame becomes outdated with time.
- (c) Errors and changes in area boundaries. These may arise from errors in identification of the boundaries and boundary changes after the frame was prepared. The unit boundaries as defined in maps or descriptions may differ from the boundaries of units on which other relevant information is available in the frame (information such as size and density of the population), or from boundaries of the actual sampling units.
- (d) Inappropriate type and size of units. The available units may be too large, too small, or too variable in size to serve as efficient sampling units.

- (e) Lack of auxiliary information. Information on size and other characteristics of the units, required for efficient sample selection, may be inaccurate or simply unavailable. Area frames for household surveys based on the population census often suffer from outdated information on population size of the area units. The requirements and problems of auxiliary information can be much more serious in the case of area frames for surveys of small-scale establishments and other economic units.
- (f) High cost. Area frames are generally expensive to create. Usually the investment is justifiable only when the frame is to be used repeatedly over a relatively long period of time.

753. Serious coverage errors occur in the use of area frames when the boundaries of the enumeration areas (EAs) cannot be clearly identified during the fieldwork for the survey. This is quite a common problem in surveys based on outdated area frames, or when the available maps and descriptions are of inadequate quality. To understand this problem one needs to look at the precise way in which EAs are identified. The following is an illustration of what may happen in practice: in theory, an EA is identified simply by delineating its boundaries on a map. These boundaries are supposed to follow natural features such as streams or roads. But obviously such features are not always available, so the mapper often draws straight lines on the map that have no basis on the ground. S/he is quite likely to draw the line through an inhabited area, reasoning that since no one lives there it does not really matter exactly where the boundary lies. For greater clarity, the mapper is normally instructed to insert on the map the location and name of all settlements (rural) or streets (urban) lying within the EA. The household listers for the survey, who may arrive on the scene a few years later, soon learn that the map boundaries, often straight lines, cannot be located on the ground, and so turn their attention to the names written on the map. Very quickly they come to see the job as that of listing the households in these named places and streets. Obviously, wherever the population is growing this will lead to error. New settlements and new streets will tend to be omitted by the listers. What can be done to reduce this problem?

754. Firstly, it can be helpful to try and isolate the problem. It often happens that the problem with area boundaries is serious only in some domains. For instance, census EAs in rural areas may correspond to individual localities and it may therefore be relatively easy to allocate these to the correct a real unit even when the precise boundaries are not easily identifiable; and in the other domains the problem may exist only in the more densely populated (and more rapidly changing) urban areas.

755. Secondly, it can help to draw on all the information available, not just the maps. Often census maps are accompanied by helpful descriptions. More importantly, census household lists, even if outdated for actually selecting a sample of households, can be extremely useful in defining and identifying the original boundaries of census areas.

756. Next, we can look for different, larger, area units as PSUs in place of individual EAs, units that are likely to have less serious boundary problems. From a given frame of basic area units such as EAs, larger units can be created in two ways: by selecting EAs in

contiguous groups rather than individually from the EA list; or by introducing a higher stage of selection such as taking whole towns or villages, each composed of a number of EAs, as the PSUs. The larger areas can then be sub-sampled as appropriate. For instance, the selected units may be segmented and one or more segments selected per area in an additional sampling stage, or the whole area may be listed for the direct selection of households.

### 3.2. *Problems with list frames*

757. Completeness of the frame is a most critical requirement (and perhaps also the most common problem) of list frames. Occasionally it is also important that the list contains pertinent and accurate information on the size and other characteristics of individual units so as to permit efficient stratification and control of the selection process.

758. Problems can arise in the absence of one-to-one correspondence between listings (which are the units actually subject to the selection process), and the elementary units (obtaining a probability sample of which is the actual objective). The lack of correspondence can arise in several forms.

- (a) Presence of blanks in the list, meaning that some listings represent no real units.
- (b) Clustering of elements, meaning that the same listing may represent more than one unit.
- (c) Duplications, meaning that more than one listing represent the same unit. Sometimes the problem arises from the nature of the frame: as for example in the selection of households from an electoral roll (listing all eligible voters in each household) or from telephone directories; selecting parents from a list of school children; or selecting clients or patients from records of visits to the facility. Much more difficult is the problem of unsystematic duplications in the list, usually resulting from the failure to identify the fact that different listings actually represent the same unit. This can happen, for example, if the same unit is recorded in the list several times with slight differences in name, address or description. In such cases painstaking work to eliminate all duplications in the list may be the only solution.
- (d) Failure to locate units, i.e., the failure to identify which unit(s) a selected listing represents. This is a common problem in the absence of clear and complete description in the frame for identifying units in the field. It can also be caused by insufficient effort during the survey field work. The problem is often confused with that of blanks - units not located being indiscriminately reported as non-existing - which among other things, causes difficulty in correctly computing the response rates actually achieved.

- (e) Change in units and unit characteristics, that is, the unit itself or characteristics of the unit associated with the listing have changed. To overcome the problem resulting from mobility of the ultimate units, many surveys use the “population present” coverage definition. This means taking the sample selected to be a sample of addresses or location, rather than that of particular households or establishments; whoever is found to be present at the selected location is enumerated in the survey. The problem can be more serious in establishment surveys where the sample may have to be finely tuned to produce estimates for each of a very large number of branches of activity (industries), with possibly only a small sample allocated to any one of them. Special procedures are often needed to deal with reclassification of units across industry or economic sector boundaries. Special procedures are also required for dealing with changes in boundaries of units in area frames.
- (f) Under-coverage, i.e., units not represented in the frame. This is the most serious and difficult problem and biases the results of many surveys. No simple or cheap solutions to the problem of under-coverage exist.

759. The lack of one-to-one correspondence between listings in the frame and survey units to be selected affects the probabilities of selection of the units. In more serious situations, this can damage the scientific (probability) nature of the samples which can be drawn.<sup>111</sup>

760. The frames problems increase with the interval between the creation of the frame and conduct of field work for the survey. List frames are generally much less durable than area frames. It is for this reason that the census is normally the source for area-based frames rather than list frames, for sample surveys in the following inter-censal period.

#### **4. *Physical representation and maintenance of the frame***

761. Physical representation of the frame should meet the objectives of easy access, use and manipulation, including the production of summary statistics that can help in sample design and estimation. These objectives are served best if the frame of area units is computerised. Each unit must have a unique identifier, and also information that can be used for its classification, ordering, and reordering, etc., as required for different purposes. It may or may not be reasonable to computerise the lists of households or similar ultimate units. Often a household listing is made just prior to and specifically for the purpose of a particular survey. Computerising the list in such cases would almost certainly not be useful; indeed, it may do harm by distracting from more important tasks in the survey.

762. Maintenance and updating of the frame is a difficult task and cannot be discussed here in any detail. Some general principles may be however be noted.

---

<sup>111</sup> The presence of blanks in the list is an exception to this in that, it does not affect the selection probabilities of the units. However, it can be inconvenient and inefficient.



- (a) There is a clear distinction between the task of updating a frame of the more stable area units, and the lists of households, establishments or similar small and less stable units; the two differ in the amount of resources and frequency of updates required.
- (b) A basic requirement is to develop plans and procedures for frame updating, and allocate resources necessary for the purpose.
- (c) It is a good principle to control the frame updating operations centrally and to standardise the procedures to the extent possible. This applies in particular to decisions about urban-rural and other reclassification of area units, adjustment of area boundaries, or addition, deletion, merging and segmentation of area units, etc. As far as possible, final decisions on such matters should not be left to the field workers involved.
- (d) Full use should be made of the opportunity offered by the population census in undertaking major periodic updating of the frames for surveys following the census.

763. Updating the frame of area units, covering the whole country, is clearly a major undertaking. Its main source is the decennial population census. Creating the basis for updating area frames for inter-censal household surveys (and possibly also for surveys of small establishments based on area samples) should be organised as an important objective of the census. Between censuses, area frame updating is best kept to a minimum, for example once at the mid-censal period, and confined to some specially fast growing peri-urban areas or other well-defined strata where the updating is both important and feasible. In addition, there are certain improvements which can be confined to only the selected units, such as improving the area maps, clarifying the boundaries prior to listing of households within sample areas, segmentation of sample PSUs to permit further stages of sampling, compiling information which can be useful in the planning and control of field operations. This work may be done in the areas selected for a particular survey, or on a larger 'master sample', which serves a number of surveys.

764. By contrast, lists of household (or other ultimate units) need to be updated or redone much more frequently, usually specifically for and immediately prior to each major survey. Often this is feasible since the listing is confined to the ultimate area units actually selected into the sample. It is highly desirable that operationally the listing is kept separate from the final sample selection and the main fieldwork interviewing - otherwise serious biases can result.

765. Listing is an expensive operation, and often it is tempting to use pre-existing lists (available from other sources) or old lists (prepared for earlier surveys), or use the same lists for a number of surveys. The last mentioned is clearly a desirable option, but its feasibility depends on the relationship in design and timing of the surveys concerned. Usable pre-existing lists from administrative and other sources are hardly ever available in developing countries, at least for surveys aimed at covering the general population.

766. In relation to using old lists prepared for earlier surveys or censuses, the crucial

question is: how old can the lists be before the sampling biases become unacceptably large? In the absence of sufficient empirical information and diversity of the conditions encountered, only rough indications can be mentioned. When the subject matter is closely linked to family formation and current economic activity (as in many demographic and labour force surveys), it is highly desirable to keep the time-lag between listing and interviewing to a minimum, not exceeding 6 months in any case. For some other surveys, the maximum tolerable gap can be somewhat larger: less so in urban areas where the population tends to be more mobile, and possibly longer in rural areas; less so for lists of households, and somewhat longer when more stable structural units like dwellings or addresses are used.

## 5. *Master samples*

767. In any survey with a multistage design, each stage of the sampling process involves the task of frame preparation and sample selection, until finally a sample of the required ultimate, or lowest-stage, units is obtained. For economy and convenience, one or more stages of this task may be combined or shared among several surveys.

768. The sample resulting from the shared stages is called a master sample. For example, in a three-stage design with EAs, there are segments (smaller areas within EAs), and households as the first-, second-, and third-stage units respectively. A large sample of EAs may be selected, each sample EA divided into segments, and some or all of the segments retained in the sample. The resulting large sample of segments can serve as a master sample from which smaller samples of segments and households can be drawn for individual surveys. Thus the task of frame preparation and sample selection up to the stage of segments is shared among different surveys using the master sample.

769. The master sample therefore provides a common sample of units up to a certain stage, from which further sampling can be done to serve the needs of individual surveys. Its purposes are to minimise the costs of developing sampling frames and materials and of sample design and selection, by combining these operations for different surveys; to facilitate substantive as well as operational linkages between different surveys and survey rounds; and to facilitate, as well as to restrict and control when necessary, the drawing of multiple samples for various surveys from the same frame.

770. At a minimum, the concept of the master sample involves the construction of a common (and sufficiently large) sample of primary sampling units (PSU's) that have been selected for use for several surveys. For a particular survey, some or all of these PSUs may be taken to continue the sampling process to lower stages as required for the survey. The primary units initially selected may be subject to further subsampling for individual surveys, and the subsamples drawn for different surveys may be independent or related, distinct or overlapping. In this example the task of constructing the primary frame and selecting an initial sample of primary units is common to all surveys using the

master sample. As noted, the initial sample of PSUs for multiple uses is the minimum requirement implied in the concept of a master sample. It may however be useful to carry on the shared sampling process to lower stages, such as to ultimate area units or even to the ultimate sampling units. Sampling for individual surveys will be confined within the lowest stage of units obtained in the master sample.

771. The objectives of using the master sample approach include the following:

- (a) to economise, by sharing between different surveys, on costs of developing and maintaining sampling frames and materials, and costs of sample design and selection;
- (b) to simplify the technical process of drawing individual samples;
- (c) to facilitate substantive as well as operational linkages between different surveys, in particular successive rounds of a continuing survey; and
- (d) to facilitate, as well as restrict and control as necessary, the drawing of multiple samples for various surveys from the same frame.

772. The saving of cost is achieved to the extent that office and field operations necessary to obtain samples for different surveys are combined in the form of a master sample. The saving is the greatest in the situation when the common sample of PSUs (and other higher stage units) is used repeatedly and over an extended period of time, and especially if considerable fieldwork is required after selecting the areas to update their maps, descriptions and boundaries, or segment them to create more suitable units.

773. In practice master samples also have their limitations or disadvantages:

- (a) The saving in cost can be negligible when the master sample concept cannot be extended to lower stages of sampling, where the units involved are less stable and the corresponding frames or lists need frequent updating. A case in point is the list of households that often need to be prepared or updated for individual surveys.
- (b) Reasonable saving can be obtained only if the master sample is used for more than one, preferably several surveys. Effective use of a master sample requires long-range planning and determining the size and other basic design parameters of the surveys planned for its use. Such planning is not achieved in the circumstances of many countries. In certain cases the investment made in constructing the master sample can actually be wasteful, if for instance the planned surveys which were to use it fail to materialise, are radically changed, or if the master sample turns out to be unsuitable/inadequate for other reasons.
- (c) Another basic problem can be the lack of flexibility in designing individual surveys to fit a common master sample. The constraints and requirements (concerning sample size, clustering, stratification, survey timing, etc.), imposed by the available master sample can be difficult to accommodate and result in loss of efficiency of the individual surveys.

- (d) Drawing subsamples from a master sample and correctly computing the selection probabilities of the units selected is not always a straightforward operation. It requires continuity and detailed and accurate maintenance of documentation on master sample design and use.

774. In conclusion, although the use of master samples is not necessarily advantageous or easy in all circumstances, in many situations, a properly planned and maintained master sample can have significant advantages of convenience, economy and control. The instrument can be applied flexibly: it is not necessary that a single master sample should meet all requirements of all surveys, or that a uniform sampling scheme be used for diverse surveys.

*5.1. Possible relationship to a large sample attached to the population census*

775. Often there is a close relationship between the large sample attached to the census for supplementary data collection and the master sample required for subsequent surveys. Hence a close relationship can exist between the designs of these two. For instance, if special attention and resources have been devoted to staff deployment and training or to the development of maps, other materials, and data for the areas included in the attached sample, it will be efficient to ensure that the master sample overlaps with the attached sample as much as possible; and the design of the latter will have to take this requirement into consideration.

776. Sometimes the attached sample can itself serve as a master sample for subsequent surveys — in which case it should be designed as such, with replications and other features that will facilitate sub-sampling from it at a later stage. In general, however, because of their differing objectives, the two sample systems may not fully coincide. For example, to provide acceptable estimates for the whole country, as well as for numerous large and small census domains, the attached sample may have to involve greatly varying sampling rates among domains. For ordinary sample surveys, however, the reporting domains will be typically larger, fewer, and probably with less extreme variations in size. Consequently, more uniform sampling rates will be appropriate for the master sample. For similar reasons, it is likely that a master sample may not have to be as large as the attached sample, and therefore it may not be necessary to undertake the taxing and expensive operations of maintaining and updating the full attached sample for future use. Both of these examples imply that in many situations the master sample itself requires sampling, often at variable rates, of areas in the attached sample; and this requirement too has to be considered in the design of the attached sample.

777. Conversely, the master sample may be especially intensive or large (for instance, to serve intensive cost-of-living surveys in a few urban centres, or surveys to monitor the effects of intensive programmes confined to particular areas) and therefore samples have to extend beyond the areas included in the attached sample. Whether either or both of the sample systems should or even can be modified to minimise their incompatibility depends on circumstances.

## **C Census-based area frames for labour force and other household surveys**

778. The nature and requirements of area frames can differ considerably between population-based surveys, such as on the labour force, and surveys of other types of units such as economic establishments. This section describes the more specific technical aspects of using the population census as a source of frame for the labour force survey and other population-based household surveys. Special considerations in the case of economic surveys of small-scale establishments will be taken up in the next chapter.

779. The construction of sampling frames for household surveys from the population census is determined by the types, size and structure of the required samples in the post-censal period. Especially, in the case of continuous surveys or periodically repeated such as the LFS, it is also necessary to take into account the sample patterns over time (relation between samples over survey rounds and across related surveys panel designs etc.).

780. The use of the population census as a frame for population-based surveys involves at least four crucial steps:

- (a) The construction of an area frame of census enumeration areas (EA's), covering the whole country and summarising the census information to the block level;
- (b) Possible reconstruction of area units which can more appropriately serve as the primary sampling units (PSUs) for the surveys, in so far as they differ from the original census enumeration areas (EAs);
- (c) Stratification of the area units (specifically, the PSUs), making use of the information available in the frame;
- (d) The selection of a 'master sample' of the area units, which can meet the common requirements of diverse surveys for a number of years, and can provide a framework for the integration of survey designs and operations.

781. Additional steps are necessarily involved in the use of the frame of census EAs or other types of area units constructed from those for individual surveys, such as: area sampling for specific surveys; listing of households or other units in the areas selected for specific survey or surveys; and selection of the final sample of households or other units for enumeration. In the following, the three steps are discussed further in turn.

### **1. *The frame of census EAs***

782. Among the above, the first step is the most critical and important. This in fact comprises the first step in the processing of census data, and is therefore crucial in its own right. Summarising the census information to the EA level will give marginal distributions of characteristics enumerated in the census, classified by geographical/administrative divisions, against which more detailed tabulations from the census can be checked. And it will also provide an integrated frame for household – and

possibly other types of – surveys. This frame involves a complete list of Population Census EA's covering the whole country, with the various population characteristics aggregated to the EA level. Ideally, most of the items in the census questionnaire should also be aggregated to the EA level. At a minimum the following types of items will be useful for constructing a sampling frame (see also paragraph 823 of Chapter 15):

- (a) Geographic and administrative information: province, district, locality etc.
- (b) Characteristics of the locality: apart from urban-rural classification, this may include information on the size and type of locality in which the EA lies, subdivision or sector within large cities, identification of rural areas with urban characteristics, etc. (In principle, some of this information may come from sources other than the census, such as administrative records.)
- (c) Number of households and number of persons disaggregated by sex in the EA;
- (d) Number of persons by level of education, disaggregated by sex, in the EA;
- (e) Breakdown of the population by activity status and sex, separating out persons below the working age;
- (f) If available, classification of the working population, i.e., number of persons (women and men) by main sectors (e.g., agriculture versus non-agriculture; or agriculture, industry and services, etc.), and sub-sectors;
- (g) Classification of the working population by status in employment and by occupation.

## **2. *Defining primary sampling units***

783. The second step may involve combining or segmentation of original census EAs to create area units more appropriate for efficient sample design and implementation. Normally, the first choice will be to use the existing census area as they are or at least to confine their restructuring to a small part of the population, as far as possible. This is because of the high cost involved in such an operation for the whole population.

784. The type of units chosen to serve as the PSUs can have a great influence on the quality and cost of the survey. For an area sample, the units need to be well defined, with clear boundaries, good maps and descriptions for identification and demarcation, and up-to-date information on size and characteristics. The areas should cover the survey population exhaustively and without overlaps. Their stability over time is another important requirement, especially if their use must extend over a long period of time. The PSUs should be of an "appropriate" size, in line with the arrangement and cost-structure of the survey's data collection operation. If the units are too large, it may not be possible to include a sufficient number of them to obtain a good spread of the sample; furthermore, the costs of listing, sub-sampling and data collection within big units may also become excessive. On the other hand, if the units are too small and compact, it may be difficult to ensure sufficient spread within the units to obtain an efficient sample. Small area units also tend to lack clear boundaries and stability over time.

785. What is the "appropriate size" of units to serve as the PSUs depends upon circumstances and objectives of the survey. Various practical considerations, of cost, quality control, administration, availability of the frame for sample selection, and efficiency of the resulting design, need to be considered simultaneously. A thorough understanding of the diverse considerations can be obtained only on the basis of a good knowledge of sampling theory and plenty of practice. At this stage, it is instructive to note that several patterns can be identified from national survey practice in different countries, in particular in relation to labour force surveys:

- (a) At one extreme, some surveys (mostly in developed countries) use very large PSUs. The total number of units in the population is often not large and only a small number of them are selected for the sample. Each unit represents a substantial investment in mapping, listing, recruiting and stationing field staff. Often one or more enumerators are assigned to work exclusively in each area. Therefore, while little or no travel between PSUs may be involved, extensive travel and mobility is often necessary within each (relatively large) area, requiring cheap and convenient transport facilities (such as personal car) at the local level. For example, some surveys in the United States use fewer than 100 counties as PSUs, a county being a large unit extending 50-100 km across. Within each selected PSU, the sample of smaller areas and households is, of course, widely dispersed.
- (b) At another level, many samples are based on numerous very small PSUs. Each PSU may be a small cluster of households (say 5-15 households), some or all of which are taken into the sample. Such a system may be suitable in densely populated urban areas where lists of housing units and/or very detailed maps of small area segments are available, and where, because of well-developed transport facilities and short distances involved, travel between units presents no particular problems. Several examples of this type exist in household surveys in developed countries, as well in urban areas in some developing countries.
- (c) On the other extreme, is the use of direct sampling of addresses, households or persons for their labour force surveys (an approach, which a number of countries are increasingly adopting). This trend is primarily the result of increasing use of telephone interviewing not requiring person visits to households. Increasing cost of interviewer time and the drastic reduction of transport cost (virtually zero, with telephone interviewing), also reinforce this tendency.

786. Neither of these two extreme variants is common in surveys in still many developing countries because the basic conditions for their efficient operation are not met. In many developing countries, while there is a sufficiently developed network of major trunk roads to facilitate travel between areas, local travel can be extremely difficult and time consuming, even if the physical distances involved are short. This diminishes the potential cost-saving benefits of employing very large and extensive PSU, since locally recruited enumerators would have a more difficult time to cover the units in their

area, relative to those from other areas served by the more developed trunk road network.

787. On the other hand, the use of numerous small PSUs (and even more so, of completely unclustered samples) is also precluded in the absence of good maps and other materials to define suitable small area units; the number of PSUs also has to be kept limited to control travel and supervision costs. Indeed, a common requirement in the choice of the design is to ensure that each PSU yields a large enough sample to keep the enumerators occupied for a sufficient length of time (say a few days) in each area.

788. It is for such reasons that so many surveys in developing countries use census EAs as the PSUs, an EA typically consisting of 100-300 households. Nevertheless, there are examples in developing countries of surveys using a small number of very large units (such as whole provinces, counties or cities) as PSUs. Often the reason is some perceived administrative convenience of this type of design. It is doubtful whether the choice of such large units as PSUs is actually justified or efficient. Examples are the use of provinces in some surveys in Thailand, and the use of counties as PSUs in surveys in China. Of course, there are situations in which administrative, supervision or travel considerations dictate the use of fewer and larger PSUs than typical census EAs, but still not as few and as large as whole provinces, districts or counties. Examples are the use of 'intermediate level' units such as communes or townships in China, tehsils in India, kacamaten (sub-districts) in Indonesia, or amphoe (communes) in Thailand in surveys requiring particularly intensive data collection and supervision activities.

789. It is important to understand the possible effect of using very large areas as PSUs. Apart from considerations of sampling efficiency, there is an added reason for the undesirability of using very large and few areas (especially large administrative areas such as provinces or whole districts or counties) as PSUs. This is because often such large units represent collections of different types of areas (or other types of units), which need to be separately represented in the sample proportionately or in some other controlled way. However, the use of large areas, mixing different types of units, as the PSUs does not readily permit control over the representation of these different types of units into the sample. An important example is the mixing of urban and rural areas in big administrative units. If the latter are used as PSUs, and the total number to be included in the sample is small and if stratification by the urban-rural composition of the areas cannot be used in their selection, then it cannot be ensured that the urban-rural composition of the resulting sample reflects that in the population. This can have serious implications for any survey, but especially in a situation where there are likely to be very large urban-rural differentials in many of the indicators to be measured.

790. The main implication of the above discussion is that in the circumstances of many developing countries, census enumeration areas do constitute basically appropriate primary sampling units for labour force and similar household surveys. However, the provision of suitable units for sample surveys is by no means the only consideration in the choice and construction of census EAs (the first objective in a census has always been to provide a complete coverage of the population and facilitate practical implementation and control of the census operations). Hence it can be expected that some restructuring of



the census EAs may be required to create more suitable PSUs. Grouping of census EAs when larger units are required is relatively easy since such operations can often be carried out in the office, without involving field work. Segmentation of EAs which are too large to serve as PSUs is usually more costly and time consuming, and therefore has to be undertaken more selectively.

### 3. *Construction of strata*

791. The third step – stratification – is important for the efficiency of the designs based on the Population Census EA frame. Stratification is generally a very effective way of improving efficiency in multi-stage (clustered) designs of the type in many surveys, which may be subject to large design effects. A common system of stratification for all household surveys will be convenient and facilitate their integration. Each different type of surveys (household, economic, agricultural) has its own requirements in relation to stratification. However, this does not preclude the use of a common framework, at least in part.

#### 3.1. *Primary stratification*

792. At the primary (highest) level, stratification must correspond to the reporting domains – e.g. regions or provinces, with appropriate urban-rural classification. An urban-rural classification should correspond to administrative (legal) definitions, wherever possible, because of reporting requirements. It may be more than a simple dichotomy, for instance distinguishing big cities from other urban areas, and rural localities by size category.

793. Often a common system of primary stratification can be used for all types of surveys and censuses: whether household, economic or agricultural. Hence up to this point at least, a common EA frame can be constructed for general use. It may become necessary to distinguish between different types of uses in further, more detailed, stratification.

#### 3.2. *Further stratification*

794. It is often necessary to go beyond the simple geographic stratification. Within each primary stratum as defined above, further stratification is desirable on the basis of characteristics of the PSUs. In creating strata, a number of decisions have to be taken. What characteristics should be used for stratification? How many strata should be created? How should the strata boundaries be determined? Apart from explicit strata, what characteristic or characteristics should be used for ordering the areas to provide implicit stratification when systematic sampling is used?

795. The important point to keep in mind in relation to these questions is that the answer may differ from one part (primary stratum) to another. It is neither necessary nor efficient to seek a uniform system for the whole country, or even a whole province.

### 3.2. *Stratification variables*

796. When a range of EA characteristics are available as potential variables for further stratification, from among those a choice has to be made of the most useful characteristics. The first step will be to examine the range of variation (in relative terms) or the coefficient of variation (CV) of each characteristic, separately for each primary stratum. This involves examining the frequency distributions and CVs where appropriate (the unit of analysis being the census EAs or other type of areas serving as the PSUs), for each important characteristic within each primary stratum. Characteristics with high CVs or ranges of variation are candidates for stratification. Note that it is generally more effective to use a number of relevant characteristics simultaneously, each classified coarsely, than to use fine classes of a single variable. However, it is not useful to include all if the concerned characteristics are highly correlated. For this reason, correlation between all variables with high ranges or coefficients of variation should be examined.

797. In general, the variables applied to are dependent upon the focus of the survey to be undertaken. Therefore, in addition to the geographic or administrative primary stratification the following criteria are likely to be useful for further stratification of household survey frames beyond:

- (a) in urban areas, more detailed information on type and size, apart from geographic location (administrative subdivisions);
- (b) in rural areas, percentage of employed persons in non-agricultural activities, and/or percentage of females literate.

798. More sophisticated approaches - than the above rather simple one of examining potential stratification variables one by one or in pairs or other combinations - are possible, such as using multivariate analyses. However, the suggested simple approach should suffice for many purposes. This is because the primary stratification is already likely to be quite detailed, and the scope for further stratification in the case of household survey frames is limited. Economic and agricultural surveys have their own requirements of controlling the sector of activity in detail.

### 3.3. *Number and size of strata*

799. Within each primary stratum defined above, the number of further (sub) strata to be created depends on (i) the number and distribution of the units (EAs) in the population, and (ii) the likely number of units to be taken into the (master) sample.

800. The basic rule is to try and create strata roughly equal in population size. Hence the strata boundaries will depend on the distribution of units according to the stratification variable(s). However, this basic rule can be modified to take into consideration the likely differences in the sampling rates, which normally will be applied in different domains. It is useful to create smaller (i.e., more numerous) strata in domains, which are likely to be sampled at higher rates.

## **D Selecting a master sample of EA's for household surveys**

801. The fourth step– the construction of a master sample of areas (EAs) – has a number of objectives. It helps to: (i) reduce the cost of frame maintenance; (ii) facilitate sample design and selection for individual surveys; (iii) control sample rotation over time for continuing surveys; and (iv) permit proper integration and separation between the different surveys as required.

802. The idea of a master sample is to select a large sample of areas: a sample which is large enough and is distributed and structured so as to permit easy drawing of representative subsamples as and when required for particular surveys or survey rounds. Because of all these cost, operational and statistical advantages, the construction of a master sample – especially for household surveys – is highly recommended. Separate master samples may be constructed for economic and agricultural surveys and censuses.

803. The guiding principles in constructing a master sample are the following: (i) In size and distribution, the master sample must be able to meet known and unexpected sampling requirements of all the surveys to be conducted using it. (ii) The master sample should be constructed such that samples for individual surveys can be drawn from it quickly and cheaply in the required relationship with each other (such as sample independence, overlap, rotation etc.).

#### **1. *Size of the master sample for household surveys***

804. As noted above, the overall size (the number of PSUs selected) of the master sample should be large enough to meet the requirements of all the samples which may be drawn from it. It is better to err on the liberal side than to fall short: in practical samples it is much easier to reduce the sample size by dropping units, than to increase it by selecting supplements from the population which has already been subject to previous sample selection.

#### **2. *Allocation across domains***

805. The master sample should be allocated among the reporting domains, regions, provinces, etc., to reflect the allocation of the main surveys, which will be based on it. In large countries, and also in other situations requiring sub-national results, this may involve disproportionate allocation across regions or provinces, and increasingly even among districts in view of the increasing need in many countries for decentralised reporting.

#### **3. *Selection of master sample areas***

806. Within strata, master sample areas should normally be selected with probabilities proportional to population of the area (number of persons or households). This is the common scheme in household surveys, and is consistent with the proportionate allocation within provinces. The actual selection probabilities will necessarily vary in line with the

provincial allocation.

#### **4. Replications**

807. It is common and convenient to select the master sample in the form of independent replications, each representative of the whole population. Normally, the replications will be identical in design (stratification, clustering etc.) and size. The purposes of such replicated design include the following:

- (a) to facilitate the selection of samples for individual surveys;
- (b) to permit the required patterns of sample rotation in continuing surveys;
- (c) to permit the desired level of overlap or separation between different surveys;
- (d) to reduce or enhance the sample size at a short notice; and
- (e) to permit flexibility in changing the allocation of the sample across domains.

808. Occasionally, replicated designs have been used to produce quick and approximate estimates of variances, which is possible when a reasonable number of replications is available. In practice, replications can be created simply by systematically dividing a single systematic sample into the required number of parts. Taking for instance the  $K$ -th unit systematically from the master sample will give one of the  $K$  possible replications.

809. The basic idea is that the sample for any particular survey can be constructed simply by taking one or more replications into it; also that the rotation of a continuous survey can be obtained simply by replacing one or more replications from one round to the next; or that a sample allocation different from that in the master sample can be achieved by taking different numbers of replications in different domains, etc. In short, in place of individual PSUs, whole replications – each representative of the entire population or the domain of interest – become the basic “units” for the construction of individual samples of the required size and structure.



## **Chapter 15      The Population Census As A Frame For Economic And Agricultural Surveys And Censuses {ILO COMMENT: THIS CHAPTER TO BE REFERRED TO FAO FOR COMMENT}**

### **A      Introduction**

810. As noted earlier, one of the main uses of the population census for sample surveys is to provide sampling frames for different types of surveys following the census. This is widely recognised in the case of labour force and other household or population-based surveys. However, it is less commonly appreciated that this applies equally to economic surveys, in particular of small-scale establishments, in both the agricultural and non-agricultural sectors. Indeed, this is true also of what are often referred as agricultural and economic ‘censuses’. Outside the population field the term "census" or "sample census" is commonly used to refer to operations that are not in fact a complete enumeration of all units in the study population, but are based on a probability sample. For example, most agricultural "censuses" are conducted on a sample basis, albeit often on a large scale. Similarly, although censuses of economic establishments cover large and medium-size units on a 100 percent basis, it is often reasonable as well as practically unavoidable for them to cover the numerous small establishments only on a sample basis.

811. Surveys and censuses of small-scale economic units have special characteristics and requirements. In most situations, the samples have to be selected in multiple stages using area-based designs. To be efficient, the sampling designs have to take into account the patterns of geographical distribution of the units, which is often very uneven. Special information is required for the construction of such frames. The census of population can be a major source of required information for constructing sampling frames for what may be termed *economic surveys*, as distinct from population-based surveys such as the labour force surveys, provided its provision is carefully taken into account in the design and planning of the census.

812. By economic surveys is meant sample-based enquiries (whether termed ‘surveys’ or ‘censuses’) concerned with the study of characteristics of economic units, such as agricultural holdings, household enterprises, own-account businesses, or other establishments covering diverse types of units, products, activities or sectors of the economy. The reference here is to sample surveys of economic units which, like households, are small-scale, numerous and widely dispersed in the population, and which may be unstable as well as informal (see paragraphs 368-371). Units, which are medium-to-large in size, few in number or are not widely dispersed, may require different approaches, often based on list frames.

813. There are various technical considerations involved in the design and use of the population census as frame for economic and agricultural surveys and censuses. It is useful for clarifying the role of establishment surveys in the provision of information on economic characteristics and activity of the population: identify the basic similarities and

differences of sampling for small-scale economic establishments and typical household surveys such as the labour force survey. It is important to appreciate these similarities and differences in order to understand how the results from the population census may be used for the construction of sampling frames for economic and agricultural surveys and censuses, and what steps must be taken in the design and implementation of the census to ensure that it yields the information required for the construction of such sampling frames.

## **B Censuses and sample surveys of establishments: Similarities with household survey design**

814. The type of sample designs used in 'typical' household surveys provides the point of departure in this discussion of sampling of other small-scale units. Indeed, there may often be a one-to-one correspondence between such economic units and households, and households rather than the economic units as such may directly serve as the ultimate sampling units. Nevertheless, despite much common ground with sampling for population-based household surveys, sampling small-scale economic units involves a number of different and additional considerations.

815. National or otherwise large-scale household surveys are typically based on multi-stage sampling designs. Great variations are of course possible in the type of designs used in different circumstances and for different surveys. Here we consider briefly a most commonly used type of design.

816. A most commonly used type of design is: first a sample of area units is selected in one or more stages, and at the last stage a sample of ultimate units (dwelling, households, persons, etc.), is selected within each sample area. Increasingly – including, and especially in, developing countries – a more or less standard two-stage design is becoming common. In this design the first stage consists of the selection of area units with probability proportional to some measure of size, ( $M_k$ ), such as the estimated number of households or persons in area  $k$  from some past source providing such information for all areas in the sampling frame. At the second stage, ultimate units are selected within each sample area with probability inversely proportional to size. The overall probability of selection of a unit in area  $k$  is:

$$f_k = \left( \frac{a \cdot M_k}{M} \right) \left( \frac{b}{N_k} \right) = f \left( \frac{M_k}{N_k} \right), \quad [1.1]$$

where ( $a$ ,  $b$ ,  $M$  and  $f$ ) are constants. Here  $a$  is the number of areas selected, when  $M$  is the total of  $M_k$  values in the population;  $b$  is the expected number of ultimate units selected per sample area; hence  $(a \cdot b) = n$  is the expected sample size; and  $f$  is a constant defined as:

$$f = \frac{a \cdot b}{M} = \frac{n}{M} \quad [1.2]$$

817. It is common in national household surveys to aim at self-weighting or approximately self-weighting designs. This often applies at least within major geographical domains such as urban-rural or regions of the country. The denominator  $N_k$  may be the same as  $M_k$  (the measure of size used at the first stage), in which case we get a *self-weighting* sample with  $f_k = f = \text{const}$ . Alternatively,  $N_k$  may be the actual size of the area, in which case we get a 'constant take' design, i.e., with a constant number 'b' of ultimate units selected from each sample area irrespective of the size of the area. It is also possible to have  $N_k$  as some alternative measure of size, for instance representing a compromise between the above two designs. In any case,  $M_k$  and  $N_k$  are usually closely related to – and are meant to approximate – the actual size of the area.<sup>112</sup>

818. The selection of ultimate units within each sample area requires a listing of these units. As discussed above existing lists may have to be updated or new lists prepared for the purpose of capturing the current situation. No such lists are required for areas not selected at the first stage. The absence of up-to-date lists of ultimate units for the whole population is a major reason for using area-based multi-stage designs.

819. Surveys of small-scale economic units, such as agricultural holdings or other types of household enterprises, share many of these features. Just like households, such units tend to be numerous and dispersed in the population. Indeed, households themselves may form the ultimate sampling units in such surveys, the economic units of interest coming into the sample through their association with households. Similar to the situation with household surveys, typically no up-to-date lists of small-scale economic units are available for the entire population. This requires resorting to an *area-based multi-stage design*, just as in the case of a typical household survey as outlined earlier.

## **C Special features of sampling for economic surveys**

820. However, despite similarities noted above, there are certain major differences in the design requirements of population-based household surveys and surveys of small-scale (often household-based) economic units. These arise from differences in the type and distribution of the units and in the reporting requirements. The main difference is that economic surveys generally require major departures from self-weighting designs.

### **1. Heterogeneity**

821. Household surveys are generally designed to cover the entire population uniformly. Different subgroups (such as households by size and type, age and sex groups in the population, social classes, etc.), are often important analysis and reporting categories, but (except possibly for geographical classes) are rarely distinct design domains. By contrast, economic units are characterised by their heterogeneity with respect to characteristics with much more uneven spatial distribution. The population comprises of multiple 'sectors', often with great differences in the number, distribution,

---

<sup>112</sup> Throughout, 'size' refers to the number of ultimate units in the area, not to its physical size.



size and other characteristics of the units in different sectors – representing different types of economic activities to be captured in the survey, possibly using different questionnaires and even different data collection methodologies. Separate and detailed reporting by sector tends to be a much more fundamental requirement, than it is in the case for different population subgroups in household surveys. The economic sectors can, and often do, differ greatly in size (number of units in the population) and in sample size (precision) requirements, and hence in the required sampling rates. Therefore it is necessary to treat them not only as separate analysis and reporting categories, but also as *distinct design domains*.

## **2. *Uneven distribution***

822. These aspects are accentuated by uneven geographical distribution of economic units of different types, in terms of relevant characteristics. Normally, different economic sectors to be covered in the same survey are distributed very differently across the population: varying from (i) a few areas having high concentration of some sectors, to (ii) most areas having a representation of some sectors (i.e., sectors widely dispersed across the geographical areas), but with (iii) many sectors in the intermediated or mixed situation, which are neither highly concentrated nor widely and uniformly dispersed. These patterns of geographical distribution have to be captured in the sampling design. True, population subgroups of interest in household surveys can also differ in their distribution (as in the typology of geographical, 'cross' and 'mixed' subclasses defined by Leslie Kish and others<sup>113</sup> in the analysis of design effects), but normally type (ii) rather than (iii) predominate there. By contrast, often situation (iii) predominates in economic surveys; and furthermore, as noted above, such 'mixed' sectors need to be treated as distinct design domains.

## **3. *Sampling versus survey units***

823. There are a number of other factors, which make the design of economic surveys more complex than that of household surveys. Complexity arises from the possibility that the ultimate units used in sample selection may not be of the same type as the units involved in data collection and analysis. The two types of units may lack one-to-one correspondence. For instance, the ultimate sampling units may be (often are) households, each of which may represent none, one or more than one type of economic activity of interest. For instance the same household may undertake different types of economic activities, e.g., agriculture as well as processing and trade of the resulting products. Hence, seen in terms of the ultimate sampling units (households), different sectors (substantive domains) are not disjoint but overlapping. This gives rise to two possible design strategies: (1) An integrated design based on a common sample of households, in which all sectors of activity in which a selected unit is engaged, would be covered simultaneously. (2) Separate sectoral designs, in which the sector populations (in terms of

---

<sup>113</sup> Kish, L., R. M. Groves, and K. Krotki (1997).

the sampled units, households) and hence the samples may overlap. In each sectoral survey, activity of the selected households pertaining only to the sector concerned would be enumerated.

824. Separate surveys are generally more costly and difficult to implement. The overlap between the sectoral samples may be removed by characterising the sampling units (households) in terms of their *predominating* sector. This helps to make the sampling process more manageable. However, this precludes separate sectoral surveys: in so far as the sample for any particular sector is restricted only to the households in which that sector predominates over all other sectors, the coverage of the sector remains incomplete.

#### **4. *Sampling different types of units***

1. In practice it is often costly, difficult and error-prone to identify and separate out the ultimate survey units into different sectors and apply different sampling procedures or rates by sector. Hence it is desirable, as far as possible, to absorb any differences in the sampling requirements by sector at preceding area stage(s) of sampling, so as to avoid having to treat different types of units differently at the ultimate stage of sampling. The cost of such (very desirable) operational simplification is however the increased complexity of the design that this may involve.

#### **D Data required for sample design and selection**

825. It follows from the above that the frame for sampling small-scale economic units has similar but more complex requirements. The required sampling frame will typically consist of area units (which form the primary sampling units (PSUs) in a two-stage design, or the 'ultimate area units' (UAUs) in a design with multiple area stages). These units can be similar, or even the same as those for household surveys: typically the population census enumeration areas (EAs) described in previous chapters.

826. However, in addition to the information on simple population counts by census EA at the time of the population census, information is also required on economic characteristics of each area unit, for instance on the number of economic units of different types and sectors present in the area, the number of persons employed in different sectors and occupations, etc.

827. This information requirement by economic sector of activity is more elaborate than a single measure of population size of the type normally required for probability proportional to size (PPS) sampling in household surveys. It is important, especially in the context of countries with limited administrative sources, that potential sources such as population, agricultural and economic censuses are designed to yield such information, required for efficient design of surveys of small-scale economic units.

**1. Information on economic characteristics of area units and corresponding measures of size**

828. National surveys of small-scale economic units in developing countries surveys have to be typically based on multi-stage sampling designs. Great variations are of course possible in the type of designs used in different circumstances and for different surveys. Here we consider briefly a most commonly used type of design. The objective is to point out the characteristic differences of such economic surveys from household survey such as the labour force survey (LFS).

829. As noted, the population of units in economic surveys comprises a number of 'sectors', such as different types of holdings, agricultural, processing and trading activities, or products. Sample size requirements generally have to be met separately for each sector. The available sampling frame has to consist of area units (which form the primary sampling units (PSUs) in a two-stage design, or the ultimate area units (UAUs) in a design with multiple area stages), for each of which information is required on the expected number of economic units of different types.

830. In essence, the overall selection equation (following equation [1.1] above) is of the form:

$$f_k = f \cdot \left( \frac{M_k}{N_k} \right), \quad [1.3]$$

where  $M_k$  is some measure of size assigned to the area in its PPS selection (i.e., probability proportional to size), and the selection of ultimate units within the area is with probability inversely proportional to the size, i.e., measure  $N_k$  in the frame, assumed to estimate the actual size of the area.

831. The design weight to be applied at the estimation stage is inversely proportional to the overall selection probability:

$$w_k = \frac{1}{f_k} \propto \left( \frac{N_k}{M_k} \right). \quad [1.4]$$

832. The *expected* number of units contributed to the sample by area  $k$  is

$$f_k \cdot N_k = f \cdot M_k. \quad [1.5]$$

Summing over all areas  $k$  in the population in [1.5] gives the total expected sample size

$$n = f \cdot M. \quad [1.5a]$$

The expected number of units of a particular sector  $i$  contributed to the sample by area  $k$  is

$$f_k \cdot N_{i,k} = f \cdot M_k \cdot \left( \frac{N_{i,k}}{N_k} \right) = f \cdot M_k \cdot P_{i,k} \text{ , [1.6]}$$

The basic design problem is to determine the 'modified' size measures  $M_k$  such that the sample size requirements

$$n_i = f \cdot \sum_k (M_k \cdot P_{i,k}) \text{ [1.7]}$$

or in terms of relative quantities more convenient for numerical work

$$\frac{n_i}{n} = \sum_k \left( \frac{M_k}{M} \right) \cdot P_{i,k} \text{ [1.7a]}$$

are satisfied for all sectors  $i$  simultaneously in the most efficient way.

833. In [1.7] the sum is over all areas in the population (and not merely the sample). Note also that the above formulation assumes that at the ultimate sampling stage, units within a sample area are selected at a uniform rate, inversely proportional to  $N_k$ , irrespective of the sector. This is a very desirable feature of the design in practice. It is often costly, difficult and error-prone to identify and separate out the ultimate survey units into different sectors and apply different sampling procedures or rates by sector. The preceding sampling stages are assumed to absorb any difference in the required sampling rates by sector through incorporating those in the definition of the size measures  $M_k$ .

834. The most convenient (but also the most unlikely) situation in the application of [1.7] is when units of different types (sectors) are geographically completely segregated, i.e., when each area contains units belonging to only one particular sector. In reality the situation is more complex because areas generally contain a mixture of units of different types (sectors). Clearly, we should inflate the size measure (and hence inflate the selection probabilities) for areas with proportionately more units from sectors which need to be over-sampled, and vice versa.

## **2. *Collecting information on economic characteristics of population census enumeration areas (EAs)***

835. It is clear that to apply a design like the above, it is necessary to have in the sampling frame information not only on some single (and simple) measure of size such as the number of persons or households in each sample area at some time in the past (such as at the time of the population census) which is related to the actual size of the area at the time of sample selection. It is also necessary for surveys of small-scale economic units to have information on the number of units of different types in each area.

836. At a minimum, the Population Census frame involves a complete list of EAs with the various population characteristics aggregated to the EA level. These include (i) geographic, administrative and other information on characteristics of the area; and (ii)

some demographic and social information on the population in the area. The usefulness of the population census to serve as a frame of area units for economic surveys is greatly enhanced if in addition (iii) information is also collected, and compiled to the EA level, on some basic economic characteristics of the population.

837. This could include, for instance:

- (a) Information on distribution of the population according to broad sectors of activity, which is useful for constructing frames and selecting samples for censuses and surveys on these topics. Even only this type of information can be of great value in increasing the efficiency of the sample designs for economic surveys, which may be constructed from the frame of census enumeration areas.
- (b) Information on the number of economic units by main sector of activity at the time of census in each EA. This type of information is normally not obtainable from the population census without some very careful preparation for the data collection and processing procedures. In particular, questions on either place of work or branch of activity that include the name and address of the employing establishment (see example of questions in Chapter 7) should have been asked in the census and the relevant information processed and tabulated as indicated above. (Questions have been included in the population censuses of some countries, for example Belize 2001 (“Do you own any land for farming?”) that might be useful for this purpose in an agricultural census.)
- (c) Information from past economic and agricultural censuses for the areas that were selected for the latter types of ‘censuses’. Some times, it may also be possible to incorporate into the population census relevant information of this type. A crucial requirement for this is that the economic and agricultural censuses be based on the same system of areas as the population census EAs. It is an objective of good planning of the census system to ensure this. This has been done successfully in a number of countries, notably in India and Indonesia among the largest developing countries.<sup>114</sup>

838. The information on economic characteristics for individual areas (such as census EAs) is required for determining the sampling rates needed to obtain the required sample sizes for different sectors of economic activity, for efficient stratification of the area units, and also to determine the probabilities of selection of the areas to meet the requirements for reporting by sector in economic surveys. Such information can also be useful at the stage of estimation from the survey (see Chapter 14). Stratification and variations in sampling rates have to be ‘targeted’, i.e., determined on the basis of composition of the area units in terms of the relative size of the different economic sectors present in the area. Such targeting is required to ensure efficiency and control in the sample design.

---

<sup>114</sup> See for example Murphy and Roy (1970), and Verna (1998).

## **E Primary sampling units and stratification**

839. Because of the great variety of surveys of establishments and other economic units, little can be said in general about the appropriate choice of area units to serve as the PSUs. The census EAs are often quite suitable units for the purpose of such surveys. Sometimes, it is necessary to create larger areas to serve as PSUs so as to 'capture' sufficient numbers of units of interest. This is normally not a major problem since existing area units can be easily combined to define new units for sampling.

840. As to stratification, economic surveys tend to involve units greatly varied in type and size, and hence usually require much more detailed stratification than household surveys. The main stratification required is in terms of the predominating sector of activity in the area. Such information for the purpose of stratification need not be highly accurate or up-to-date, so long as it is reasonably well correlated with the area's current economic characteristics. Hence, information on the number and distribution of establishments by sector of activity collected during census listing can continue to serve a useful purpose for a number of years following the census.



## Chapter 16      Use Of Census Data For The Production Of Survey Estimates

### A      Introduction

841. Several requirements need to be taken into account when population census data are used in the production of estimates from sample surveys.

- (a) A comprehensive survey such as a continuous labour force survey is used to produce a wide range of estimates for different populations at different levels of aggregation, and over different time periods.
- (b) Several types of variables may be involved, such as the rates of labour force participation, unemployment rates, employment by status, occupation and industry. The frequency and precision requirements may differ for different types of variables.
- (c) Estimates are normally required for different types of units, such as individuals, households, families and communities.
- (d) Estimates are required not only at the total (national) level but also for separate geographic and administrative areas, urban and rural classification, and for different subpopulations such as such as demographic classifications.
- (e) Estimates are required for proportions, rates, means and other ratios (where the numerator and the denominator are obtained from the same source), as well as for population aggregates (where the survey results have to be inflated to the population total using information internal or external to the sample survey). Often surveys of moderate size are able to yield estimates of population totals (such as the total number unemployed) with much less accuracy in terms both of variance and bias, than estimates of various types of ratios (such as the unemployment rate).

842. Furthermore, various types of estimates can be distinguished in terms of the time dimension. These include

- (i.) estimates of current levels for each reporting period (month, quarter, year) of the survey;
- (ii.) estimates averaged or aggregated over longer periods, such as annually from a quarterly survey;
- (iii.) estimates of trends or net change from one period to another; and possibly also
- (iv.) estimates of gross change at the level of individual person or household, such as flows among categories of economic activity status (although such statistics are still rather uncommon), and in addition,
- (v.) more complex 'composite' estimates may be produced.



843. In the production of any of these different types of estimates of population parameters, the sample data need to be appropriately weighted. Most common estimates from a survey such as on the labour force are in the form of weighted ratios. The most basic forms in which the sample weights appear in estimation are as follows:

$$r = \frac{\sum_i w_i \cdot y_i}{\sum_i w_i \cdot x_i} = \frac{y}{x} \quad [2.1]$$

where  $r$  is the ratio of two weighted sample aggregates  $y$  and  $x$ , with each of these aggregates formed by summing up the individual sample values (on two survey variables, for instance,  $y_i$  and  $x_i$ ), for each sample unit  $i$  weighted by its associated weight  $w_i$ .

844. The estimate  $r$  may be the ratio of two substantive variables, such as wage rate ( $r$ ) as the ratio of (weighted aggregate of) wages received ( $y$ ) to the (weighted aggregate of) number of hours worked ( $x$ ). Many labour force survey (LFS) statistics, such as means, proportions or percentages, are merely simpler forms of this. For instance, average hours worked ( $r$ ) is the ratio of the (weighted) aggregate of hours worked to the weighted number of workers; the denominator in this case being simply the weighted count of cases (meaning that  $x_i$  in the denominator is identically equal to 1 for all sample cases involved in the estimation). Similarly, in computing the unemployment rate,  $y_i$  is a simple dichotomy (unemployed=1, employed=0), with  $x_i=1$  as simply the count for all persons in the labour force for which the unemployment rate is being computed.

845. Labour force and other surveys are often also used to estimate population totals, such as the total number of persons unemployed. These generally take the form:

$$\hat{Y} = r \cdot X = \frac{y}{x} \cdot X = \left( \frac{\sum_i w_i \cdot y_i}{\sum_i w_i \cdot x_i} \right) \cdot X \quad [2.2]$$

where  $X$  is an estimate of the denominator ( $x$ ) from some more reliable source external to the survey. (In principle, the sample quantity  $y$  also provides an estimate of the above, but a 'ratio estimate' in the above form is generally more precise – see below.)

846. The population census is often a valuable basis for the information required to compute appropriate sample weights  $w_i$ , and even more so for obtaining the external 'inflation factor'  $X$ . The following sections discuss some of the technical issues involved in weighting and estimation from surveys using external information from the census and other sources.

## **B Weighting of sample data**

847. The objective of weighting sample data is to improve the representativeness of the sample in terms of the size, distribution and characteristics of the study population. For example, when sample units have been selected with differing probabilities, it is common to weight the results inversely proportional to the unit selection probabilities so as to

reflect the actual situation in the population. In a survey selected from a good frame and well implemented with high response rates, application of the above mentioned 'design weights' is all that may be required.

848. In practice, however, the situation is usually more complicated because of shortcomings in the selection and implementation of the sample (which introduce biases in the results), and also because of the need for (and possibilities of) introducing improved estimation procedures to reduce variances. The need for more complicated estimation procedures tends to be greater in surveys suffering from high non-response and coverage errors; inconsistencies in the definitions of units used at different stages in the survey operation; departures from representative (probability) sampling; small sample sizes; and when estimates for many separate subpopulations have to be produced. The need (and the opportunity) is also greater in the presence of more extensive and more reliable external information for the purpose.

849. Issues involved in sample weighting can be complex, and the 'best' solution may be situation-specific, depending on the nature of the data at hand, sources of error which need to be controlled, knowledge of the specific circumstances and limitations of the survey, and existing practices and preferences of the survey organisation. Nevertheless, there are major advantages in following certain basic standards and a systematic approach, as outlined below.

#### *1. Source of information*

850. In applying the weights, the best use has to be made of the information available, both internal to the sample and from external sources. The primary role is given to information internal to the survey; external information is introduced to the extent judged useful for further improving the representativeness of the sample.

851. Five types of information sources may be noted, which can be used in a systematic manner to apply weights in a step-by-step procedure:

- (a) Sample design, i.e., the design probabilities of selection of each household.
- (b) The sampling frame, which may provide additional information on sample areas and on all responding and non-responding households.
- (c) Sample implementation, i.e., response rates and information on non-respondents.
- (d) Other, significantly larger survey(s) with better coverage, higher response rates and more reliable information on certain characteristics of households and/or persons. Indeed the labour force survey itself is often a good source of information for weighting in order to improve the representativeness of more complex and difficult surveys with smaller samples and/or poorer response rates.
- (e) Population projections from the census (and/or current registers where available) providing information on characteristics of the population and their distribution.

852. When the same or similar information is available from more than one source, priority should first be given to the source internal to the survey; for instance weighting to compensate for differences in selection probabilities and known incidence of non-response should always be applied before corrections on the basis of external data are introduced. In using external information, it is necessary to ensure that the information is significantly more reliable than the available information internal to the survey, that the items of information used are comparably defined and measured, and that the coverage and scope are the same. For instance, if the survey is confined to the population residing in private households, external information on the population used for sample weighting should be similarly restricted.

## 2. *A step-by step procedure*

853. For achieving common standards, as well as for clarity and convenience, it is desirable that a step-by-step procedure be adopted, which separates out the different aspects of weighting. As a rule, each step should be applied separately so that its contribution to the final weights can be identified. For a household survey of the general population, for instance, the basic steps involve the application of following weighting factors in sequence, which are described in turn in the next sections.

- Design weights.
- Non-response weights.
- Weights correcting the distribution of units in the sample, to agree with more reliable external information
- Where applicable, appropriate ‘inflation’ of the sample estimates to agree with external control totals, such as the current size of the total population.

## C **Design weights**

854. Each household or person in the sample is weighted in inverse proportion to the probability with which it was selected. The design weights are introduced to compensate for differences in the probabilities of selection into the sample, the weight given to each sample unit being inversely proportional to its probability of selection.

855. For instance, with  $p_i$  as the unit’s overall probability of selection into the sample, the design weight may be computed as:

$$w_i = \frac{1}{p_i} \cdot \left( \frac{n}{\sum (1/p_i)} \right) \quad [2.3]$$

where the summation is over  $n$  units successfully enumerated in the survey. (Incidentally, the above form has the convenient property  $\sum w_i = n$ , meaning that the weights have been scaled to average 1.0 per unit enumerated in the sample. Clearly, relative rather than

absolute values of the selection probabilities matter. With multi-stage sampling design, the reference is to the overall selection probabilities of the ultimate units; other aspects of the sample structure such as clustering and stratification do not enter the equation.)

856. With probability sampling (i.e., units selected with known non-zero probabilities) the design weights can be computed (in principle and mostly also in practice) from information on sample design and outcome internal to the survey, without requiring external information from census or other sources. However, in certain circumstances it is useful (and necessary) to incorporate into the design weights a correction for known exclusion or gross under-coverage of some parts of the study population, which may have occurred because of defects in the sampling frame or other reasons to do with the procedures of sample selection and implementation. One way to apply such a correction would be to reduce the design probabilities of selection (i.e., to increase the weights correspondingly) in proportion to the coverage rates in the affected domains, or to incorporate a compensation in other covered domains similar to the one(s) excluded. For instance, in countries where certain remote islands or areas have been excluded, the weights for certain other, perhaps similarly remote but still included, areas may be increased in proportion to the populations involved.

#### **D Non-response weights**

857. These are introduced to reduce the effect of differences in response rates achieved in different parts of the sample. These weights are based on characteristics that are known for responding as well as non-responding units. Weighting for non-response is particularly important when rates of non-response are high and generally variable from one part of the population to another. Such defects make it necessary to weight the data for non-response to correct for at least gross distortions in the distribution of the achieved sample according to various characteristics of the analysis units. Of course, weighting for non-response only corrects for distortions resulting from differential non-response in terms of the variables used in the classification. It cannot take into account the effect of the absolute levels of non-response - only of the relative levels across the classes.

##### ***1. Defining weighting classes***

858. Essentially, weighting for non-response involves estimating units' "propensity to respond" as a function of unit characteristics, and weighting the responding units inversely to that propensity. In a simpler and commonly use form, weighting for non-response involves the division of the sample into certain appropriate 'weighting classes', and within each weighting class, the weighting-up of the responding units in inverse proportion to the response rate, so as to 'make up' for the non-responding cases in that class. Specifically, the following steps are involved:

- (a) Division of the sample units into groups or ‘weighting classes’, showing the number of units (households, persons etc) selected and the number of interviews completed in each class. Depending on the number of control variables involved, several parallel divisions may be involved, each providing marginal distribution of the whole sample according to some classification variable or some cross-classification of a set of variables, for instance marginal distributions by geographic location or stratum, by some other classification of sample areas, by tenure, household type and size, or other characteristics of the households.
- (b) Computation of the response rate for each category in the classification.
- (c) Assigning a uniform weight to all units in a category, in inverse proportion to the response rate in the category.

859. The effectiveness of the procedure depends on the extent to which non-responding units within each class are similar to the responding units in that class in terms of all the other variables not used in the classification. Differences in unit characteristics and in response rates should be maximised across the weighting classes chosen.

860. Given this requirement, it is still necessary to choose the appropriate number and size of classes to be used for this purpose. The use of many weighting classes has the possible advantage of reducing non-response bias by creating relatively small and homogeneous weighting categories within which characteristics of respondents and non-respondents can be assumed to be similar. On the other hand, the use of many small weighting classes can result in the application of large and variable weights that can greatly increase the variance of the sample estimates. A compromise is therefore required. The choice will depend on how variable the response rates are across different parts of the sample, and how these variations are related to the characteristics of units.

861. When a number of classification variables are available, each variable (or each cross-classification of two or more variables) may be used to divide the whole sample into parallel sets of weighting classes. Each set will provide the distribution, according to the classification variable(s) concerned, of the numbers of units selected and the numbers successfully interviewed on the basis of which response rates can be computed in each category of the classification.

## **2. *Data sources***

862. It is obvious that weighting classes can be defined on the basis of only those characteristics that are available for both the responding and non-responding units. While the non-responding cases are identified from the survey records, normally only limited information is available from the survey on non-responding units. External sources, including the population census, are often the primary source of information on these units.

863. In a multi-stage sample of households, for instance, some of the characteristics

involved may pertain to sample areas, and others to individual households and persons, including both responding and non-responding units.

864. Area-level characteristics refer to characteristics relating to areas or other aggregates, such as geographic location (administrative divisions), type of place of residence (e.g., an urban-rural classification), and various socio-economic characteristics of the areas. Some such information is always available from the sampling frame itself (specifically, the geographic location and other information used for stratification of sampling areas). Additional information may also come from local area statistics from the census or administrative sources.

865. Other characteristics relate to individual households or persons, such as household size and type, tenure of accommodation, income, socio-economic status and other characteristics of the household head or reference person, and demographic characteristics, activity status and other characteristics of individual household members. Several characteristics sources of such information are possible.

- (a) Where the survey records can be linked at the individual level with administrative sources or the population census, the latter can provide information on characteristics of non-responding units in the survey. When such external information is used for the present purpose, the same source should be used for the classification of both non-responding and responding households, even if for the latter the same type of information is also available from the survey itself. This is necessary to retain consistency in the classification.
- (b) In situations where the sample is drawn from lists which include relevant information for the classification of the selected units, the information in the frame can be used to provide common classifications for responding and non-responding units. At a minimum, geographical location and other information used for stratification of the sample areas must be used to appropriately define weighting classes for non-response adjustment in household surveys. Other classifications based on additional information of the types described below should be created where possible.
- (c) In cases where the current sample is based on some previous survey or the census, these can provide information on non-respondents to the current survey.
- (d) In complex surveys prone to high rates of non-response in particular, it is desirable to make a special effort to collect at least a few basic items of information on each unit selected into the sample, irrespective of whether or not the unit is successfully enumerated in the main survey.

## **E Correcting the sample distribution with more reliable external information**

866. After the sample data have been adjusted for differential sampling probabilities and response rates, the distribution of the sample according to the number and characteristics of the units will usually still differ from the same distributions available

from more reliable external sources such as the population census, projections, registers or other large-scale surveys. Normally, the precision of the estimates is improved by further weighting the sample data so as to make the sample distributions agree with the external information.

867. In a household survey, the distributions in various categories may involve numbers of households and/or aggregates of some variables measured on each household. With a survey involving households as well as persons as the units of data collection and analysis, it is necessary to control the sample distributions in terms of both types of units.

868. Several considerations are highly relevant in using auxiliary information from a past census in the production of current estimates from surveys. Obviously, the external information has to be sufficiently accurate and up-to-date, so that its imposition on the sample actually results in improving of the latter. Another basic requirement is that the characteristic(s) used for matching the sample distribution to the external control distribution be the same in the two sources, i.e., defined and measured in exactly the same way. This cannot always be taken for granted in using the census data in the weighting of sample surveys. Furthermore, when more than one external sources are used (as is normally the case in practice), the information from the various sources must be consistent. Re-weighting on the basis of external control distributions does not require matching of the sample and the external source at the level of individual households or persons. The weighting adjustments are made on the basis of comparison of sample and external distributions at the aggregate level.

869. The following two points are relevant in relation to the use of population census data for the purpose.

- (a) Tabulated census data often provide distributions by one or, at the most, a small number of individual characteristics at a time, rather than by a wide range of variables; yet in most situations the distribution by only a single characteristic is not sufficient. On the other hand, while it is desirable to control all-important characteristics simultaneously, having too many controls results in small adjustment cells and large variations in the resulting weights. It is also preferable to control for a number of marginal distributions simultaneously, rather than to use a single variable with very detailed classification, or a full 'post stratification' involving many cells in a cross-tabulation of several variables. Also, the census data can be out-of-date; however, this tends to affect actual levels more than the underlying distributions.

- (b) Another point to note is that external information on the relative distribution of units by characteristics is often more readily available than information on the actual numbers of units in different categories in the population. Also, the relative distributions tends to be less sensitive than absolute numbers to any differences between the survey and other sources in terms of the definitions used.<sup>115</sup>

## F Estimating totals

870. It is statistically convenient to 'scale' the weights computed at each of the preceding steps to average 1.0. Or these can be expressed to any arbitrary scale. The scaling from sample to population aggregates can then be specified in terms of an overall inflation factor, common to all sample units, which inflates the sample numbers or estimated totals to the corresponding aggregates in the population. This does not affect the survey estimation of proportions, means or other ratios, but only the estimation of totals or aggregates. In principle, different inflation factors may be involved in the estimation of aggregates for different types of quantities, and for different types of units such as households and persons. However, as explained below, additional considerations are involved in estimating aggregate population values from sample data.

871. The 'simple unbiased estimation' (i.e., with sample data weighted simply by inverse of sample selection probabilities), while often suitable for means and other ratios, usually needs modification when the objective is to estimate population aggregates. This is especially the case for surveys with a multi-stage design and small sample size. This is because with multi-stage sampling design, the resulting sample size varies at random, and therefore aggregates directly estimated from the survey can have a large sampling variance. The problem is even more serious when estimates are required for population subclasses the selection of which is not explicitly controlled in the multi-stage design.

872. An equally important problem arises from the fact that estimates of aggregates are biased in direct proportion to the magnitude of the coverage and related errors. By contrast, this effect on estimates of proportions, means, other ratios, and more analytic statistics is often less marked.

873. The appropriate procedure for estimating population aggregates is generally as follows. In place of simple inflation of the form

$$\hat{Y} = F \cdot y, \quad [2.4]$$

i.e., inflating the sample aggregate  $y$  by  $F$ , the inverse of the overall sampling fraction,

---

<sup>115</sup> A convenient method of adjusting the sample distribution to a number of external controls simultaneously is the classical Iterative Proportional Fitting or Raking method originally proposed in Deming and Stephan (1940), "On a least square adjustment of a sample frequency table when the expected marginal totals are known", *Annals of Mathematical Statistics*, 11, pp. 427-444. A more accessible source is Deming, W. E. (1943) *Statistical Adjustment of Data*, Chapter VII. New York: John Wiley. Paperback edition of this book has been published by Dover Publications New York, first in 1964.



the required aggregate may be expressed in the form of a ratio-type estimate, shown in equation [8.2] above, namely:

$$\hat{Y}_r = \frac{y}{x} \cdot X \quad , \quad [2.5]$$

where  $y$  and  $x$  are estimated totals from the sample;  $y$  being the variable of interest, and  $x$  an auxiliary variable for which a more reliable population aggregate value  $X$  is available from some external source.

874. The value and applicability of this procedure depends on several factors, and considerations that are highly relevant in using auxiliary information from a past census in the production of current estimates from surveys. First, the correlation coefficient between  $y$  and  $x$  must be positive and preferably large, say greater than 0.6 or 0.7 at least. Secondly,  $X$  should be available with higher precision than the simple estimate of the population aggregate that can be produced directly from the sample itself. Thirdly,  $X$  in the population and  $x$  in the sample should be based on essentially similar measurement on the same population; a difference between the two would introduce a bias into the estimate. This often requires that values of the variable  $x$  for individual units - unless they are simply a count of the cases, as in the case of an ordinary mean - are taken from the external source rather than directly from the measurements in the survey, though of course that must be for the actual units included in the sample.

## Chapter 17      The Census As A Basis For Evaluating Survey Data

875. Another potential use of census data is to serve as a benchmark for the evaluation of results from post-censal sample surveys. Due to differences of timing, content, methodology, practical conditions of data collection, etc., the comparability between the two types of sources are limited, and hence a past census cannot fully serve as a benchmark for more current sample surveys. Nevertheless, the population census can be useful in a number of ways for this purpose. The questions to be addressed are whether and to what extent data on economic characteristics and activity of the population from the labour force and other sample surveys can be evaluated on the basis of information from the population census and projections? It is useful to distinguish between content and coverage aspects of the evaluation.

### A      Survey content

876. Surveys such as the labour force survey (LFS), often contain much more detailed information on economic characteristics and activity of the population than it is possible (or desirable) to collect in a population census. This limits the scope to which census data can be used to evaluate the results of such surveys. It is important to begin by recognising the limitations of census data on activity status and other economic characteristics.

#### 1.      *Limitations of census data on labour force characteristics*

877. The main limitation of the census data result from the fact that, because of its large size, the census can include only a short series of questions on the labour force. Consequently, individuals, especially those at margins of the labour force categories, may not be classified with the same degree of precision as is normally possible in a labour force sample survey.

878. In relation to the correct identification of persons in employment, the main problems are likely to be the treatment of persons temporarily absent from work and the correct application of the concept of “work”. To determine whether a person is employed or is unemployed or economically inactive, a labour force survey normally includes the necessary questions on duration of and reasons for absence from work, whether payment is received, formal job attachment, etc. Typically, a census cannot incorporate such a detailed series of questions, and consequently persons away from work are much more likely to be automatically classified as employed.<sup>116</sup> Similarly, detailed prompts and probes are needed to ensure that all the different types of economic activities are included in the concept of “work” (including for example, volunteer work for profit-making enterprises, water carrying, wood gathering, making handicrafts). Such detailed efforts are not always possible in a population census and persons solely engaged in such

---

<sup>116</sup> See for instance, Carew, J., R. Woods, and B. Brady (1999). *1996 Census: Labour Force Status*. Australia Bureau of Statistics, Census Working Paper 99/2.

activities may be classified incorrectly as not employed.

879. Similar considerations apply to the measurement of unemployment. To measure this variable properly, taking fully into account the recommendations of the International Labour Organization (ILO), requires a fairly elaborate series of well-tested and well-administered questions. A labour force survey of limited size can be expected to incorporate such a series fully, but a population census will normally need to rely on some appropriately simplified version. Future starts (those waiting to start a job) tend to be especially difficult to be classify correctly.

880. Furthermore, the interviewers in a survey can be expected to be better trained to administer the questions in the required form, and are usually allowed more time to probe in-depth when complicated situations so require. By contrast, censuses have to be conducted with more limited resources per unit contacted, and in any case may have to give priority to other aspects, such as getting the basic population counts right. Hence it is not entirely unexpected that censuses often produce different (generally lower) participation, employment and unemployment rates than do many labour force surveys. As a general rule, we can expect that, despite being subject to sampling error, a labour force survey produces more reliable and comparable estimates of activity status and other labour force related characteristics than a population census, at least at the national level and major subnational domains.

881. In comparing census and labour force survey results, it is also important to identify (and adjust for where possible) differences in scope, coverage, treatment of non-response and other differences in the mode and conditions of data collection. For various such reasons, differences observed between the census and the surveys on nominally the same items may be quite expected and legitimate, without necessarily one being ‘better’ than the other. An important consequence of the limits in comparability between survey and census data, even for nominally the same items, should be noted in the context of weighting of survey data on the basis of the census discussed in the previous chapter. In so far as the data are not comparable, it would be a mistake to ‘force’ – as is sometimes done – the survey data to conform to distributions from the census.

## **2. *Potential for comparison with and evaluation of LFS data***

882. Many censuses nevertheless do cover broad aspects the population’s economic characteristics and activity, such as activity status (employment, unemployment, inactivity), status in employment, industry, possibly also occupation, sector of employment, working time (full-time, part-time), etc.

883. A good census can provide a good basis against which the overall plausibility of the survey results can be judged, and at least any gross errors identified. It is always worth checking the results of a survey against the census when available. Furthermore, some of the differences between the two sources have less impact on differentials between population groups than on overall levels. This happens to the extent that

different population groups in each of the two sources are subject to similar measurement biases, thus reducing their impact on the observed differentials between groups. Examples are differences in participation or unemployment rates by age, sex, geographical region, urban-rural classification, etc.

## **B Survey coverage and ‘representativeness’**

884. In contrast to the limited scope for the evaluation of substantive results of sample surveys against the census, information on the size, composition and distribution of the population from the census and projections based on it can provide a solid basis for the evaluation of coverage and representativeness of the sample.

885. It has been commonly found that, because of the special conditions and political context of their implementation, censuses are often able to achieve more complete coverage of the population than sample surveys. This applies both to the size and spatial distribution of the population (by geographical and administrative location, urban-rural, etc.). Normally this also applies to the distributions by basic demographic characteristics such as age, sex, and household type and size.

886. Sample survey data, inflated by inverse of the selection probabilities of the units in the sample, should always be compared where possible with the totals and distribution in the sampling frame and with the census and census projections. This is termed as preparing ‘simple unbiased estimates’ from the survey. These estimates are a valuable means to evaluate coverage and representativeness of samples, as explained below.

## **C The importance of preparing simple unbiased estimates**

887. It is very important to be able to prepare "simple unbiased estimates" from the survey data, even though these may be refined and modified subsequently in the production of the final estimates. In practical survey work, the term "simple unbiased estimates" is used in the sense that: (i) the estimates are produced directly from the survey results without recourse to data external to the survey, by weighting each observation in inverse proportion to its probability of selection into the sample; and that (ii) at least with moderate to large sample size, the estimates so produced are approximately unbiased in the statistical sense (i.e., apart from any sample implementation and non-sampling errors).

888. Such estimates can be prepared only with probability sampling, i.e., for samples selected in such a way that each element in the population has a known and non-zero probability of being selected. To prepare good estimates, it is also necessary that problems of sample implementation, such as non-response and under-coverage, do not significantly distort these probabilities. Good simple estimates would also imply that any adjustments, which may have to be made subsequently to improve their precision, will not turn out to be large. In short, being able to produce good simple unbiased estimates indicates that the survey has been designed and implemented properly: hence the importance of such estimates in survey practices.

889. In a survey with a two-stage sample, for example, involving the selection and household listing of sample areas, followed by the selection and enumeration of households within each sample area, the number of households listed within each sample area, would be inflated by the inverse of the area's selection probability, and summed over all areas in the sample to provide an estimate of the total population (number of households in this case). This can be directly compared with the expected population size from census projections or other sources. The ratio of the sample estimate of the total population to the expected population size provides a good indication of completeness of coverage of the listing operation. In many surveys it has been found that the ratio falls short of 1.0 by a significant margin, indicating the need to improve the quality of household listing operations and using more up-to-date list for sample selection.

890. The above exercise should also be carried out by major geographical classifications of the survey population, by type of place (urban-rural) and other major domains, and also by main demographic classifications such as by age group, sex, household size and type.<sup>117</sup>

---

<sup>117</sup> Note that in estimating population aggregates actual values of the sampling probabilities are required, while for checking on population distribution, relative values of these probabilities suffice.

## Chapter 18      **Combined Uses Of Census And Survey Data: Current Estimates For Small Domains**

### **A      The need for small domain estimates**

891. The census can provide geographically detailed, but infrequent and not very current statistics, while statistics based on sample surveys can be more frequent and up-to-date, but lack geographical and other ‘small domain’ details due to sample size constraints. Sample survey results are widely used to derive reliable estimates for totals and means for large areas and domains. However, despite great developments in survey capability and practice, the usual ‘direct’ survey estimators for a small domain, based only on observations of sampled units in the domain, are likely to yield unacceptably large sampling errors due to the limited sample sizes involved.

892. Yet demands are growing everywhere for more timely and varied statistics for lower-level administrative units and other small domains. These needs can be met only by producing estimates that are both more frequent and more up-to-date than can be provided by censuses conducted at long intervals but which can be classified in much greater detail than is possible from sample surveys of limited size. Small area estimation methods, aimed at providing current yet detailed information, have been developed for this purpose. The basic idea of these procedures is to borrow and combine the relative strengths of more than one source in the production of better (more accurate and hence more useful) estimates.

893. In the past few decades, there have been major and sophisticated developments in the methodology of small domain estimation procedures. The objective of this chapter is not to review or evaluate those procedures, but merely to provide an introduction to the issues and some basic approaches.<sup>118</sup>

### **B      What are ‘small domains’?**

894. Firstly, it should be noted that the term ‘domain’ or ‘estimation domains’ is used to refer to the population or any part of the population for which separate statistics are required. The term ‘small area’ or ‘local area’ is commonly used to denote a geographical area which is considered “small” in some sense, such as provinces, counties, districts, other smaller administrative divisions, localities, or even census divisions or enumeration areas. The term “domain” is more general; it may refer to geographical areas, or to other sub-populations of interest such as specific groups by age, sex, nationality, race or other characteristics.

895. Secondly, it is important to note that “small” does not refer to smallness of the

---

<sup>118</sup> For an appraisal of recent developments in the field, see, for instance: Ghosh, M., and J.N.K. Rao, (1994). Small area estimation: an appraisal. *Statistical Science*, 9(1), pp. 55-93; Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25(2), pp. 175-186.

population size of the domain of interest, but to the smallness of the sample (number of observations) available on it. Small in one context may be very different from small in another context. For example, in statistically more developed countries where large and frequent sample surveys and/or administrative sources are available; the reference may be to very local areas or small population groups. In many less developed countries, the production of useful statistics even for large provinces or districts may require special ‘small domain estimation’ techniques because of the impossibility to sufficiently expand the sample sizes of national surveys for the purpose. The possibilities and suitability of specific small domain estimation procedures may greatly differ between these two situations.

896. Classification of domains according to size category can be useful in keeping the distinctions clear. Adapting a rough classification proposed by Kish,<sup>119</sup> we may distinguish between:

- (a) Major domains, say 1-10 divisions of the population, such as major regions of the country, major groups by occupation, industry, sex, age, etc. Sample surveys are usually designed to provide useful direct estimates for such divisions.
- (b) Minor domains, say 10-100 divisions of the population, such as individual provinces in Indonesia and Thailand or individual districts in Kenya, two-way classification by occupation and sex-age groups. In many developing countries, the primary interest is to extend the available statistics to this sort of level of detail.
- (c) Mini domains, say 100-1000 divisions of the population, such as individual counties in the United States or China. The production of reliable estimates at this level of detail is often beyond the capacity of many countries.
- (d) Rare domains may be used to refer to say 1000 or more subdivisions, such as rare populations or multi-way classifications.

### **C Diversity of methods**

897. Earlier methods of small area estimation focused on demographic methods for population estimation in post-censal years, and on estimation of the size of labour force status categories and other such characteristics of interest. Many of these methods used current data from administrative registers in conjunction with related data from the latest population census. Essentially, these ‘symptomatic estimation’ methods exploit the relationships between “symptomatic” variables (such as locally recorded number of births, deaths, school enrolment, housing units, etc.) and variables of interest to be estimated (such as the local population size). Generally such methods are very specific to the situation, and depend on the quality and type of administrative data available for the purpose.

---

<sup>119</sup> Kish, L. (1980). Design and estimation for domains. *The Statistician*, 29(4) pp. 209-222.

898. A second class of methods, the so-called ‘synthetic estimation’, borrow the structural relationships between variables available in detail (but which are not current) in the census, and impose those relationships in some appropriate manner on the less detailed but more current survey data to produce estimates which are current and detailed at the same time. The quality of the resulting estimates depends on the validity (stability) of the imported (postulated) relationships.

899. It can also be useful to combine indirect estimates with direct estimates from the sample, to construct ‘composite estimates’ as an appropriately weighted sum of direct and indirect estimates.

900. More recent and sophisticated techniques include procedures such as ‘empirical Bayes’, ‘hierarchical Bayes’, and ‘empirical best linear unbiased prediction’ procedures.<sup>120</sup> These methods have made a significant impact on small area estimation practice in recent years.

901. Some aspects and forms of the ‘synthetic estimation’ techniques referred to above appear both feasible and practical for application in the statistically less developed country circumstances. The basic idea behind this sort of approach will be outlined in the next section. Beyond that, the procedures mentioned above will not be reviewed here any further. Rather, in this practical *Handbook*, it is useful to emphasis the correct approach to the development and use of small domain estimation procedures. The suitability of any technique is situation-specific and is determined by several criteria:

- (a) Availability of data
- (b) Accuracy of the estimates
- (c) Practicality
- (d) Acceptability among the users.

902. The following summarises a number of practical lessons in the application of such techniques.

- (a) “From a variety of available methods several lessons may be learnt. First, that one may find among them a better method than the one he is arbitrarily using at present for small domains; this is often the passive ‘null’ method of continued reliance on the last decennial census that may be 12 years out-of-date.

---

<sup>120</sup> See for example Ghosh and Rao (1994) and Rao (1999).



- (b) Second, there is no single method that is best for all situations. Great differences between countries exist in the sources and quality of data available; the scope and quality of its census; the extents, contents and sizes of its sample surveys; and especially the scope and quality of its administrative registers. However, passive and negative attitudes are generally unjustified since every country has some resources, and ingenuity and effort can find unused resources of data. These may be of apparently different origins, but potentially useful because of high correlation with population sizes.
- (c) Furthermore the choice of sources and methods should vary with and depend on the nature of the statistics, on the desired estimates, and also on the domains to which they pertain. Note also the effects of the lapse of time since the last census. More generally, the balance between biases of a census and the variance of a large sample survey will move in favour of the latter during the 10 years between decennial censuses. The balance will also move in favour of less accurate but more current registers. The balance of sample surveys versus censuses or registers should depend on the sample size, but a fixed sample size has relative advantages in smaller populations.
- (d) Finally, the choice between methods is more difficult because the ‘best’ is often not clear even after the event. Errors in the estimates arise from biases chiefly, and ‘true values’ are usually not available for measuring the biases directly. Tests must be combinations of empirical and model bases, often depending eventually (and uncertainly) on results from decennial censuses. Better methods and criteria must be pursued with several methods and over the long range with an evolutionary approach and with patience”.<sup>121</sup>

“We may ... sketch the bare outline of present and future developments:

- (i.) Methods are useful and used now for post-censal estimates for local area statistics.
- (ii.) These methods will be used for other statistics also and in other situations.
- (iii.) The present methods can and will be improved.
- (iv.) The relative strength of different methods is difficult to predict and it depends on specific circumstances; they may be discovered in specific empirical trials.
- (v.) Success depends on first using better data and second on better methods. ... Good data sources are the principal means to better statistics ... [we need to] work towards the collection of other and better data [and also consider] strategies for cumulating data from

---

<sup>121</sup> Purcell, N.J., and L. Kish. (1980). Postcensal estimates for local areas (or domain). *International Statistical Review*, 48, 3-18.

samples for small areas.”<sup>122</sup>

### 1. *Illustration of a procedure*

903. Finally, this section provides a simple illustration of a potentially useful synthetic approach. As a simple illustration of the procedure, consider that a sample provides current estimate of some quantity or count (such as numbers by activity status)  $y_g$ , by sex-age group ( $g$ ). From the census, detailed but less current distribution of the population  $Y_{gh}$  by sex-age ( $g$ ) and small area ( $h$ ) is known. This latter provides an estimate of the distribution of an sex-age group by small area:

$$[Y_{gh}/Y_g] \quad [3.1]$$

904. On the assumption that this distribution (the ‘structural relationship’) is still valid, small area estimates of variable  $y$  of interest are given by summing over all sex-age groups  $g$ :

$$y_h = \sum_g (Y_{gh}/Y_g) y_g \quad [3.2]$$

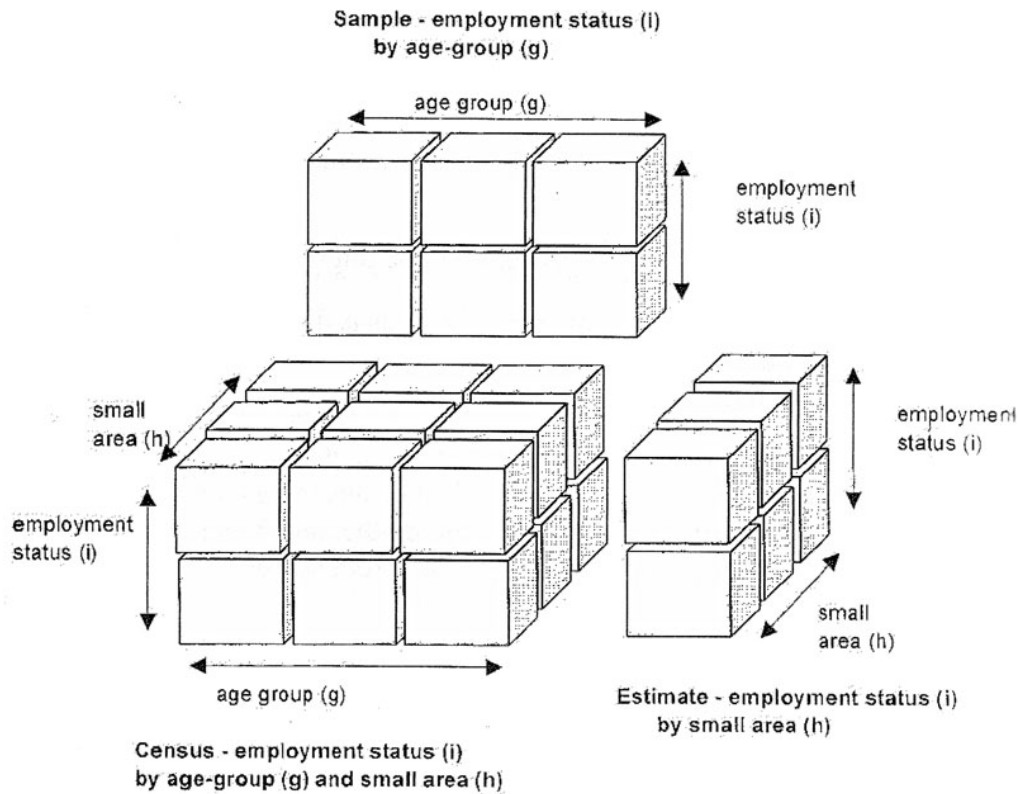
905. The basic idea of a more general form of the procedure, termed ‘structure preserving estimation’ (SPREE) by its authors, is illustrated by Diagram 6.<sup>123</sup>

---

<sup>122</sup> Kish, L. (1987). Discussion, in *Small Area Statistics: An International Symposium*, Platek et al (ed.). Wiley.

<sup>123</sup> Purcell, N.J., and L. Kish. (1980). See above.

**DIAGRAM 6: ILLUSTRATION OF A SMALL AREA ESTIMATION PROCEDURE**



906. Suppose that in the census, detailed but less current distribution  $X_{igh}$  of activity status ( $i$ ) by age-group ( $g$ ) and small area ( $h$ ) is available. Current but less detailed distribution  $Y_{ig}$  of activity status ( $i$ ) by age-group ( $g$ ) is available from the sample. This corresponds to the marginal distribution of the full census distribution, summed across small areas in the country. The interest is in estimating the distribution  $Y_{i,h}$  of activity status ( $i$ ) by small area ( $h$ ), i.e., the other margin of the full census distribution as shown in the diagram. This estimate is given by

$$Y_{i,h} = \sum_g \left( \frac{X_{igh}}{X_{ig}} \right) Y_{ig} \quad . [3.3]$$

907. This is based on the assumption that the term in the parentheses, namely the distribution of employment status by age-group ( $i$  by  $g$ ) across small domains ( $h$ ) has remained unchanged since the census. The corresponding distribution is 'imposed' on each current (sample) value by  $i$  and  $g$ , and then summed over all age-groups to obtain small area estimates of employment status (without classification by age-group of course).

908. Different forms of the procedure can be developed depending on the completeness of the information available from the census, and the types of constraints implied by the available information from the survey that must be satisfied by the resulting small area estimates.

## **ANNEXES**



## EXAMPLES OF COMPLETE SETS OF ECONOMIC QUESTIONS

It is difficult to find census questionnaires where the 'economic' questions cover all the topics included in the *Principles and Recommendations, Revision 2*. This is partly because there are other competing demands for space on the census questionnaire and also because some of these topics (and changes in conceptual boundaries) have been added since the 2000 census round and thus few countries have had a chance to even consider adopting them.' A few reasonably comprehensive examples of the 'economic' block of questions are presented here to show how they have been structured.

Examples from a range of countries, mainly English-speaking countries (one from a French-speaking country, two from Spanish-speaking countries; and several for which English translation was available), have been included at this stage. The sets of questions from the developed countries Australia, Canada, France, Italy and Japan, are also provided, although it is recognized that the questionnaires of these countries are quite lengthy, and in most cases will take far too much space for use for the census in the countries that represent the major audience for this Handbook. They however present a good basis for other countries considering a review of their questions and/or testing of new questions, to experiment with other formulations adapted to, their specific contexts.

The examples are presented under four groupings, following the format of the questionnaire from which they were extracted:

- (a) Questionnaires in the landscape format, one or two pages used for several individuals a specific household, with the listing of individuals along the rows and the questions presented in the columns.
- (b) Questionnaires using the portrait format, one or more pages used for several individuals within a specific household, with the listing of individuals along the columns and the questions presented along the rows.
- (c) Questionnaires using the portrait format, with one or more pages devoted to an individual/respondent; and the question set presented for each individual on one or more pages.
- (d) Questionnaires from developed countries (which follow the second and third types of formats) are presented separately.

## **A. Questionnaires using the landscape format**

The landscape format has been particularly favoured by countries in Africa, 'which have very few resources for censuses and often try to fit all questions onto one side of a large questionnaire. In most cases the topics included and the space allowed for each topic are severely restricted. Four examples of the block of economic questions in this format are presented (Botswana, 2001; Malawi, 1987; Namibia, 2001; South Africa; 2001).

Botswana 2001: Has a total of six questions, three on activity status (employment, unemployment, and job seeking); the other three on employment status, occupation and industry in that order.

Malawi 1998: Has three effective columns, one tries to combine 'current activity (including active, inactive unemployment and first-time job seekers) and 'status in employment'. The remaining two deal with occupation and industry.

Namibia 2001: There are four questions, the first dealing with activity status (unemployed and employed) as well as reason for non-economic activity. The others cover occupation, status in employment and industry.

South Africa 2001: Devotes ten columns to economic characteristics: the first two deal with activity status and reason for non-economic activity; the next two deal with unemployment, and the following four cover status in employment, industry (two related questions), and occupation. The last two are on hours worked and place of work.

**Botswana 2001**

ALL PERSONS		12 YEARS AND OVER			
ECONOMIC		ACTIVITY	OCCUPATION	INDUSTRY	JOB SEEKER
What has .... been doing mainly since Independence day 1999?	What was .... doing mainly in the past 7 days?	What was .... working as during the past 7 days?	What type of work did .... do in the past 7 days?	What was the main product, service or activity of .... place of work?	What steps did .... take during the last 30 days to seek work?
<b>Seasonal work</b> 01 Paid 02 Unpaid <b>Non-seasonal work</b> 03 Paid 04 Unpaid <b>Other</b> 05 Job Seeker 06 Home Maker 07 Student 08 Retired 09 Sick -- Other (specify)	01 Worked for payment (Cash or in-kind) 02 Worked in family business (unpaid) 03 Worked at lands/cattle post (unpaid) 04 Job seeker: GOTO A26 05 Home maker 06 Student 07 Retired 08 Sick Other (specify)	01 Employee - Paid cash 02 Employee - Paid in kind 03 Self-employed (no employees) 04 Self-employed (with employees) 05 Members of producers' cooperatives 06 Working in family business (unpaid) 07 Working at land/cattleposts (unpaid) 08 Apprentices	(Probe as necessary, use two or more words to describe the occupation)	(Probe as necessary, use two or more words to describe the industry)  GOTO A27 for Female	1 Register at Labour Office 2 Applied directly 3 Visited work sites 4 Asked friends/relatives Other(specify)
A21 (2)	A22(2)	A23	A24	A25	A26



Malawi 1998

PERSONS 18 YEARS OR MORE			
Marital Status	Type of activity last week	Economic Activity	
		Occupation	Industry
Married	What work... doing last 7 days? (Clear code)	What is this person's main occupation? <i>For example teacher, nurse, driver, machanic, reporter, etc.</i>	What is this person's trade or business? <i>For example construction, agriculture, transport, manufacturing, community services, etc.</i>
B16	B17	B18	B19

<b>ACTIVE</b>	<b>INACTIVE</b>
Male = 01	Nonmarket (Never worked before and not seeking work) = 08
Employee = 02	Homeowner = 10
Family business worker = 13	Student = 11
Self-employed = 04	Other = 12
Employer = 05	
Unemployed	
Worked before, seeking work = 06	
Worked before, not seeking work = 07	
Never worked before, seeking work = 08	

If female, job is 220  
 Other sex, job is the next person's job  
 Part C

**Namibia 2001**

For persons aged 8 years and above					
D	For those coded 01, 02 and 03 in D1				
	During the last 7 days prior to census night, did (name) work for at least one hour for pay, profit or family gain? 01 Yes, Worked 02 No, but has a job or a business 03 Unemployed (worked before) 04 Unemployed (first time jobseeker) 05 Student 06 Housemaker 07 Income recipient 08 Severe Disabled 09 Retired 10 Old age 11 Other, specify 99 Don't know	What type/ kind of work did/does (name) do at his/her place of work?  Write the main work done	For official use only	In your main job did you work as 01 Subsistence/Communal farmer (with paid employees) 02 subsistence/Communal farmer (without paid employees) 03 Other employee (with paid employees) 04 Other own account worker (without paid employees) 05 Employee (Government or parastatal) 06 Employee (private) 07 Unpaid family worker (Subsistence/Communal farmer) 08 Other unpaid family worker 09 Other, specify 99 Don't know	What kind of activity is carried out at (name's) workplace  Describe the activity
D1	D2	D3	D4	D5	D6
	-----			-----	
	-----			-----	
	-----			-----	
	-----			-----	
	-----			-----	
	-----			-----	

**South Africa 2001**

**ASK FOR ALL PERSONS AGED 10 YEARS AND OLDER (BORN BEFORE 10 OCTOBER 1991)**

ANY WORK IN THE 7 DAYS BEFORE 10 OCTOBER (P-18) In the SEVEN DAYS before 10 October did (the person) do any work for PAY (in cash or in kind) PROFIT or FAMILY GAIN, for one hour or more?	DID NOT HAVE ANY WORK	HAD WORK
	REASON WHY NOT WORKING (P-18a)	ACTIVE STEPS (P-18b)
1 = Yes: formal registered (non-farming) 2 = Yes: informal unregistered (non-farming) 3 = Yes: farming 4 = Yes: has work but was temporarily absent 5 = No: did not have work  # YES go to P-19	If NO to P-18 <b>What is the main reason why (the person) did not have work in the seven days before 10 October?</b> 1 = Scholar or student 2 = Home-maker or housewife 3 = Pensioner or retired person/ too old to work 4 = Unable to work due to illness or disability 5 = Seasonal worker not working presently 6 = Does not choose to work 7 = Could not find work  If more than one reason, write the code of the MAIN (most important) reason.	If NO to P-18 <b>In the PAST FOUR WEEKS before 10 October has (the person) taken active steps to find employment?</b> Y = Yes N = No For example, (the person) went to visit factories or other employment places, placed or answered advertisements, looked for land or a building or equipment to start own business or farm.
		AVAILABILITY (P-18c) If offered work, how soon could (the person) start? 1 = Within one week 2 = More than 1 week, up to 2 weeks 3 = More than 2 weeks, up to 4 weeks 4 = Some time after 4 weeks 5 = Does not choose to work <b>Go to P-20</b>
		WORK STATUS (P-19) If YES to P-18 <b>How can one best describe (the person's) main activity or work status?</b> 1 = Paid employee 2 = Paid family worker 3 = Self-employed 4 = Employer 5 = Unpaid family worker 6 = Other (specify)
		BUSINESS/COMPANY NAME (P-19a) If YES to P-18 <b>What is the FULL name of the business/company or organisation for whom (the person) works?</b>  If the person works for him/herself, and the business does not have a name, write SELF in the appropriate row. If doing PAID domestic work in a private household, write DOMESTIC SERVICE.  Use CAPITAL LETTERS only

**ASK FOR ALL PERSONS AGED 10 YEARS AND OLDER (BORN BEFORE 10 OCTOBER 1991) WHO HAD WORK**

COMPANY/BUSINESS ACTIVITY (P-19b)	OCCUPATION (P-19c)	HOURS WORKED (P-19d)	PLACE OF WORK (P-19e) (P-19f)
If YES to P-18	If YES to P-18	If YES to P-18	If YES to P-18
<b>What does the business do (main economic activity)?</b>  Write the MAIN INDUSTRY, economic activity, product or service of (the person's) employer or company. For example, gold mining, road construction, supermarket, police service, healthcare, hairdressing, banking. <b>OR</b> Write the activity of the person if self-employed. For example, subsistence farming.  If doing PAID domestic work in a private household, write DOMESTIC SERVICE.  Use CAPITAL LETTERS only.	<b>What is the main occupation of (the person) in this workplace?</b> Occupation refers to the type of work (the person) performed in the seven days before 10 October.  Use two or more words.  For example, street trader, cattle farmer, primary school teacher, domestic worker, fruit vendor, truck driver, warehouse manager, filing clerk, etc.  Use CAPITAL LETTERS only.	<b>How many hours did (the person) work in the seven days before 10 October?</b>  If (the person) was absent from work those seven days, but usually works, write the number of hours s/he usually works.	<b>Does (the person) work in the same sub-place in which s/he usually lives?</b> Y = Yes N = No Dot the appropriate box.  If NO, where is this place of work?  If NOT the same place, write PROVINCE P R, MAIN PLACE (city, town, tribal area, administrative area) and SUB-PLACE (suburb, ward, village, farm, informal settlement).  If another country, write the name of the country in the box below.

## **B. Questionnaires using the portrait format (type 1) (several individuals listed on one or more pages)**

The portrait format that list individuals along the columns and has the questions in the rows, is a major improvement over the landscape format with respect to space (both for formulating questions and recording of answers). Two examples of the block of economic questions are presented (Mauritania, 1988; and Sri Lanka, 2001).

Mauritania 1988: Although the questions are presented on several pages with more than one individual listed on a page, the format is similar to the landscape one, in which the individuals are recorded along the rows and questions/characteristics presented along the columns. There is however a lot more space allocated per individual/per question than in the landscape format. There are four questions on economic characteristics. One on activity status, and one each on occupation, status in employment and industry.

Sri Lanka 2000: The five questions that are presented deal with status in employment, occupation, industry, institutional sector of employment and principal non-economic activity, during the twelve-month reference period.

**Sri Lanka 2001**

P 18	<b>ACTIVITIES IN THE LAST 12 MONTHS AND THEIR DURATIONS (IN WEEKS):</b> A. Paid employment B. Employer C. Own account worker D. Contributing to family enterprises (unpaid family worker) E. Available/ Seeking work F. Non economic activities (Student, house keeping, income recipient/pensioner, unable/too old to work, child not attending school, other non economic activity)  If not engaged in any activity write '00' for that activity. The sum of all durations should be 52 weeks.	Activity Weeks A. Paid <input type="text"/> <input type="text"/> B. Emp. <input type="text"/> <input type="text"/> C.O.A.W. <input type="text"/> <input type="text"/> D.Family <input type="text"/> <input type="text"/> E. Available <input type="text"/> <input type="text"/> F.N.E.A. <input type="text"/> <input type="text"/>	Activity Weeks A. Paid <input type="text"/> <input type="text"/> B. Emp. <input type="text"/> <input type="text"/> C.O.A.W. <input type="text"/> <input type="text"/> D.Family <input type="text"/> <input type="text"/> E. Available <input type="text"/> <input type="text"/> F.N.E.A. <input type="text"/> <input type="text"/>	Activity Weeks A. Paid <input type="text"/> <input type="text"/> B. Emp. <input type="text"/> <input type="text"/> C.O.A.W. <input type="text"/> <input type="text"/> D.Family <input type="text"/> <input type="text"/> E. Available <input type="text"/> <input type="text"/> F.N.E.A. <input type="text"/> <input type="text"/>	Activity Weeks A. Paid <input type="text"/> <input type="text"/> B. Emp. <input type="text"/> <input type="text"/> C.O.A.W. <input type="text"/> <input type="text"/> D.Family <input type="text"/> <input type="text"/> E. Available <input type="text"/> <input type="text"/> F.N.E.A. <input type="text"/> <input type="text"/>	Activity Weeks A. Paid <input type="text"/> <input type="text"/> B. Emp. <input type="text"/> <input type="text"/> C.O.A.W. <input type="text"/> <input type="text"/> D.Family <input type="text"/> <input type="text"/> E. Available <input type="text"/> <input type="text"/> F.N.E.A. <input type="text"/> <input type="text"/>
P 19	<b>PRINCIPAL OCCUPATION OR KIND OF WORK DONE IN THE LAST 12 MONTHS:</b> If P18A to P18D '00' go to P 22 State clearly exact occupation or kind of work done during the last 12 months. eg. Paddy cultivator, Tea plucker, Accounts clerk, Mathematics teacher, Vegetable seller (retail), Domestic helper etc. If the person had more than one occupation/work, give the principal occupation/work on which most time was spent.	If P18A to P18D '00' go to P 22 _____ _____ _____ _____ _____ _____	If P18A to P18D '00' go to P 22 _____ _____ _____ _____ _____ _____	If P18A to P18D '00' go to P 22 _____ _____ _____ _____ _____ _____	If P18A to P18D '00' go to P 22 _____ _____ _____ _____ _____ _____	If P18A to P18D '00' go to P 22 _____ _____ _____ _____ _____ _____
P 20	<b>INDUSTRY, BUSINESS OR SERVICE:</b> Describe the kind of products made or services rendered at the work place of the principal occupation/ work. Write the name of the work place. eg. Paddy cultivation, Tea cultivation, Education department, Textile shop (retail), Garment factory etc.	_____ _____ _____ _____	_____ _____ _____ _____	_____ _____ _____ _____	_____ _____ _____ _____	_____ _____ _____ _____
P 21	<b>EMPLOYMENT STATUS AND SECTOR:</b> 1. Government employee 2. Semi government employee 3. Private sector employee 4. Employer 5. Own account worker 6. Unpaid family worker	1.Govt. 4.Empl. 2.Semi govt. 5.Own acc.wor. 3.Pvt.sec. employ. 6.Unpaid fam.wor	1.Govt. 4.Empl. 2.Semi govt. 5.Own acc.wor. 3.Pvt.sec. employ. 6.Unpaid fam.wor	1.Govt. 4.Empl. 2.Semi govt. 5.Own acc.wor. 3.Pvt.sec. employ. 6.Unpaid fam.wor	1.Govt. 4.Empl. 2.Semi govt. 5.Own acc.wor. 3.Pvt.sec. employ. 6.Unpaid fam.wor	1.Govt. 4.Empl. 2.Semi govt. 5.Own acc.wor. 3.Pvt.sec. employ. 6.Unpaid fam.wor
P 22	<b>PRINCIPAL NON ECONOMIC ACTIVITY IN THE LAST 12 MONTHS:</b> If P18F is '00' go to P 23 1. Student 2. Household work 3. Income recipient/ Pensioner 4. Unable/Too old to work 5. Child not attending school 6. Other non economic activity	If P18F is '00' go to P 23 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other	If P18F is '00' go to P 23 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other	If P18F is '00' go to P 23 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other	If P18F is '00' go to P 23 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other	If P18F is '00' go to P 23 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other 1.Student 4.Unable/ 2.H.h. Too old work 5.Child 3.Inc.rec./ not att. 6.Other

**Mauritania 1988**

**ACTIVITE ECONOMIQUE: PERSONNES AGEES DE 10 ANS ET PLUS**

SITUATION D'ACTIVITE	POUR LES «OCCUPES ET SANS TRAVAIL DEJA TRAV. CHER»		
	PROFESSION PRINCIPALE	SITUATION DANS LA PROFESSION	BRANCHE D'ACTIVITE ECONOMIQUE
18	19	20	21
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____
1 <input type="checkbox"/> Occupé (e)      5 <input type="checkbox"/> Menagere 2 <input type="checkbox"/> Sans trav. déjà trav. cher.      6 <input type="checkbox"/> Handicapé (e) 3 <input type="checkbox"/> Sans trav. cher. 1 <sup>er</sup> trav.      7 <input type="checkbox"/> Retraité (e) 4 <input type="checkbox"/> Etudiant-élève      8 <input type="checkbox"/> Rentier 9 <input type="checkbox"/> Autre	_____ _____ _____ _____	1 <input type="checkbox"/> Indépendant      5 <input type="checkbox"/> Apprenti 2 <input type="checkbox"/> Employeur      6 <input type="checkbox"/> Aide familial 3 <input type="checkbox"/> Salaire permanent      7 <input type="checkbox"/> Autre 4 <input type="checkbox"/> Salaire temporaire	_____ _____ _____ _____

### **C. Questionnaires using the portrait format (type 2)(one or more pages of questions for each individual)**

This portrait format provides the most ample space for both the questions responses to them. Often, the whole set of individual questions (as opposed to the household questions) are contained in a couple of pages for each individual. Few developing countries adopt this format due to cost considerations, but there are an increasing number of them. The examples of the block of economic questions presented are from Argentina, 2001; The Bahamas, 2000; Belize, 2000; Jamaica, 2001; Mexico, 2000; and Zambia, 2000.

Argentina, 2000: There are 13 questions dealing with economic characteristics; the first four are on activity status, employment, and unemployment. The next three cover industry and occupation (two questions). Status in employment and institutional sector are covered in three questions, and of the remaining questions, two deal with retirement benefits and the last covers the size of the establishment (with respect to the number of employees that it has).

The Bahamas, 2000: the 13 questions are presented in the section that deals with economic characteristics, and in addition a question in a separate section that deals with amount and sources of income. The first two questions deal with activity status and unemployment. These are followed by a question on multiple jobs and hours worked; and all these questions were in reference to the past week. The next four questions are on activity status and number of weeks worked in a twelve-month reference period. Another three questions are on industry (two questions) and occupation. Status in employment and a question on “informal trade” are the next set.

Belize, 2000: The questionnaire contains 17 questions (and sub-questions) on economic characteristics. The first seven questions deal with activity status, including one on hours worked, and another on reasons for non-economic activity. Five other questions deal with occupation and industry (present and previous for those currently unemployed), and the address of the place of work. Two questions cover present and previous status in employment, and the last two are on income.

Jamaica, 2001: The first four questions of the set of 20 questions cover activity status and the fifth is on hours worked. Status in employment, occupation and industry take up the next three, with the following two covering place of work. The size (employment) and names of the establishment take up two questions, and are followed by a question on income and one on the last employment held by the non-economically active. Two questions cover usual activity status (twelve-month reference period), and two more deal with permanent layoffs. The two last questions are on types of benefits from work. This questionnaire was administered to a 10 per cent sample of the census population.

Mexico, 2000: The nine economic characteristics questions presented in the

questionnaire are broken down as follows: three on activity status, a pair on occupation, one on status in employment, another pair on industry and the final one on income.

Zambia, 2000: There are five questions, one on current activity status (one-week reference period) and another on usual activity status (twelve-month reference period). The three remaining questions cover status in employment, occupation and industry in the twelve-month reference period.





**SECTION 8: ECONOMIC ACTIVITY (PERSONS 15 YEARS AND OVER)**

**P33. DID YOU DO ANY WORK AT ALL, FOR ANY LENGTH OF TIME DURING THE WEEK OF APRIL 24TH THRU 30TH**

(This would include helping in a family business/farm, street vending etc.)  
 1  YES (Skip to P35)      2  NO

**P34. WHAT WAS YOUR MAIN ACTIVITY DURING THAT WEEK?**

- |   |  |
|---|--|
| 01 <input type="checkbox"/> HAD A JOB BUT DID NOT WORK  | 05 <input type="checkbox"/> VOLUNTARY WORK WITHOUT PAY |
| 02 <input type="checkbox"/> LOOKED FOR WORK   | 06 <input type="checkbox"/> HOME DUTIES                |
| 03 <input type="checkbox"/> LOOKED FOR WORK DURING THE PAST 4 WEEKS                               | 07 <input type="checkbox"/> STUDENT                    |
| 04 <input type="checkbox"/> DID NOT LOOK BUT WANTED TO WORK AND WAS AVAILABLE DURING PAST 4 WEEKS | 08 <input type="checkbox"/> RETIRED                    |
|   | 09 <input type="checkbox"/> DISABLED                   |
| 10 <input type="checkbox"/> OTHER _____ (Specify)   |  |

(ALL PERSONS ANSWERING P34 SKIP TO P37)

**P35. DURING THAT WEEK, HOW MANY PAID JOBS DID YOU WORK AT?**  
 1  ONE      3  THREE OR MORE  
 2  TWO      4  NONE

**P36. HOW MANY HOURS DID YOU WORK ON YOUR MAIN JOB DURING THAT WEEK?**  
 1  1 - 8      4  33 - 44  
 2  9 - 15      5  45 & OVER  
 3  16 - 32      6  NONE

(ALL PERSONS ANSWERING P36 TICK 1 AT P37 AND SKIP TO P40)

**P37. DID YOU WORK AT ALL, FOR ANY LENGTH OF TIME DURING THE PAST TWELVE MONTHS?** (This would include selling newspapers, peanuts and other items, helping in a family business or farm, summer employment, etc.)  
 1  YES (Skip to P40)      2  NO

**P38. WHAT WAS YOUR MAIN ACTIVITY DURING THE PAST 12 MONTHS?**

- |  |                 |
|--|-----------------|
| 01 <input type="checkbox"/> HAD A JOB BUT DID NOT WORK (Skip to P40)       | } (Skip to P46) |
| 02 <input type="checkbox"/> LOOKED FOR WORK                                |                 |
| 03 <input type="checkbox"/> DID NOT LOOK BUT WANTED WORK AND WAS AVAILABLE |                 |
| 04 <input type="checkbox"/> VOLUNTARY WORK WITHOUT PAY                     |                 |
| 05 <input type="checkbox"/> HOME DUTIES                                    |                 |
| 06 <input type="checkbox"/> STUDENT  |                 |
| 07 <input type="checkbox"/> RETIRED  |                 |
| 08 <input type="checkbox"/> DISABLED                                       |                 |
| 09 <input type="checkbox"/> OTHER _____ (Specify)                          |                 |

**P39. HAVE YOU EVER WORKED OR HAD A JOB FOR AT LEAST TWO (2) WEEKS?**  
 1  YES      2  NO

(ALL PERSONS ANSWERING P39 SKIP TO P46)

**P40. HOW MANY WEEKS DID YOU WORK IN THE PAST TWELVE MONTHS?**

- |                                    |                                    |
|------------------------------------|------------------------------------|
| 1 <input type="checkbox"/> 1 - 4   | 5 <input type="checkbox"/> 40 - 48 |
| 2 <input type="checkbox"/> 5 - 13  | 6 <input type="checkbox"/> 49 - 52 |
| 3 <input type="checkbox"/> 14 - 26 | 7 <input type="checkbox"/> NONE    |
| 4 <input type="checkbox"/> 27 - 39 |                                    |

**P41. WHAT IS THE NAME OF THE COMPANY/BUSINESS WHERE YOU WORK OR FOR WHICH YOU LAST WORKED?**  
 (This and following questions refer to main job)

**P42. WHAT KIND OF BUSINESS OR ACTIVITY TAKES PLACE THERE?**  
 (Describe the kind of business e.g. retail store, primary school, law firm, brewery, etc.)

\_\_\_\_\_

**P43. WHAT TYPE OF WORK DO YOU/DID YOU DO?**

(Describe your job as accurately as possible e.g. sales clerk, typist, doctor, auto mechanic, civil engineer, taxi driver, housemaid, etc.)

(Do Not Say engineer, mechanic, teacher, supervisor, clerk, etc. Be more specific e.g. sales clerk, primary school teacher, auto mechanic, etc.)

\_\_\_\_\_

**P44. ARE/WERE YOU SELF-EMPLOYED OR WORKING FOR SOMEONE ELSE IN YOUR MAIN JOB?**

- |   |                 |
|---|-----------------|
| 1 <input type="checkbox"/> SELF-EMPLOYED (NO PAID HELPER)         | } (Skip to P46) |
| 2 <input type="checkbox"/> SELF-EMPLOYED (1-4 PAID HELPERS)       |                 |
| 3 <input type="checkbox"/> SELF-EMPLOYED (5 OR MORE PAID HELPERS) |                 |
| 4 <input type="checkbox"/> EMPLOYEE (GOVT/GOVT CORP)              |                 |
| 5 <input type="checkbox"/> EMPLOYEE (PRIVATE - 1-4 WORKERS)       |                 |
| 6 <input type="checkbox"/> EMPLOYEE (PRIVATE - 5 OR MORE WORKERS) |                 |
| 7 <input type="checkbox"/> UNPAID FAMILY WORKER                   |                 |

**P45. DO YOU MOVE ALL YOUR GOODS DAILY; E.G. FRUITS, PEANUTS, NEWSPAPERS, CLOTHING, EQUIPMENT?**  
 1  YES (INFORMAL TRADER)      2  NO

**SECTION 9: INCOME (PERSONS 15 YEARS OF AGE AND OVER)**

**P46. DURING THE PAST TWELVE MONTHS DID YOU RECEIVE INCOME FROM ANY OF THESE SOURCES? IF SO, STATE AMOUNT IN THE SPACE PROVIDED.**  
 (B\$. to the nearest whole number e. g. 12565.80 = 12566)

(PRIMARY JOB)  
 1. WAGES, SALARY, COMMISSION, TIPS, ETC.

\_\_\_\_\_

4. RETIREMENT PENSION

\_\_\_\_\_

7. GIFTS AND DONATIONS

\_\_\_\_\_

(SECONDARY JOB)  
 2. WAGES, SALARY, COMMISSION, TIPS, ETC.

\_\_\_\_\_

5. OLD AGE PENSION

\_\_\_\_\_

8. INVESTMENTS

\_\_\_\_\_

3. OWN BUSINESS

\_\_\_\_\_

6. GOVERNMENT ALLOWANCES

\_\_\_\_\_

9. OTHER SOURCES

\_\_\_\_\_

10. TOTAL INCOME DURING LAST TWELVE MONTHS

\_\_\_\_\_

**SECTION 11 ECONOMIC ACTIVITY**

FOR PERSONS 14 AND OVER PERSON 2

- 11.1 Do you/Does . . . own any land for farming?**  
 ① Yes    ② No    ③ Don't know/Not stated
- 11.2 Did you/. . . do any work for pay, profit, or family gain for at least one hour during the week ending May 12th? This includes helping in a family business or farm, street vending, or work at home.**  
 ① Yes (SKIP TO Q 11.5)    ③ Don't know/Not stated  
 ② No
- 11.3 Did you/. . . engage in any economic activity on the following list, for pay, profit or family gain, for at least one hour, during the week ending May 12th?**  
 ① Yes (SKIP TO Q 11.5)    ③ Don't know/Not stated  
 ② No

**INTERVIEWER: READ THE FOLLOWING LIST**

- |                            |                     |
|----------------------------|---------------------|
| Babysitting                | Nurse's Aid for pay |
| Laundry, ironing for pay   | Subsistence farming |
| Cleaning yard/cutting yard |                     |

- 11.3 list continued.**
- |  |                            |
|--|----------------------------|
| Selling food or snacks at market/bus-stop/school | Sewing for pay             |
| Bicycle/cart deliveries                          | Cleaning of offices        |
| Selling food from home                           | Car washing                |
| Selling sweets from home (fudge, etc.)           | Taxi-driver                |
|  | Lottery vendors            |
|  | Any other similar activity |

- 11.4 Did you/. . . have a job during that week, from which you were temporarily absent?**  
 ① Yes  
 ② No (SKIP TO Q 11.6)  
 ③ Don't know/Not stated (SKIP TO Q 11.6)

- 11.5 How many hours did you/. . . work during that week?**

	0	1	2	3	4	5	6	7	8	9
	0	1	2	3	4	5	6	7	8	9

99 Don't know/Not stated

(SKIP TO Q. 11.7a)

(Belize continued)

ALL PERSONS ECONOMIC		12 YEARS AND OVER ACTIVITY		OCCUPATION	INDUSTRY	JOB SEEKER
What has .... been doing mainly since Independence day 1999? Seasonal work 01 Paid 02 Unpaid Non-seasonal work 03 Paid 04 Unpaid Other 05 Job Seeker 06 Home Maker 07 Student 08 Retired 09 Sick -- Other (specify)	What was .... doing mainly in the past 7 days? 01 Worked for payment (Cash or in-kind) 02 Worked in family business (unpaid) 03 Worked at lands/cattle post (unpaid) 04 Job seeker: GOTO A26 05 Home maker 06 Student 07 Retired 08 Sick Other (specify)	What was ... working as during the past 7 days? 01 Employee - Paid cash 02 Employee - Paid in kind 03 Self-employed (no employees) 04 Self-employed (with employees) 05 Members of producers' cooperatives 06 Working in family business (unpaid) 07 Working at land/cattleposts (unpaid) 08 Apprentices	What type of work did .... do in the past 7 days?  (Probe as necessary, use two or more words to describe the occupation)	What was the main product, service or activity of .... place of work?  (Probe as necessary, use two or more words to describe the industry)  GOTO A27 for Female	What steps did .... take during the last 30 days to seek work? 1 Register at Labour Office 2 Applied directly 3 Visited work sites 4 Asked friends/relatives Other(specify)	
A21 (2)	A22(2)	A23	A24	A25	A26	

**Jamaica 2001**

**PC01C**

**SECTION 4**

**ECONOMIC ACTIVITY**

*(For persons 14 years and over)*

- 4.1 Did you/did ..... work for at least one hour during the first week of September 2001 ?  
 Yes ( Go to Q4.5 )    No    Not Stated
- 4.2 Did you/did ..... do anything like farming, buying and selling during the first week of September 2001 ?  
 Yes ( Go to Q4.5 )    No    Not Stated
- 4.3 Did you/did ..... do any type of odd job or hustling during the first week of September 2001 ?  
 Yes ( Go to Q4.5 )    No    Not Stated
- 4.4 What were you/was ..... doing for most of the time during the first week of September 2001 ?  
**(READ CATEGORIES)**
- Working in Agriculture or any other business without pay
  - With job not working ( Go to Q4.6 )
  - Seeking first job ( Go to Q4.15 )
  - Seeking a job which was not the first ( Go to Q4.7 )
  - Did not seek work but wanted work and was available ( Go to Q4.7 )
  - Student ( Go to Q4.14 )
  - Did Home Duties ( Go to Q4.14 )
  - Retired did not work ( Go to Q4.14 )
  - Disabled unable to work ( Go to Q4.14 )
  - Not interested in work ( Go to Q4.14 )
  - Other ( Go to Q4.14 )
  - Not Stated

- 4.5 How many hours did you/did ..... work during the first week of September 2001 ?
- |  |  |
|--|--|
|  |  |
|--|--|
- 
- Not Stated

**Confidential**

- 4.6 Which of the following categories best describes your/ ..... 's main employment?  
**(READ CATEGORIES)**
- Paid Government Employee
  - Paid Employee in Private Enterprise
  - Paid Employee in Private Home
  - Unpaid Employee in Agriculture or in any other type of business
  - Self Employed with Employees
  - Self Employed without Employees
  - Other
  - Not Stated

- 4.7 What kind of work do you do/does ..... do/did you last do/did ..... last do ?

--	--	--	--

- Never Worked ( Go to Q4.19 )    Not Stated

- 4.8 What type of business is/was carried on at the work place ?

--	--	--	--

- Not Stated

→ Go To Q 4.14 If Respondent Is Not Currently Employed Otherwise Continue

- 4.9 Where is your/ls ..... 's place of work located ?  
**(READ CATEGORIES)**

- In own home or yard    On a Farm
- In another home or yard    Not Stated
- Not in a private home

- 4.10 In which parish do you/does ..... work ?

- Kingston    Hanover
- St. Andrew    Westmoreland
- St. Thomas    St. Elizabeth
- Portland    Manchester
- St. Mary    Clarendon
- St. Ann    St. Catherine
- Trelawny    More than one parish
- St. James    Not Stated