STATISTICAL PAPERS Series M No. 71 (Part II)

RECOMMENDATIONS FOR THE 1983 WORLD PROGRAMME OF INDUSTRIAL STATISTICS

Part Two

Organization and conduct of industrial censuses

UNITED NATIONS

ST/ESA/STAT/SER.M/71(PartII)

DEPARTMENT OF INTERNATIONAL ECONOMIC AND SOCIAL AFFAIRS
STATISTICAL OFFICE

STATISTICAL PAPERS Series M No. 71 (Part II)

RECOMMENDATIONS FOR THE 1983 WORLD PROGRAMME OF INDUSTRIAL STATISTICS

Part Two

Organization and conduct of industrial censuses

UNITED NATIONS
New York, 1981

NOTE

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

ST/ESA/STAT/SER.M/71(Part II)

UNITED NATIONS PUBLICATION

Sales No. E.81.XVII.12

Price: $U.S. 11.00

Inquiries should be directed to:

PUBLISHING SERVICE
UNITED NATIONS
NEW YORK, N.Y. 10017

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

Annexes

INTRODUCTION

A. General remarks

1. The Statistical Commission of the United Nations at its twentieth session, in February-March 1979, endorsed a 1983 World Programme of Industrial Statistics and requested the Secretary-General, inter alia, "to prepare, as soon as possible, a document describing the purposes and the content of the Programme" as well as "a complete set of draft recommendations for the Programme" recognizing "the need for methodological guidelines". 1/

2. In response to that request, the recommendations for the 1983 Programme have been prepared in two parts. Part One outlines the general objectives and describes the nature and content of the Programme. The present publication (Part Two) is a manual designed to serve as a practical guide to the planning, management and conduct of bench-mark censuses and correlative annual inquiries of industrial activity. Both parts deal with the problems confronting countries at an early stage of their statistical development. Responding to the Statistical Commission's views that "... because of the rather disappointing results achieved in the 1973 Programme, ... caution should be exercised by not attempting too ambitious a programme for 1983", 2/ the items of data recommended for inclusion in the census, listed in Part One, have been reduced to the very minimum so far as developing countries just starting industrial statistics programmes are concerned.

3. The Commission emphasized that "the 1983 Programme might be more successfully implemented in the developing countries if limited to a full-scale, full-coverage canvassing operation of recognizable industrial establishments, with only a few basic items of data gathered". 3/ The bench-mark inquiry is regarded as only one aspect, although an important one, of a comprehensive, ongoing system of periodic and current industrial statistics. A continuing annual inquiry on industrial activities occupies the key role in that system. For developing nations the bench-mark census can serve as the first step in the development of the integrated programme, acting as the forerunner to the annual inquiry.

4. The recommendations (Parts One and Two) should be regarded as a revised and updated edition of Parts I and III of the Recommendations for the 1973 World Programme of Industrial Statistics. 4/ Still useful as a point of reference is

---

1/ Official Records of the Economic and Social Council Session, 1979, Supplement No. 3 (E/1979/23), paras. 13 and 10.

2/ Ibid., para. 7.

3/ Ibid., para. 6.

4/ Recommendations for the 1973 World Programme of Industrial Statistics, Part I, General Statistical Objectives, Statistical Papers, Series M, No. 54 (Part I) (United Nations publication, Sales No. E.71.XVII.13); Recommendations for the 1973 World Programme of Industrial Statistics, Part III, Organization and Conduct of Industrial Censuses, Statistical Papers, Series M. No. 54 (Part III) (United Nations publication, Sales No. E.72.XVII.10).

Part II of those 1973 recommendations, 5/ which was not modified for the 1983 Programme. Part II of the 1973 publication contained a standard list of individual commodities for which, it was recommended, production and consumption data be compiled in industrial censuses. Other publications which provide reference material of value to a programme of bench-mark inquiries are listed in the bibliography to the present publication. For example, the conceptual framework of the inquiries proposed in the 1983 Programme manual is outlined in the International Recommendations for Industrial Statistics. 6/

5. For the purposes of the present publication, a census is regarded as a comprehensive inquiry conducted at infrequent intervals, usually of five or 10 years. In addition, an integrated industrial statistics programme includes an ongoing annual inquiry, as well as monthly and quarterly surveys. The census and more frequent surveys encompass the activities of mining, manufacturing and the supply of electricity, gas and water, as defined in the International Standard Industrial Classification of All Economic Activities (ISIC), 7/ major divisions 2, 3 and 4. The units for which statistics are compiled are industrial establishments (factories, mills, works, plants, mines and quarries) within the territorial boundaries of the country. Establishments may be single-unit enterprises or parts of multi-unit enterprises.

B. Organization of the manual

6. The present manual (Part Two) discusses the steps that should be taken in planning and conducting an industrial sector inquiry. Its content has been organized to follow the sequence in which operations should be carried out in the inquiry. In many aspects, the methods used in bench-mark censuses and annual inquiries are quite similar.

7. Chapter I describes the initial steps and major considerations involved in the selection and organization of the inquiries. Chapter II covers the needs and uses of industrial statistics. Chapter III is devoted to a discussion of general planning and organization for the industrial census. Chapter IV deals with the industrial directory, small establishments and household units, sampling, and pre-testing. The topics of management and budget are treated in chapter V. Chapter VI examines issues related to the design of the questionnaires and related instructions. Chapter VII considers methods and problems of data collection. Chapters VIII, IX and X deal with the steps associated with data processing, tabulation, review and correction. The final chapter (XI) relates to

---

5/ Recommendations for the 1973 World Programme of Industrial Statistics, Part II, List of Selected Products and Materials, Statistical Papers, Series M, No. 54 (Part II) (United Nations publication, Sales No. E.71.XVII.16).

6/ Statistical Papers, Series M. No. 48 (United Nations publication, Sales No. E.68.XVII.10).

7/ Statistical Papers, Series M, No. 4, Rev.2 (United Nations publication, Sales No. E.68.XVII.8).

the publication of preliminary and final results. A detailed treatment of practical sampling techniques to be used in industrial-type inquiries is provided in annex I. Annex II displays specimen questionnaires for consideration by countries adopting either the minimum or the complete programme. There is a bibliography of United Nations publications that provide background material for the 1983 World Programme and are often cited in the present publication.

8. The present manual should be regarded as a set of practical suggestions on organizing and carrying out industrial inquiries. The approach emphasizes the most basic steps to be kept in mind: for example, early planning, the development of tabulation plans, care in designing the questionnaire and phrasing the inquiries, the need for pre-testing and quality control of the processing. In an undertaking as large and complex as an industrial census, failure to carry out properly any one of these actions may mean serious trouble for the national statistical agency. It is obviously impossible to include in a single publication everything that one needs to know about taking an industrial census. This manual assumes that the statistical authority has knowledgeable personnel, familiar with the subjects of data collection, computer processing, sampling, forms design, and the like. Countries in the beginning stages of a statistical programme may have to rely heavily upon the literature available on the different topics and to supplement their own resources with outside technical assistance. Each country will need to study these recommendations in depth and to adapt them to its particular conditions and resources.

9. Although international recommendations are intended to reflect common practices in the majority of countries, those interested in a more detailed discussion of certain topics or in the particular experience of certain countries should communicate with such countries to determine the problems they might have faced and to benefit from their solutions. The first industrial census resembling the modern type in its comprehensive coverage of establishments was conducted in the United States of America for the year 1810. Since that time, many countries have conducted industrial censuses and much of the experience of the government agencies managing them has been recorded. Today, a country contemplating its first industrial census can take advantage of a considerable body of technical advice to assist it in planning and managing the inquiry. Advice on particular problems may also be obtained by writing to the Director, Statistical Office, United Nations, New York, N.Y. 10017.

Chapter I

PRELIMINARY STEPS AND MAJOR CONSIDERATIONS

A. Preliminary steps

10. While the international recommendations suggest the kinds of data to be collected on industrial activity, it is necessary for each country to review for itself the data already being collected and what its bench-mark census objectives should be. This requires a thorough investigation into the data requirements of both the public and the private sectors. In that review of data needs, which is discussed in the next chapter, it should be kept in mind that there are two major and highly integrated statistical programmes that contribute to satisfying those needs. The first consists of a comprehensive census, taken every five or 10 years, which attempts to develop statistics in comparatively great detail as a basis for understanding the structure of that portion of the industrial sector covered by the census. An industrial census should be sufficiently complete and accurate to provide a satisfactory directory of establishments and, appropriately updated, to serve as the basic framework for surveys of the industrial sector taken at more frequent intervals and with more limited goals.

11. The second type of programme is the annual survey of industrial activity, usually conducted on a sample basis and often supplemented by monthly or quarterly surveys dealing with short-term measures of change, such as employment, shipments, unfilled and new orders and stocks on hand. While the samples used in the annual inquiry and other current surveys do not necessarily have to be selected from complete census enumerations, they will be more reliable and useful if so based. Such linkage ties the two sets of data together and acts as a check on the functioning of the sample between census periods.

12. Some countries with well developed statistical systems are able to maintain an industrial directory without infrequent censuses through their annual programme, supplemented by information made available by revenue or social security authorities. If such a directory is available, giving up-to-date information about the name, address, type of organization, industrial classification and indication of size, such as turnover or employment of each establishment, sampling methods can be used to provide estimates of the structure of the industrial sector, which must otherwise be sought in the infrequent censuses. The directory, whether based on a bench-mark census or an up-to-date register of establishments, should be sufficiently complete and accurate to serve as a sampling frame for surveys of the industrial sector.

13. Part One of the present publication includes recommendations on the concepts to be followed in undertaking the annual and bench-mark industrial inquiries, as well as on the items of data to be gathered. It is not expected that these recommendations will be found appropriate in every detail by all countries. Appropriate modifications should be made dependent on each country's special circumstances. Each national programme should, however, be fitted into the framework of the international recommendations so that meaningful comparisons can be made from country to country and world statistics can be further developed on the basis of standard concepts. As the world's economies become more interdependent, standardization can lead to a better understanding of the interplay of economic forces and a more intelligent solution of economic problems.

B. Major considerations

1. Need for advance planning

14. Because of the complexity of both the bench-mark census and the annual inquiry, it is critical that all aspects of the work to be performed be understood and planned far ahead of time. The importance of advance planning of the collection, processing, tabulation, review and publication of the data cannot be overemphasized. Considerable sums of money and much effort have been expended in trying to correct errors which could have been avoided by careful planning. It is necessary to consider carefully each operation and its interactive effect to make sure that the final product will be acceptable.

15. The elements in a census are closely interrelated, and an initial decision on one aspect of the programme is likely to affect several others. To mention only two examples: questionnaires cannot be designed until the content of the census has been decided, and editing and coding procedures can only be determined in the context of considering costs in relation to standards of accuracy judged to be desirable and attainable. The development of a well-thought-out programme, adapted to the needs of the particular country, is important. Statistical systems that are successful in one country may be inappropriate for use in another country.

2. Budget for the census

16. A realistic budget is a function of careful advance planning and a thorough understanding of cost allocation procedures. It is important that a competent budget analyst review all operations to make sure that the budget is proper and sufficient. Contingency reserves must be provided in order to cover unforeseen emergencies occurring during the conduct of the census. Finally, the entire budgetary procedure must be thoroughly reviewed to ensure that nothing of significance is omitted.

17. The amount of funds made available for the census will largely determine what can be expected in the way of results. It should also be noted that the depth and quality of the census planning operation has a great effect on whether the budget permits the achievement of the desired objectives. One element that tends to be overlooked as a budgetary consideration is the simple matter of timing - that is, the overall time-limit set in advance for accomplishing the task. Because of poor planning and budgeting, some countries have found that their funds for the census were largely expended in the collection phase and insufficient resources were left to complete the processing of the questionnaires, the review of the tabulated data and the publication of the results.

3. Limiting requests for information

18. There is a widespread awareness today of the need for industrial statistics in economic and social planning and in decision-making by businessmen. Data users, however, are often either unaware of or unsympathetic to the practical problems attending the collection and compilation of statistics and may suggest collecting information which is difficult or impossible to obtain. The number of queries on the census questionnaire may make it so long and detailed that it arouses opposition among respondents. An excessively long questionnaire adds greatly to the cost of collection, editing and tabulation. In the past, some countries have overloaded their census programme to the point where the entire system has collapsed because

of its complexity. It is especially important in initiating a new programme to keep the content of the questionnaire as simple as possible to assure the successful completion of the census.

4. Pre-testing

19. For these and other reasons, it is imperative that the questionnaires developed for a census or annual inquiry be tested before being used, regardless of the method of enumeration to be employed. Such tests need not be costly or elaborate. Often, the information obtained from interviewing a small number of respondents under actual field conditions will be of immense value in deciding on the final content and wording of the questionnaires. In addition, it is desirable that pre-testing be done on all aspects of the survey, from the development of the respondent universe to publication of the results. Over-all planning of the census programme should allow for timely analysis of the results of these tests and for the necessary revision of plans and procedures.

5. Co-operation of respondents

20. To have a successful survey, it is essential to obtain the co-operation of the respondents. One way of obtaining such co-operation is through legal compulsion and penalties for non-response are therefore generally included in the legislation for censuses. The legal compulsion to report does not, however, prevent unco-operative respondents from filing late or reporting carelessly or erroneously. Advance publicity and the training of enumerators to explain the merits of the census convincingly are important ways of increasing the co-operation of respondents.

21. The degree of co-operation exhibited by respondents can greatly affect the cost of collecting the reports: for example, several field visits may be required to obtain information from a reluctant respondent, while a friendly and co-operative respondent will supply the information on the first visit. It behoves the census organization to work with business groups and representatives of firms of every size (large, medium and small) to establish a favourable climate for the conduct of the survey. An essential element in eliciting good response is that the data requested be obtainable from the business records of the respondents.

6. Industrial directory

22. The existence of an industrial directory in advance of the actual enumeration can be of great assistance in conducting the census. A directory facilitates the checking of field enumeration and the control of coverage in the collection phase and aids in ensuring completeness at all stages of processing. The prior establishment of a directory will also contribute significantly to the efficient use of sampling in covering the smaller units in the census. A mailing list may be available from regulatory agencies or licensing authorities, or can be developed from their records. Such a list frequently contains an industry classification of the units and sometimes certain data, such as volume of sales.

7. Data collected by other government agencies

23. A review of governmental functions may reveal that a considerable amount of data is already available in connexion with the operations of the revenue authorities, the administration of labour laws or the operation of social security

programmes. These data need to be examined to determine whether they are usable in statistical programmes as well as in the regulatory or administrative agencies that originally required them. It may be feasible to ask the administrative or regulatory agencies to expand or modify its requests for data in order to accommodate some of the statistical needs of the census programme. However, sometimes legal difficulties may be encountered in acquiring access to the data of other agencies, particularly those of the revenue authority.

24. Use of the records of other governmental agencies will usually require skillful negotiation, since such agencies are not primarily concerned with statistics. It will also necessitate careful planning to make certain that such records become available on time and in the proper form for use in the census. A problem may arise because data from a regulatory agency may have been collected under a promise of confidentiality. Since the statistical agency should also be collecting its data under a law that promises confidentiality, the use of the data from the regulatory agency can be justified so long as the statistical agency personnel are sworn to confidentiality and that agency's publications will contain no data that disclose the operations of any firm. The converse, however, is not true; that is, information on individual firms collected by the statistical organization cannot be transferred to a regulatory agency, since such information should as a rule have been obtained with the guarantee that it would not be used for regulatory or taxation purposes.

8. Use of sampling

25. A basic consideration in any economic census is whether to use sampling or to conduct the census on a complete enumeration basis. When the scope of the census is broad and a large number of establishments is to be covered, a size cut-off may be desirable, whereby the smaller establishments are accounted for by sampling a predetermined percentage of them on a randomly selected basis. Alternatively, it may be preferable, considering the demands for data and budget limitations, to ask all respondents the priority questions only and to limit the respondent panel for additional items to a sample that is heavily weighted towards the larger establishments. In either case, the sample design must represent the universe which is being covered and have measurable reliability.

9. Organization for conducting the census

26. There is usually a data-gathering organization already in existence in most countries. A decision has to be made regarding the extent to which the different surveys (economic and demographic; current and bench-mark) can or should be integrated, inasmuch as there are usually a number of common elements. While it is possible to integrate much of the field work and data processing, other functions that require subject matter and industry knowledge, such as specialized edits, data review and review of tabular content need to be provided for in a separate division. Returns from large manufacturing and mining establishments and problem cases in those areas possess such unique and complex features that over-centralization of the editing and coding operations has proved to be self-defeating. In many countries, the subject-matter divisions of the government's statistical agency:

(a) design the census questionnaires; (b) plan and develop the census publications; (c) prepare the technical specifications for clerical operations; (d) resolve the difficult data-processing problems; and (e) develop and prepare computer systems and programs for processing the census data. In some countries, function (e) is assigned to the computer services division, which operates and manages the electronic digital computers and peripheral equipment.

Chapter II

NEEDS AND USES OF INDUSTRIAL STATISTICS

A. General remarks

27. This chapter describes the needs for and uses of the information obtained in industrial censuses and annual inquiries. An understanding of these uses will assist statistical planners not only in selecting the items to be covered and the emphasis to be given to each, but also in preparing publicity and in answering questions about the purposes of the inquiries. Especially in the case of small industrial enterprises, there is little awareness that information of this kind can be of great value in formulating national economic policies and that, in many instances, the data can be of direct help to industry managers in making effective decisions.

28. The growth of industrial activities in a country has nearly always been accompanied by a demand for facts about the structure of these activities and changes that have taken place in them, as well as the nature of their current status. Each decade has witnessed more and more business decisions made on the basis of facts rather than solely on the basis of "judgement" or "experience". The role of Governments in collecting and disseminating such facts has varied, but in all countries where comprehensive censuses have been conducted, governmental organizations have assumed the dominant role. Private groups, such as manufacturers' associations, which are generally limited in their scope and authority, are not able to conduct censuses.

29. The census should be accompanied by annual inquiries of industrial activity in the years between censuses. Both the government and the private sector want to be kept informed of changes in the status or movements in industries and regions and in shifts in inputs and outputs. The annual surveys satisfy that objective by carrying forward yearly the principal measures of employment, labour and material costs, stocks, capital expenditure and shipments. More and more, countries are responding to the increased need for extensive annual data, a need underscored by the requirements of the system of national accounts. The annual inquiry is truly to be regarded as the central inquiry of an integrated system of industrial statistics.

B. Governmental needs and uses

30. Data on manufacturing and mineral activities and on the supply of energy (electricity and gas) and water are essential to the process of developing policy (and in monitoring short-term trends and measuring the industry's response to the policies already established) in many different areas of governmental activity. In addition to being one of industry's largest customers, government is also the supplier of many of the services on which manufacturers and mineral companies rely and the Government needs information on industrial activity if these services are to be provided efficiently. If highways and employment service offices, to cite only two examples, are to be properly situated, the Government must have facts on the level of activity and whatever changes have taken place. Other programmes that may be helped include planning of regional development programmes, the drafting of

social welfare programmes, such as social security, minimum wage and unemployment insurance, the evaluation of anti-trust programmes etc. It would be difficult to provide adequate national security without industrial and regional statistics regarding the country's industrial plants.

31. The measurement of industry's response to the Government's fiscal and economic policies calls for industrial statistics, as does the current administration of most institutions regulated or operated by the Government. The data are also need needed for the computation of national income and product statistics. In forecasting and planning, it is especially useful to analyse the national product in terms of the transactions that determine its size and composition.

C. Business needs and uses

32. While the business community is the principal supplier of data on industrial activity, it is also a very important user of such data. The industrialist must turn to statistical compilation for the many facts he requires. Only through statistical compilation can he obtain the large quantities of information that are sufficiently reliable and comparable, among firms and from one period to another, to serve his needs.

33. Industrial enterprises rely upon census facts to compare their enterprise's performance with others in the same industry and to find out if the market for their products is expanding or contracting. Census facts are also used in considering plant expansion, locating warehouses and distribution outlets, selecting sales territories and checking costs. The census publications can provide the businessman with answers to such questions as: (a) What is my share of the market? How has it changed since the previous census inquiry? (b) How do my input-output costs and other plant characteristics compare with the average for my industry or my size group in that industry?

34. The answers to these standard business questions can be found in the census publications. No matter how circuitous the route - through reports from trade associations, trade publications or studies by economists - the basic data needed to provide answers come from the censuses, annual inquiries and other current surveys. Census data afford the executive a realistic understanding of the economy as a whole as well as of his own particular economic sector.

35. In addition to the industrial enterprises themselves, firms performing services for industry also find the data useful. For example, transportation and communication companies, as well as trade and service operators, can do a much better job of planning if they have a detailed knowledge of the structure and location of the units involved in manufacturing, mining and the production of electricity, gas and water.

D. Other needs and uses

36. Trade, academic and research institutions and professional associations need

census information to learn how industries with which they are concerned are changing. Local governments need to understand the industrial structure and the changes occurring in their areas. Labour organizations find the data useful in wage negotiations. The knowledge they gain of the structure and functioning of their industry contributes to setting reasonable demands and negotiating effective contracts.

37. Another basic use to which the results of the infrequent comprehensive census can be put is in the design of samples to be used for other surveys, especially those conducted annually or at more frequent intervals. The census is the ideal framework from which the most efficient sample can be selected, inasmuch as it contains such characteristics of the units to be sampled as kind, size and exact location of establishments. While private sampling organizations cannot have access to the establishment information reported to the Government, it is still possible for them to make good use of the published data, both in designing an efficient sample and in checking the results of their work against census totals.

38. The census measures also serve as bench-marks for current surveys of industrial activity and for indexes of industrial production, productivity and prices, all of which are essential in developing the general background for understanding particular economic situations.

E. Needs and uses for specific items of data

1. Employment, hours (days) worked and earnings

39. Governments are interested in programmes that promote economic and social welfare, and current facts about employment, hours worked and earnings are key indicators of the progress of such programmes. Recognizing this interest, the International Recommendations for Industrial Statistics 8/ include recommendations for the collection of employment, hours (days) worked and earnings data in current industrial inquiries. A major role of the industrial census in this connexion is to provide bench-mark data and mailing lists for use in such inquiries.

40. An additional role of the industrial census in this area is in the development of measures relating labour input to value added, gross output and other measures of industrial activity. The development of a set of data capable of reflecting the many interrelationships among the various types of activity, factor inputs and the output of products and services is often looked upon as the principal contribution of an industrial census. Productivity studies usually originate in data from industrial censuses, and such studies are needed to estimate the additional workers demanded by contemplated expansions in industry and to judge the feasibility of such expansions. In addition, data on labour productivity afford clues to the

8/ Statistical Papers, Series M, No. 40 (United Nations publication, Sales Sales No. E.68.XVII.10).

extent of automation or mechanization among industries. Other requirements for data on the labour force arise from the formulation and administration of labour laws and negotiations regarding wage and salary rates.

2. Gross output and input

41. Industrial census data on the quantity and value of particular commodities are required for supply and demand analyses. These analyses are useful in determining national policy in foreign trade matters, in studying the relationship of industrial activities to the national economy as a whole, and in dealing with surpluses or deficits in the supply of commodities and related price problems.

42. Statistics on the consumption of specific materials, distributed by industry or by location, are of value to industry management in the sectors where the materials originate. Reliable market information of this type is not usually obtainable except from industrial censuses.

43. Data on the value of the goods and services produced, as well as on the cost of the goods and services consumed in their production, are essential components of input-output tables. Such tables, which measure in some detail the flow of goods and services into and out of the various branches of industry, are of considerable value in ecomonic planning and programming; they are an integral component of the more advanced systems of national accounts.

3. Stocks

44. In addition to their use in connexion with studies of national wealth, data on inventory levels provided by industrial censuses are useful as bench-marks and as mailing list sources for more frequent surveys on stocks. Such surveys, usually covering specific products or materials of considerable importance to a country's economy, are valuable in connexion with economic analysis, particularly short-term forecasting. The ratio of stocks of materials, on the one hand, and of finished products, on the other, to production (or sales) are useful indicators of the development of the trade cycle in a country. Data on changes in inventory levels between the beginning and end of the census year are necessary for the calculation of value added when information on products and materials is gathered on a shipped or received basis.

4. Fixed capital formation

45. Since the level of fixed capital formation, particularly in the mining and manufacturing sectors, is a leading indicator of changes in activity, current information on investment in plant and equipment attracts great interest. Industrial censuses provide bench-marks, as well as mailing lists, for sample surveys designed to collect the related current information.

F. Respondent burden

46. As products, processes and organizations have become more complex, the demand for statistics has increased. In industrialized countries, statistical collection has grown to such an extent that careful attention must be paid to the value of the information collected in relation to the cost of providing it. In the case of government-sponsored statistical projects, it is important to keep in mind the fact that the cost to industry of supplying the information may be as great as or greater than the cost to the Government of collecting, compiling and publishing it.

47. It has long been recognized that small enterprises maintain less detailed records than do large ones and thus have more difficulty in supplying census-type data. It may also be observed that large firms are better equipped to use statistics and, accordingly, are more likely to recognize the usefulness of a census and more apt to co-operate with the census enumerator, or to fill out a mail questionnaire completely and conscientiously. Imposing severe limits on the questions to be asked of small enterprises is, therefore, doubly important.

Chapter III

GENERAL PLANNING AND ORGANIZATION

A. General remarks

48. The success or failure of an industrial census depends in large part on the care taken in formulating its broad objectives and in the detailed planning at an early date of each phase of data collection and review and the tabulation and publication of the results. The experience of countries with a long tradition of industrial censuses has demonstrated that advance planning for each census is essential, even though several successful censuses may have been conducted previously. The programme evaluation review technique, known as PERT, is a frequently used planning tool.

49. Countries with little or no experience in industrial census-taking should plan with even greater care, because their resources are usually limited and there may not be sufficient funds or manpower available to correct the effect of a serious misjudgement. Without adequate planning, there may be unacceptable delays in the completion of the work. This occurred in the case of several developing countries in the 1973 World Programme of Industrial Statistics. In such countries, the work of the statistical agency is hampered by the fact that the nation's industrial enterprises will have had little or no experience in furnishing statistics and will frequently have little appreciation of the value of economic data. For a country conducting its first industrial census, there is always the concern that an unsuccessful attempt may result in the refusal of the Government to sponsor similar projects in the future.

B. Approval of the census programme

50. As the initial step in the process of receiving approval for the industrial census, it is advisable to prepare a brief or justification for the proposed programme. Much of the discussion that follows provides the framework for preparing such a justification. It should include realistic estimates of programme content, cost, timing and personnel requirements. This need not be in great detail but should be sufficient to give the director of the statistical agency an indication of order of magnitude, so that he or she can decide whether to recommend approval or modification of the programme.

C. Legislative requirements

51. The legal authority for the industrial census is fundamental, because it gives the project the stability and continuity needed to carry it to a successful conclusion. It also contributes significantly to ensuring a satisfactory rate of response. The type of legal authority required depends to a large extent upon the legal traditions of the country. Industrial censuses have generally been conducted under one of three kinds of authority:

Executive decrees. Where wide powers to gather economic information are already vested in a government official, an executive decree or order would carry sufficient authority. The decrees issued have usually specified the general nature of the inquiries to be made, the obligations of respondents to furnish information, and the responsibilities of the Government in connexion with protecting the confidentiality of the individual returns and with compiling and publishing the statistical totals.

General statistical acts. Such acts pertain to statistical systems rather than to censuses alone, and may encompass demographic as well as economic surveys, and current data as well as censuses. The acts generally have clauses on the obligations of respondents and the Government with respect to them. To provide for specific industrial censuses, further action in the form of a decree or other executive action may be necessary.

Industrial census acts. These deal only with industrial censuses but in other respects have similar provisions to the more general statistical acts. They differ in the detail in which the provisions appear; for those that are not sufficiently detailed, further executive action may be necessary.

Although the guarantees of confidentiality should be very specific and penalties stipulated for failure to respond or wilful falsification of the data reported, in general the legislative authority should grant a considerable degree of flexibility to the statistical agency regarding issues of scope, coverage, frequency of the censuses etc. In particular, the acts or executive decrees should be drafted broadly enough to allow for initiating annual and more frequent surveys in progressing towards a comprehensive industrial statistics system. The "mandatory authority" provisions should be extended to embrace the annual inquiry of industrial activity.

D. Organizational structure

52. A decision as to which government organization is to be responsible for conducting the industrial census should be made at an early stage. There are decided advantages in concentrating all authority and operations in connexion with the census in a single agency of government. The resultant advantages are the central co-ordination of all detailed plans, the speed with which changes in plans can be made effective and the economies in recruiting and training personnel to work full-time on the census rather than part-time in addition to their other duties. This should not, of course, exclude officials in other agencies from participating in the planning of the census through committees. Also, in particular situations it may be beneficial to enlist the assistance of other government agencies. For example, local or provincial governments might be called upon to provide interviewers, or the services of teachers or field agents of the ministry of agriculture used to supplement the regular field staff of the national statistical organization.

53. Most of the participants in the 1983 World Programme of Industrial Statistics will have taken population censuses, and there are good reasons for the many similarities in the organizational structure adopted for the two censuses. In concept, both the economic and demographic censuses should be carried out through a chain of command, extending from the central office down to the provincial and

local supervisors and ultimately to the enumerator in the field, in those instances where the economic census is not a mail canvass. Field personnel working on demographic censuses or current surveys may be used on the industrial census. There are, however, significant differences between the censuses and therefore the adaptability of the demographic census organizational structure to the economic sector should be carefully evaluated. The population census inquiries are not so complex; nor are they subject to the differing interpretations found in the industrial sector. Furthermore, it is desirable for an enumerator to have some knowledge of accounting practices and record-keeping systems in dealing with business enterprises. The length of the enumeration period is much shorter in the case of the population canvass.

54. There are undoubted advantages for data users, especially those concerned with the national accounts, and for the statistical agency itself in conducting the industrial census in conjunction with the census of distributive trades and services or possibly even the construction industries. The statistical organization could realize economies in its industrial directory work, field organization and the use of computer facilities. On the other hand, history shows that a relatively inexperienced statistical agency is usually not able to cope with all of the resulting problems within the limits of the time and resources available, and the peak and trough work-load problems can be greatly eased if the several censuses are spread over a period of, say, 10 years. Accordingly, the agency's experience, number of trained personnel and other assets should be weighed carefully in relation to the advantages of combining censuses of several economic sectors into one programme.

55. The specialized technical and administrative skills required for the execution of many of the functions in a data collection and processing organization argue for their being kept together in one place rather than dispersed among subject-matter divisions. Thus, we generally find computer operations centralized in a computer services division, regardless of the subject-matter to be processed, the field division set up as a separate entity, a single publications division, an industrial directory division, and so on. Manual editing and coding functions are centralized in a data preparation division, with subunits therein for each major industry sector, such as manufacturing.

56. All these divisions, however, are supportive of the subject-matter division (manufacturing, mining etc.), which occupies the central role in the conduct of the industrial censuses. The many and varied functions performed by the subject-matter division consist in planning the censuses; consulting data users, determining subject-matter content; designing questionnaires and instructions to respondents; preparing specifications for processing reported data; resolving data-collection problems; planning the publications and writing texts; and reviewing the tabulated data. The staff performing these functions must be of a high calibre. The basic skills required are a thorough knowledge of statistics, economics, mathematics, accounting practices and last, but not least, a subject-matter knowledge of the industries.

E. Organizing advisory committees

57. At an early stage, the planning staff for the census should assemble two major advisory groups of data users. One committee that should be formed is a governmental advisory group. Government departments that might be represented

include the ministries of industry, external trade, labour, finance, social insurance and other agencies charged with promoting or regulating industry or with economic planning and analysis. The officials from these ministries on the committee should be in high positions, so that they can speak authoritatively for their agencies. Also represented on the committee, or on a subordinate group at the working level, should be technicians who use the census data in their day-to-day operations and who understand the nature of statistical programmes and their limitations. These government representatives should provide advice on just what information on the industrial sector is required in the public interest and for the effective functioning of government.

58. The second advisory group should consist of non-governmental users and providers of industrial statistics. This group should include top representatives of individual companies, trade association officials, heads of accountant groups, labour organization representatives, economists and statisticians from the academic community, as well as representatives from market and statistical research organizations and consumer groups. A determined effort should be made to select committee members who have a first-hand knowledge of data needs and survey work and can therefore contribute meaningfully to the planning of the census. The individuals chosen should be able to give guidance on the general pattern of the appropriate questions relating to the establishment as a whole and how these data can be related to one another and to commodity statistics, both on the input and the output side. Their views on the level of detail for the materials and product inquiries will also deserve serious attention by the statistical agency.

59. As the census planning progresses, it is often desirable to form ad hoc government or private sector committees to deal with specific issues or problems, such as the harmonization of the external trade and domestic production data. A carefully selected group of individual, company and trade association representatives should be assembled to advise on the availability, in establishment records, of the data being considered for the census and on the optimum wording of the inquiries and accompanying instructions. The meetings of these groups are usually conducted on an informal basis and the committees are terminated as the problems they consider are settled.

F. Publicity programme

60. Advance information about the industrial census will afford the business community the opportunity to prepare for the census and thus facilitate more timely reporting. This, in turn, leads to a more efficient census-processing operation and earlier publication of the results.

61. The publicity programme for the industrial census has the threefold objective of (a) encouraging the business community to lend its support during the data-collection phase of the census; (b) informing the public of the availability of published census reports; and (c) promoting the use of census data. Before, during and after the mailing of the census questionnaires, the statistical agency should distribute press releases and news stories to business and trade publications and industrial associations. These articles should stress the benefits of census data to the business community and the guarantee of confidentiality of the individual returns, while referring to the legal requirement for reporting. The trade publications and trade associations might also be provided with booklets containing sample report forms to be used in the censuses of the manufacturing and

mineral industries and the electricity, gas and water industries. The accompanying text would point out that, by reproducing the forms in their own releases, the news media and industrial associations would be enabling industry members to set up their records or reporting systems in time to file the census returns. As a minimum, the Government should issue a press release describing in some detail the content of the inquiries for the forthcoming census.

62. Announcements and order forms for obtaining the summarized results of the census should be widely circulated just before the printed reports become available. It is useful to the business community and the public if the final totals can also be provided on microforms and computer tapes. An interesting recent innovation in some industrialized countries has been the initiation of conferences of data users in cities throughout the country, at which information is presented and questions answered on various aspects of the economic censuses.

G. Decisions on scope, coverage and classification

1. Scope

63. A decision on the industries to be covered in the industrial census is fundamental because it determines the funds and personnel required and burden on respondents. The international recommendations for an infrequent industrial inquiry 9/ call for coverage of establishments engaged primarily in mining, manufacturing and electricity, gas and water supply (major divisions 2, 3 and 4 of ISIC). Those planning the census may be faced with proposals, from advisory committee members or other organizations, for coverage of additional activities as well, particularly the distributive trades and services (ISIC major division 6 and parts of 8 and 9). The census planners should consider very carefully the feasibility of such an expansion of the scope of the census because of the additional effort and resources required.

64. However, for a number of countries, especially developing nations just beginning their industrialization, the border line that separates manufacturing from the distributive trades and services is sometimes very difficult to determine in practice. For these countries, the issue of more budget and staff resources may be outweighed by the practical problems encountered in attempting to isolate one sector from the other when there is an extensive overlap.

65. Some countries may judge the combined approach to be more advantageous, that is, the "economic" census approach, including both manufacturing and mining and the distributive trades and services. There are distinct advantages in this approach, especially when field canvassing is contemplated or a directory is readily available, although the additional effort should be carefully weighed in terms of possible significant delays in the publication of the census results or even the potential failure of the whole endeavour.

9/ See International Recommendations for Industrial Statistics, Statistical Papers, Series M, No. 48 (United Nations publication, Sales No. E.68.XVII.10).

2. Statistical unit

66. It is recommended that the practice of the previous Programmes should be continued in the 1983 World Programme by designating the establishment as the statistical unit, that is, the unit to which the information collected relates. The reports of multi-establishment enterprises may originate with the headquarters of the enterprise but a separate report should be submitted for each establishment within scope of the census.

67. It is useful to supplement the establishment-type data with statistics at the enterprise level to show the various economic characteristics of the firms that own or control the establishments. Enterprise statistics can emerge as a statistical by-product of the regular census programme through a regrouping of the census data for establishments under common ownership or control. This linkage of establishments with their parent enterprises is accomplished by means of the company identification numbers assigned to them in the industrial directory operation (see chap. IV below). On the basis of the characteristics of the assembled establishments of the enterprises, codes are assigned to each enterprise to describe its industry classification, employment size and other economic characteristics. Tabulations of the enterprise data are made by combining the single-establishment company figures with the company totals derived in the computer for firms with two or more in-scope establishments.

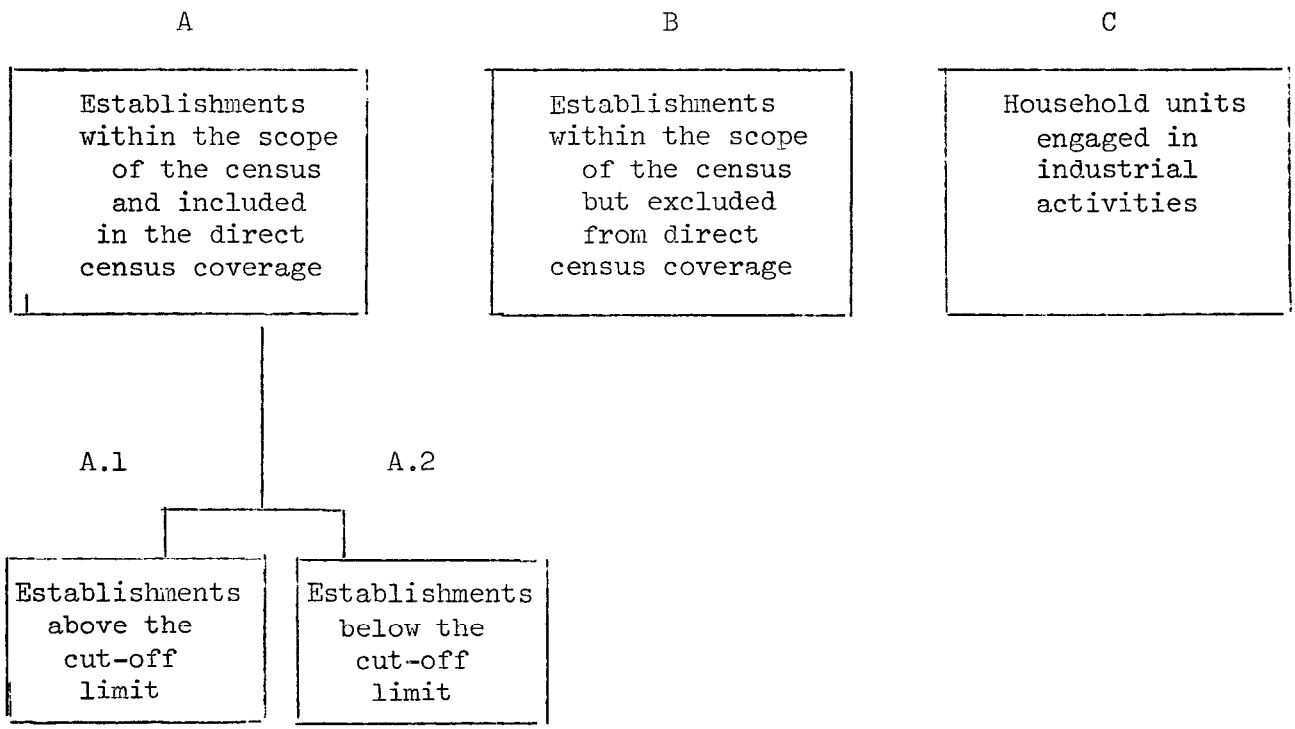
68. Certain types of economic data are only available at the enterprise level, as is the case for financial and related statistics (such as income, expenses and profits from the income statement; debt, liquidity and other balance-sheet items). An enterprise statistics programme provides the statistical link between the establishment data published in the censuses and financial or other enterprise-level statistics gathered either by the census agency or by other agencies of the Government.

3. Coverage

69. For the 1983 World Programme, it is recommended that countries should attempt to achieve full coverage of all recognizable industrial establishments, although there should only be a few basic items of data to be gathered by developing countries whose industrial statistical programmes are in their early stages. Full coverage does not imply, however, that a complete enumeration is required. It does mean that the enumeration should be such that satisfactory estimates can be prepared for the universe. A number of options are open for attaining this goal. The enumeration plan might call for coverage of all establishments above a designated cut-off size and a sample of the smaller establishments. Alternatively, all units might be canvassed and abbreviated short forms used for the smaller establishments; or data for the small-unit segment of the universe might be estimated from administrative records. Where a size cut-off is employed in the enumeration, it should be based on the characteristics of the country's industrial structure (size distribution of establishments), budget for the census, and degree of reporting burden (availability of the information in establishment records). For the 1973 Programme, the proposed cut-off for international comparability was set at establishments with five or more persons engaged. This recommendation was not generally implemented; several countries selected different cut-off points, such as 10 or 20 persons engaged (or employees), often tying in with the cut-off point used in their annual industrial inquiry.

70. Industrial activity in households, usually of an artisan or handicraft nature, should be covered wherever it accounts for a significant proportion of the total output. The level at which the operations of such household units become a significant factor should be determined by the individual country, on the basis of economic and social considerations, the difficulties likely to be encountered in obtaining reliable data and the cost involved. Household industries may be covered by a special area sample survey, based on the population census and conducted either concurrently with or immediately following the canvass of the regular industrial establishments (see the discussion of household-type units in chap. IV below).

71. The units performing industrial activities in any country may be grouped in schematic form as follows:



On the basis of the cut-off size selected, establishments in A.2 will either be covered with short forms, be represented by a sample, or be estimated by other means, such as data from administrative records. Establishments in B are those that may be engaged in activities within the scope of the census but which for reasons of location, ownership or type of operation are excluded from direct coverage (establishments owned and operated by the Government; itinerant establishments with no fixed place of business, such as small sawmills in rural areas etc.). Information from other government agencies or other sources may be used as the basis for showing the activities of these units as a supplement to the census totals. Industrial activities in households (C above) may be the subject of a separate survey of the type described earlier, or incorporated as part of the recurrent sample household survey if one exists in the country.

4. Classification systems

72. For industrial census results to be useful, the reported data must be classified in meaningful arrangements. Therefore, classification systems need to

be developed for the three principal methods of classifying establishments, that is, by industry, by location and by size. Classification systems are also required for products and materials. The census report forms must provide the information necessary for classification purposes. With respect to industrial classification, this is accomplished through the inquiries on products made and materials used, but predominantly the former.

73. To be both useful and practical, the classification systems should deal with the requirements for classifying data, the information that can be collected for classification purposes and the coding problems that might arise in using the systems. While the classification systems should make distinctions that are economically significant and are important to most data users, these distinctions should not be so precise that the information needed for coding is very difficult or impossible to collect. The classification schemes should encompass all of the industrial activities carried on in the country and, since they may be used for many years, they should be flexible enough to allow for future changes in the economy.

74. The classifications must provide an unambiguous category for each unit. For example, in a commodity coding structure, the definition of each category must clearly differentiate between the commodities included there and those included in other categories. Similar considerations apply to all coding structures; if this principle is not followed, the coding operation will be much more difficult and time-consuming, because questionable cases will frequently arise that cannot be resolved by coding clerks. The provision of residual categories ("other" or "not elsewhere classified", (n.e.c.)) at appropriate points in the industrial and commodity classifications will accommodate unforeseen units not otherwise provided for in the classifications.

5. Classification by industry

75. Uniformity in the industrial classification used is essential for the international comparability of statistics. The 1968 version of the International Standard Industrial Classification of All Economic Activities (ISIC) 10/ is recommended to countries for use in the 1983 World Programme. The industrial classifications used by many countries that have recently developed their own systems have been based on ISIC, and other countries have modified their existing systems to make them comparable. In order that internationally comparable results can be obtained, the classification scheme for a particular country should be such that the categories at the most detailed classification level in the national scheme should coincide with, or be subdivisions or combinations of, the individual groups of ISIC. The industrial classification system developed for a country should take into account the technical progress and level of statistical and economic development of that country. For example, the industrial classification structure for a low-technology country will give greater emphasis to certain types of industries than a classification designed for high-technology countries.

76. Indexes to ISIC 11/ are available for use in developing detailed coding manuals for the clerical coding of establishments by industry or for incorporation into reference tapes where the industry coding is done in the computer.

10/ Statistical Papers, Series M, No. 4, Rev.2 (United Nations publication, Sales No. E. 68.XVII.8).

11/ Indexes to the International Standard Industrial Classification of All Economic Activities, Statistical Papers, Series M, No. 4, Rev.2, Add.1 (United Nations publication, Sales No. E.71.XVII.8).

77. The industrial classification system should not specify different categories for activities commonly carried on in the same establishment. Conversely, two activities of economic significance that are ordinarily performed in different establishments should not be combined in one category. In the first instance, coding may be difficult or impossible and it may be necessary in the end to combine the two groups for publication. In the latter case, statistics for establishments in essentially different groups will be combined in one group, and whatever economic distinctions between the groups are significant will not be ascertainable from the census tabulations. If the international activity classification is followed, problems of this nature need not arise.

6. Classification of products and materials

78. The classification system for products and materials should be linked directly to the system for classifying establishments industrially. An ISIC-based system for the coding of industrial products and materials used is simple and effective, in that it admits of a direct numerical relationship between commodities and industries. Such linkage greatly facilitates the tabulation of the census data and readily permits the computer derivation of the industry codes for individual establishments.

79. Classification systems for products and materials should be compatible with the "International Standard Classification of All Goods and Services (ICGS) (draft)" (E/CN.3/493) 12/ and the Standard International Trade Classification, Revision 2. 13/ Another source for developing product and material categories is available in the List of Selected Products and Materials, 14/ created as a guide for requesting commodity information in the 1973 World Programme.

7. Classification by location

80. A location classification system should follow a country's administrative or political subdivisions or designated statistical/economic regions. The system will have to define the boundaries of the units comprising the classification. The statistical agency will need to develop the location reference files necessary to assign location identifiers to the source data collected. In developing a location classification system, where one is not already in use by the statistical agency, the classifications by areas which may be in use in other government agencies should be taken into account.

12/ To be issued as a United Nations publication.

13/ Statistical Papers, Series M, No. 34, Rev.2 (United Nations publication, Sales No. E.75.XVII.6).

14/ Recommendations for the 1973 World Programme of Industrial Statistics, Part II, Statistical Papers, Series M, No. 54 (Part II) (United Nations publication, Sales No. E.71.XVII.16).

8. Industrial classification of enterprises

81. Countries supplementing their establishment-type data with statistics at the enterprise level will need to develop a standard enterprise classification system. Its use will facilitate the tabulation, presentation and analysis of data relating to enterprises and will promote uniformity and comparability in the enterprise-level statistics of the various government agencies.

82. The structure of an industrial classification of enterprises should follow closely the system used for industrially classifying establishments, thus providing a means of linking establishment and enterprise statistics. However, the four-digit ISIC structure is not practical for enterprise statistics, since many enterprises encompass establishments in several industries. An enterprise classification system is usually structured closer to a three-digit ISIC major group or combinations of three-digit major groups. In the coding operation, an enterprise should be assigned to the industrial classification which includes the plurality contribution of its establishments as measured by value added or employment.

H. Decisions on methods of enumeration, processing and tabulation plans

1. Methods of enumeration

83. Once the industries and establishments to be covered have been determined, an important part of the census planning is to decide on the method of enumeration. Answers should be given to such questions as: To what extent will the data be gathered by field enumeration? What portion of the universe will be covered by mail? How will addresses be obtained for mailing purposes? To what extent will sampling be used? Representatives of both the planning and the processing staffs will have to work jointly on these questions.

84. There are two basic methods of enumeration employed in censuses: self-enumeration, in which questionnaires are distributed to and received from respondents by mail; and enumeration by interview or personal visit. In practice, various combinations of these methods are normally employed, although personal visits and enumeration by interview usually predominate, especially in developing countries. Some examples of enumeration methods are given below:

(a) Enumerators visit the industrial establishments and fill out the questionnaire at once, based on answers given by the establishment owners or managers - this method is suitable only if the great majority of the establishments are small and the questionnaire very simple, so that answers to the questions are not difficult to obtain;

(b) Enumerators visit the establishments, explain the purpose and importance of the census, and leave the questionnaire for the respondent to complete; at a later date, reached by agreement, the enumerators return and pick up the completed questionnaires: they may also review them for completeness at that time;

(c) Questionnaires are mailed to the respondent and field representatives call for the completed forms at a later date;

(d) Questionnaires are mailed to the respondents, who complete and return them to the statistical agency at a later date; this approach is the most common one for countries with developed statistical programmes.

2. Methods of data processing

85. With the continuing advancement in the design of data-processing equipment, the number of options available for use in processing and tabulating data continue to expand. Various new techniques for handling mass data have been made possible by powerful new computers and related electronic equipment, such as minicomputers and desk terminals. Significant developments in processing methods during the past decade include:

(a) Extended use of computer editing to replace operations previously performed manually;

(b) Substitution of magnetic tapes for punch cards;

(c) Development of automated procedures to control mail-out and follow-up operations;

(d) Computerized location coding of establishment addresses;

(e) Development of a computerized system for work and progress reporting of census operations.

86. It is, however, necessary to insert a word of caution at this point. Too often a decision is made to use sophisticated electronic equipment when it is really not only unnecessary but actually inadvisable. Unless personnel are available who have mastered the tasks of writing the computer programs and operating and maintaining the computer equipment, it would be better to use the simpler systems, such as those based on punch cards and electromechanical equipment. Careful consideration should always be given to the type of equipment available and the feasibility of using it for processing the census data.

87. Based on the final selection of the type of equipment to be used, the size and nature of the clerical and professional staff needed for editing respondent reports and preparing and reviewing tabulations must be determined. A related requirement is the evaluation of equipment and staff requirements for mailing, check-in and follow-up operations and for the establishment and maintenance of a directory of industrial establishments.

3. Tabulation plans

88. Well before the census questions are made final and the questionnaires are printed, the content of the statistical tables should be outlined, consideration

being given to coding and tabulation procedures. Some countries with limited experience have not prepared tabulation plans until the collection of data has neared completion; at this stage they have discovered that certain questions or topics should never have been included in the census, simply because the information obtained could not be coded and tabulated. A detailed consideration of tabulation plans may also reveal weaknesses in the census inquiries, such as categories that are not mutually exclusive, so that it is impossible to interpret the replies, or inaccurately phrased questions that lead to ambiguous or meaningless results.

89. Table outlines should, accordingly, be prepared at an early stage in the planning of a census or survey. Timely consideration of the plans for publishing the census results is also required to serve as a basis for estimating the cost of that phase of the census programme. Such action helps to ensure that adequate funds are allotted to publication work. It is critical that funds budgeted for the publication stage should not be used to bail out other census operations that may have fallen behind schedule and exceeded their budgets. No matter how well a census is conducted, it will be judged on the adequacy, timeliness and quality of the publication of its results.

I. Treatment of confidential information

90. The purpose of collecting and compiling statistics is to make information publicly available. Yet most countries feel that their statistical agencies should not be free to publish all of the statistical information that it is possible to compile from the respondents' returns. Experience in census and survey taking has demonstrated that adequate response and holding collection costs within reasonable bounds can only be ensured if the statistical agency provides the respondent with guarantees of the confidentiality of information reported to the agency. The guarantees should be written into the legislative acts providing for the census and current surveys. Wording along the following lines is frequently seen in these acts: "information furnished under the provisions of this Act shall be used only for the statistical purposes for which it is supplied. No publication shall be made by the Census Office whereby the data furnished by any particular establishment or individual can be identified nor shall anyone other than sworn employees of the Census Office (under oath not to divulge individual information) examine the individual reports".

91. There is a major difference in the attitude of industry, depending upon whether the confidentiality protection is a matter of law or is an administrative action only. Businessmen tend to suspect a situation where the protection depends upon the prevailing policy of a government agency or of particular individuals in that agency. Such policies can be abrogated overnight with a change of agency head or of the political party in power.

92. Traditionally, national statistical offices have been concerned with developing and applying disclosure rules which will prevent the publication of any totals that might reveal the activities of an individual statistical unit. While

the industrial census and related surveys are based upon establishment reports, the analysis of the data, to avoid a disclosure, should be based on the enterprise as the unit. This is necessary because an enterprise has access to all the reports it files and presumably has a more or less accurate knowledge of the number of companies operating in the product lines, industry or region in which it is active. In general data should not be published for the following categories: (a) those in which there are only one or two enterprises; or (b) those in which there are three enterprises, of which one or two account for a very large percentage of the category total. The application of the disclosure rules absorbs many computer hours, as well as a considerable amount of the time of the professional and clerical personnel. However, most countries consider these costs to be necessary in order to induce respondents to report promptly and accurately and, in fact, to be cost-effective in terms of the savings realized in collecting the reports.

93. Some developing countries question whether the confidentiality safeguards given to individual data were intended to deprive national planning agencies of the basic statistics required for economic policy-making. The flexibility they seek can perhaps be found in the principle that the laws or acts relating to disclosures are only for the protection of the respondent; if the respondent wishes to waive such protection, he is competent to do so. Authorizations have been secured from respondents for the transfer of individual data to other agencies for programming and policy purposes. In any country, industrialized or not, the application of the disclosure rules may result in the suppression of statistics for an entire industry or a major geographical area dominated by only one or two enterprises. To avoid the loss of critical data, countries may request permission from the enterprise(s) concerned to publish the figure which would involve disclosure of the enterprise's contribution. Quite often permission is given, but such requests should be limited to special situations where there is a strong justification in terms of the public need for the information. In following a disclosure waiver approach, due regard should be given to: (a) the administrative work-load involved in securing and monitoring the releases; and (b) the danger of revealing individual data reported by others in the same category who do not give their consent through disclosure by differences in the figures shown in one or more of the many tables published in the census.

94. A more direct and economical approach is tnat, when another agency's need for the individual data is foreseen, the survey be made a co-operative venture and the questionnaire have joint sponsorship or contain the statement that copies of the individual respondent reports will go to the other agency. Even this technique should be used sparingly, because, in countries where it has been employed, legal questions have arisen as to whether any statutory guarantees of confidentiality could follow the data even though both agencies wished to have them apply. Furthermore, any co-operative venture with the revenue or other regulatory authorities would have a seriously detrimental impact on the statistical undertaking.

Chapter IV

INDUSTRIAL DIRECTORY; SMALL ESTABLISHMENTS AND HOUSEHOLD UNITS; SAMPLING; PRE-TESTING

A. Industrial directory

95. A directory of industrial enterprises and establishments, industrially classified and maintained by a single agency, is essential for the successful implementation of a programme of industrial statistics, whatever method of enumeration is used. A reliable frame is critical to the operation, whether the inquiries are carried out by complete coverage, sampling methods or a combination of the two or whether the canvass is by mail or field enumeration.

96. If a mail canvass of all establishments is planned, a complete directory is absolutely indispensable. Even where a complete field enumeration is planned, the cost of compiling a directory may well be more than offset by savings in subsequent operations, such as the verification and control of enumeration. If they are available early enough, data compiled from the directory will also assist greatly in the development of efficient plans for the census. Information in the directory on kinds of activity can be used to decide the number and types of questionnaires, information on physical location can guide the geographical allocation of the statistical agency's resources and knowledge of establishment size can be used in sampling and coverage decisions. In later stages of the census operation, the directory can serve as the medium for imputing data for non-respondents.

97. Ideally, the industrial directory should be a universe control file covering all sectors in ISIC. For practical purposes, the scope of the directory may have to be limited to those industrial sectors for which the particular country conducts bench-mark censuses or annual inquiries. In any case, there should be one unified file from which to control the conduct of the inquiries. The annual inquiry should be integrated into the base control file and actions emanating from the annual survey during the year prior to the census year will determine the universe for the bench-mark survey.

1. Content of the directory

98. As a minimum, the directory should include the following items of information:

(a) Name and physical location of the establishment;

(b) Mailing address, which may differ from the physical location;

(c) For establishments which are part of multi-unit enterprises, name and address of the central office or headquarters of the enterprise;

(d) Establishment identification number;

(e) Industrial activity description and code;

(f) Size code, which should normally be based on the number of persons engaged;

(g) Source and date of the information.

In the directory, an enterprise which consists of only one establishment is referred to as a single-unit enterprise. Enterprises which consist of more than one establishment are referred to as multi-unit enterprises. The directory should include all industrial establishments (or all within the scope of the economic censuses) linked to the enterprise level.

2. Preparing the directory

99. The usefulness of a directory depends on the kinds of data it contains, its completeness and its accuracy. In considering possible sources for the directory, their quality is important, because the directory can be no better than its source material. There are a number of different sources for setting up a directory, most of which have some deficiencies. Each source must be examined carefully before being used and attempts made to overcome any defects. At the same time, it may be necessary to sacrifice some degree of completeness or accuracy in order to keep the costs of establishing the directory at a reasonable level. The following are some of the sources for industrial directories:

(a) A complete field listing. This is probably the most expensive way of collecting information needed for a directory. It is undoubtedly the most useful, however, especially when a country is initiating an industrial statistics programme. Trained field enumerators seek out each physically recognizable industrial establishment and collect the necessary information by direct interview and observation.

(b) Government records. These are records maintained by the Government for taxation purposes and for the administration of unemployment insurance, social security, licensing, regulation or other government programmes.

(c) Trade association or trade publication directories. Such directories tend to be incomplete; for example, trade association records are usually limited to association members only.

(d) Other potential sources. These include telephone directories or special listings prepared by telephone companies and lists prepared by private directory companies, chambers of commerce etc.

100. Except for a complete field listing planned and conducted by the census staff (see para. 99 (a) above), the use of these sources requires a careful examination of the listings for completeness, accuracy and suitability for census purposes. The other sources cited (see para. 99, subparas. (b)-(d) above) may relate to enterprises rather than to establishments. The locations given may be those of central offices or sales offices rather than of industrial establishments. They may be out of date, in the sense that they include obsolete listings and exclude enterprises formed recently. They may be incomplete simply because insufficient effort was expended in compiling them. They may contain an

/...

unacceptable number of erroneous addresses or descriptions of industrial activity. They may have no indication of the size of the establishments and may thus be inadequate for the selection of samples. As a rule, the lists will not be mutually exclusive and, used in combination, will involve the difficulties of sampling from overlapping frames. Finally, they may include establishments which are outside the scope of the census but are not so identified in the records.

101. Information on the extent and nature of the deficiencies in a given set of records may often be obtained from the organization which compiles and maintains those records. The comparison of one register with another which is presumed to cover the same industry or area should also reveal defects. And finally, it may be feasible for the industrial census staff to conduct a complete enumeration of a small area or two and use the results as a standard for gauging the adequacy of the various sources.

102. The directory should be set up using one record for each establishment in the form of plain cards, punch cards or computer tape. For multi-unit enterprises, there should also be a record for the headquarters or central office and each establishment should be cross-referenced to the central office. A separate record for each unit permits maximum flexibility in the elimination of records for establishments or enterprises going out of business, the addition of new records for those entering business, the introduction of changes into the register and the sorting of records, alphabetically or numerically, by kind of activity and geographical area.

103. Unless the number of establishments is very small, the directory listings should be recorded on computer tapes or punch cards. These methods make it possible to sort the listings quickly for special purposes; for example, in initiating a current survey it may be necessary to obtain the names and addresses of all establishments engaged in a specific activity. A master list should be printed out so that the directory can be reconstructed if any of the cards or tapes are lost or destroyed. Computer capability offers much in the way of speed and flexibility. A systems design for computer processing of the directory information may be predicated on a variable word-length record which permits additional information to be added as desired. Changes in the content of the file can, therefore, be made at any time. A computerized system will also have the capability of providing a number of customized size codes to meet unique sampling needs where standard size codes are not satisfactory.

104. Through establishment and enterprise identification numbers it is possible to link together and identify the affiliation of parent enterprises, subsidiary firms and their establishments throughout all phases of economic activity. The identification number, termed an "establishment identification number", is composed of a specified number of digits, depending upon the size and nature of the directory. In one type of identification system, the set of digits consists of two blocks or groups, the first of which identifies the enterprise (both single-unit and multi-unit). In the case of multi-unit enterprises, the same set of digits is common to the identification numbers of all the establishments that comprise the multi-unit enterprise. The second block of digits in the

identification number uniquely identifies each separate establishment. The linkage of establishments and enterprises through a unique "establishment identification number": (a) ensures that there will be no disclosure of data for individual enterprises by grouping all related companies and establishments under a master number; (b) facilitates the centralized mailing, collection, and correspondence relating to all the report forms for an enterprise; and (c) facilitates the collecting of data on an enterprise basis in addition to providing the means of summing data for the establishments of each enterprise. Generally, the statistical agency relies on the individual enterprises to identify their complete company structure.

105. Alternatively, some countries do not choose to link the establishment to its parent enterprise through the establishment identification number but include the enterprise links on computer reference tapes. This procedure has the advantage that the physical location does not change its identification number simply because the ownership changes. Future studies by census statisticians or outside researchers will be benefited if the identification numbering system is designed in a manner that will permit the tracing of establishments through time in spite of changes in ownership that might have occurred. Still other countries prefer a numbering system that indicates the establishment's industry and size classification.

3. Maintaining the directory

106. Countries which plan to follow the industrial census with annual or more frequent inquiries should make every effort to keep the directory up to date in years subsequent to the census. There should be some provision for supplementing the directory with new firms and for reflecting business deaths and changes in establishment ownership or location. The task will be simplified if maintenance of the file is planned from the beginning. The annual and other current surveys conducted by the statistical agency constitute a significant source of information on changes in ownership and the establishment composition of multi-units. The file of single-unit enterprises may be kept up to date by means of information from the Government's licensing authority, social security or other administrative programme on new enterprises, enterprises going out of business and changes in the ownership or location of establishments. Sometimes minor changes are required in the other agency's procedures in order to provide the information needed to maintain the directory.

107. A number of countries maintain their multi-unit file by transcribing the information in the directory to special forms, which are then mailed to the enterprises. The listings of establishment locations are corrected and updated by the multi-unit enterprises and returned by mail to the statistical agency. These enterprises are also asked to supply, for each newly acquired establishment, data on type of activity and number of employees.

108. It is necessary, therefore, to build into the system a method for incorporating the types of changes described above. The method should have the capability to produce listings or computer tapes of the specific changes made.

4. Using the directory

109. Once the directory is compiled, there are many ways in which it can be useful in the bench-mark census, depending upon the enumeration plan. If a sample of small industrial establishments is to be selected, the directory cards or computer tape for the small establishment portion of the universe would be segregated and every n^{th} card or establishment computer record selected, beginning with a random start. If the directory is on tape, a computer program can be written which will select every n^{th} small establishment record without prior sorting of the file.

110. If the plan calls for a direct mail canvass of certain types of establishments, such as those operated by multi-unit enterprises, the appropriate records can be printed out to form a mailing list for the designated group. The balance of the directory, or the entire directory if no mail canvass is planned, should be sorted out by area and printed out so that the field supervisor of each area can be given a list of all the establishments in his or her area to be enumerated.

111. The census may use separate questionnaires for different industrial activities, or one basic questionnaire with supplementary pages to cover specific activities. Such supplementary pages might consist of a preprinted list of products and services for each broad group of industries, with code numbers and space for the responses. Counts developed by tabulation of the directory can form the basis of an economical and efficient printing order for questionnaires or supplementary pages, since the order can be restricted to the number of forms actually needed by the establishments plus an allowance for incidental requirements. In a mail canvass, the directory can provide information for the selection of the proper questionnaire to mail to each respondent. Where, as is common among developing countries, the statistical office lacks information on the activity and size of establishments, a standard form is often used and the respondent (or interviewer) writes in the description of the activity, products and materials. These descriptions are then coded to industry and commodity classifications by the statistical office.

112. In addition to its extensive use in censuses, the directory provides: (a) a mailing list for use in conducting current surveys; and (b) a common frame for use in selecting statistical samples.

B. Coverage of small establishments

113. Many difficulties will be encountered in the collection of census data from small industrial establishments, which by their nature are usually hard to locate and identify. In industrialized countries, most of the production is concentrated in large establishments; often three quarters of the total output is accounted for by less than 20 per cent of the total number of establishments. The statistical agency in those countries can centre its efforts on securing adequate reports from the larger establishments, giving less attention to smaller units.

114. In developing countries, on the other hand, small establishments and households may be a significant factor in total industrial activity and make proportionately greater contributions to certain industries. Consequently, in such situations, effective methods have to be devised to gather a basic minimum of information from the considerable number of geographically widely dispersed small establishments.

115. In deciding upon the best method of covering small establishments, a number of points need to be kept in mind. Small-scale operations can often be managed with very little in the way of systematic records and often the small-establishment owner must be persuaded to estimate a substantial part of the data. Income and expenditure data are seldom segregated by type and the small businessman may find it difficult to distinguish expenditure for fixed assets from purchases of operating supplies. The situation is further complicated by the fact that industrial production in small establishments is often intermittent or carried on simultaneously with other activities, such as retailing or selected services. It is especially difficult to draw a line between manufacturing and repair services in many small establishments. Also, some small businessmen may not be able to supply aggregate figures covering a whole year, particularly value figures.

116. These are formidable problems and the success or failure of the industrial census in developing countries may depend on their solution. A number of approaches to the coverage of small establishments in the census are described below:

(a) Sampling. As discussed above, the census of small establishments may be based on a sample rather than on complete coverage. When a complete directory is available, the sampling may be accomplished by the application of stratification procedures. In the absence of a directory, area sampling will be necessary. In practice, when a directory is available, a combination of the two techniques is often used. An efficient area sample will limit the number of small establishments that need be contacted and has the advantage of lower travel costs over a random sample design. The personnel required to conduct the census may be considerably reduced if coverage of small respondents is confined to a sample of moderate size, such as 10 per cent or 20 per cent. In some countries, funds may be available but the lack of qualified staff may rule out the collection of data from the entire universe.

(b) Imputation. When a small proportion of total cutput (usually less than 5 per cent) is accounted for by the small establishments, coverage may be restricted to establishments above a minimum size and values for the small establishments estimated. In this procedure, loosely termed "cut-off sampling", the establishments above the cut-off point constitute the universe, which is usually enumerated completely. Since cut-off sampling is not a probability-type method, the reliability of the published aggregates cannot be measured precisely in terms of standard errors. Even though the overall percentage accounted for by small establishments may be known accurately, the percentage applicable to individual cells is usually unknown and can be expected to vary widely. Where the

importance of the small establishments below the cut-off level is known approximately, it is reassuring to know that a large error in estimating their output may have only a small effect on the total, as is shown in the following example.

	Actual output	Estimated output	Percentage error
Large establishments	980,000	980,000	-
Small establishments	20,000	30,000	+50
Total	1,000,000	1,010,000	+ 1

In this example, even though the output of the small establishments was overestimated by 50 per cent, the effect on the total was only 1 per cent. This percentage is probably within the tolerance limits expected, even in complete enumeration. Of course, the effect on the totals for smaller published data cells will be greater.

(c) Short forms. Short questionnaires for small establishments have been used very successfully in many industrial censuses. Such forms might include only the items basic to the census, such as the location of the establishment, a description of its industrial activity, the number of persons engaged, wages and salaries and the total value of products and services sold. Although it should be possible to obtain adequate answers to these questions from most small establishments, it may be necessary to modify some of them if advance pilot studies or discussions with small respondents point to serious reporting problems. The use of the short form, however, raises the problem of what should be done with statistics appearing only on the regular questionnaire. There are various possibilities. If the aggregate contribution of small establishments to the statistics in question is not large, their share may be imputed prior to publication. In computerized tabulation, it has been found possible to allocate such imputations to the individual small establishment records, thus making them readily available for geographical and other cross-classifications. The contribution of small establishments to some aggregates, such as that dealing with the generation of electricity, can probably be imputed as zero. In other cases, it can be imputed at a fixed ratio to total output, or at different ratios depending on the industrial activity of the establishment. If the contribution of small establishments to particular items omitted from the short-form questionnaires is large, a sample of small establishments can be selected for long-form coverage. The collection of long-form data from the small respondents included in the sample might then be entrusted to specially trained enumerators or to field supervisors. Further refinement of the results might be obtained by imputing values for these items for all other small establishments using ratios based on the sample results. Another type of imputation arises when the short form calls for the same information as the long form, but in less detail. As an example, the short form might request the total amount expended for new fixed assets, while the long form might require separate figures on expenditure on (a) buildings and other construction, (b) transport equipment and (c) machinery

and other equipment. While the distribution between the three items could be estimated for small establishments, on the basis of the proportions reported by large establishments, a better procedure might be to publish the short-form data separately. The captions for this purpose might read:

Total expenditure on new fixed assets:

(a) Buildings and other construction;

(b) Transport equipment;

(c) Machinery and other equipment;

(d) Expenditure on new fixed assets, kind not reported.

Item (d) would consist largely of short-form data, but might also include data from large establishments which were able to report the total but not the detail. It would be desirable to show in a footnote the amounts from each source, as well as to provide a measure of the size of the small establishment universe such as the number of persons engaged. At this point, consideration should be given to the eventual processing of the short-form statistics and the imputed information that would be required through its use. The level at which imputations are to be incorporated (individual establishments or aggregates), and the approach to be used for each item not covered, will affect subsequent operations.

C. Household industries

117. The sector comprising household manufacturing, mining, trade, transport and services accounts for a substantial proportion of the gross domestic product of many developing countries. Developing countries are committing themselves to promoting the activities of this sector in order to provide the sizeable segment of the population dependent upon such activities with more employment opportunities and improved earning capacity. This development effort is, however, seriously handicapped by the scarcity of information and several developing countries are looking to household surveys to fill the vital data gap.

118. Household units (that is, units where artisan, handicraft or other activity classifiable as manufacturing or mining is carried out on household premises) should accordingly, be covered as part of the industrial census if their production of goods and services for sale or exchange accounts for a significant percentage of total industrial activity. The determination of what is significant may vary from one country to another and may hinge upon whether household units are dominant in certain industries, such as sawmilling, furniture-making or metal-fabricating, which are important in the country's economy. As a rough "rule of thumb", 5 or 10 per cent of the total activity is suggested for coverage of household industries.

119. In evaluating the feasibility and cost of measuring output from household units, a number of points should be brought into sharp focus. Building up a

register of household industries is a very difficult task, because most of such units are unregistered with any government agency; many of them are mobile in nature and have no fixed location. Lists become out of date quickly, owing to the high proportion of new and discontinued businesses; the turnover can run to as much as 20 per cent a year. Books of account are rarely kept by household enterprises; where they do exist, the household and business activities frequently cannot be separated with any degree of accuracy. The valuation of inputs may present problems; raw materials may consist of discarded and waste materials not bought in the marketplace. Response is further hampered by a lack of confidence in the privileged character of the information reported and concern lest it be made available to the taxation authority.

120. Decisions on the method of covering household units, when their role justifies such action, have to be made by each country in the light of its individual circumstances. Household industries may be surveyed as part of a country's regular household survey or in a special survey designed for that purpose.

121. Countries with an ongoing multi-subject household survey system could profitably integrate the coverage of household industries into that system. Information on the degree of area concentration of such activities could be obtained through the population census. Households would be subjects for canvassing when the population census reply to "What kind of work was this person doing?" described a manufacturing or mining activity and the individual described himself or herself as "Self-employed on household premises". Industrial activity in households consists in purchasing raw materials, or obtaining them by farming, fishing, mining, or salvaging and reclaiming waste or discarded items, then performing some processing or fabricating and selling or exchanging the resulting products. "Homeworkers" should be distinguished from household units engaged in industrial activity. The term "homeworkers" designates individuals whose fabricating activities are directed by a "regular" industrial establishment which supplies the raw materials, dispenses contract payments to the homeworkers and sells the resulting product. Homeworkers should be omitted from the household industries survey; where such activity is significant, it should be reported in the questionnaires filed by "regular" industrial establishments, payments to homeworkers being shown separately from plant pay-rolls.

122. The population census information on industry and occupation can be used by arranging the area segments in the country's current household survey according to the degree of concentration of household industrial activity and selecting a systematic sample of those segments. That sample could be augmented by an additional sample of other area units in the country with a high degree of concentration of household industry. Countries not conducting a current household survey might conduct a special survey with a sampling design totally oriented towards household industrial activity.

123. For covering household industry, an area sample offers the best solution; usually a relatively small sample will suffice. In an area sample, certain specific areas are selected and these areas are completely enumerated. The savings in travel costs alone compared with a random sampling plan are very great;

no travel to the non-sample areas is necessary, for example. A possible plan for an area sample of household industries may be summarized as follows:

(a) The areas to be sampled would be enumeration districts, as defined in the most recent population census;

(b) All of the country's counties (or appropriate political or administrative subdivisions) would be classified into three groups, based on the proportion of households with industrial operations, as revealed by the population census;

(c) For counties with little or no industrial activity in households, an extremely small sample of enumeration districts would be selected. There should be a larger sample of counties having a moderate degree of such activity. The largest sample of all is to be chosen from the group of counties that indicated a substantial amount of industrial activity in households in the previous population census.

124. As a special survey not integrated with the country's regular household survey, the household industry canvass can take place either concurrently with or directly following the field work in the industrial census. In any event, the field organization for the household survey should consist of the individuals who work on the regular industrial census. To a considerable extent the interviewing skills required are much the same. The industrial activity in many households will resemble that of the short-form establishments in the regular census. The household industries survey will be affected by a similar lack of records, a lack of respondent understanding of census concepts and a failure to comprehend the need for industrial statistics. The experience of enumerators in the regular census becomes extremely useful in the conduct of the household industries survey.

125. In addition to other issues, there is the problem of choosing the reference period for the sample survey, particularly where household industrial activities are seasonal in nature, which is often the case. Furthermore, short reference periods are necessary in order to reduce reporting error owing to the respondent's inability to recall activities with the passage of time. Several developing nations have expressed the view that the only feasible method of measuring household activity is in terms of a short reference period. The limited experience with household industry surveys has shown that 30 days preceding the date of the survey is an effective reference period.

126. On the other hand, ways must be found of reducing the bias resulting from projecting data for short reference periods to annual estimates. To lessen the effect of seasonality in deriving annual estimates and to reduce the size of the sampling error, the technique of moving samples - the survey spread over the entire year with samples distributed over different months - can be advantageously used.

127. In enumerating household activity as an adjunct to the regular industrial establishment coverage, the watchword should be to keep the survey as simple as

possible. The sample size should be held to a statistically acceptable minimum of observations consistent with predetermined publication goals and standards of accuracy. Consideration should be given to omitting households with industrial activity below a certain level based on receipts. The information to be obtained from households should be called for on a specially designed short form. The questionnaire content should be limited to a very few inquiries; households cannot be expected to maintain records of the type kept by business concerns.

128. It is recommended that only the following four questions should appear on the household industries questionnaire:

(a) Physical location of the household unit;

(b) Number of persons engaged;

(c) Value of receipts;

(d) Description of the goods made or services performed (for example, knitting sweaters, sawmilling, furniture making).

129. In addition to national estimates, the household survey should produce estimated totals for the provinces and individual industries or industry groups in which household industry is a significant factor. These estimates should be published even though they are accompanied by relatively high sampling errors. The sampling errors (standard errors of estimate) should be published together with the statistics.

130. In concept, the information on industrial activity in households could be obtained in a country's population census. In practice, this does not usually prove to be feasible. The population census is generally taken several years apart from the industrial census, to which the household industries data should be closely related. Just as importantly, the specialized and detailed nature of the household activity inquiry rule against its inclusion in the generalized population census. However, as mentioned previously, that census can provide an efficient sampling frame for conducting an area canvass of households to measure their industrial activity.

D. Sampling

1. Use of samples

131. A basic consideration in any economic census is whether to use sampling in the enumeration and processing or to conduct these operations on a complete coverage basis. In taking a survey, either bench-mark or current, the survey manager attempts to design the survey so that the required data will be produced at a minimum cost, both to the respondent and the statistical agency. Since sampling is one method by which costs can be reduced, it is advantageous to review

each programme to determine to what extent and in what manner this technique can be used. 15/

132. The use of sampling can be seen in almost every phase of the census programme and in the more frequent inquiries as well. In addition to the possibilities of sampling the regular industrial establishments by one of the methods discussed below, a small area sample is likely to be the best means of canvassing industrial activity in households where coverage of such units is part of the bench-mark census effort. Also, samples are frequently used in pre-tests conducted prior to the census, in quality control of the census processing and in the verification of check-in, coding and editing operations and data entry.

133. Sampling, rather than complete coverage, is usually selected as the method of conducting the annual industrial inquiry, which is the core of an integrated system of statistics on industry. The sample design adopted for the annual inquiry by many countries consists of a stratified sample with selection proportional to size; that is, a sample that includes all establishments engaging more than a specified number of workers (say 50 or 100 persons) but smaller proportions of the other establishments. The sample selection procedure should give each establishment in the sampling frame an independent chance of selection. While this design does not ignore the contribution of the smaller plants to total activity, it greatly reduces the number of respondents. This reduction in the number of returns, often in a ratio of four or five to one, is sufficient to bring the annual survey within the bounds of a manageable undertaking, even though most of the returns included are the more difficult ones to process. Some countries elect to canvass only those establishments above the cut-off size and impute data for the remainder based on administrative records or past censuses or surveys of the statistical agency. The annual survey sample panel, or subsets of it, can be used for special surveys on subjects not covered by the annual inquiry. Samples, either of the cut-off or probability type, are also employed in current surveys of employment and earnings, acquisitions of new fixed assets, stocks and the production of individual important commodities.

2. Advantages of sampling

134. Because the size of a sample means that fewer reports need to be collected and processed, the use of sampling has a number of advantages. Among other things it relieves many respondents of the burden of reporting and reduces the costs and personnel requirements of the statistical agency. This lower agency workload permits a greater degree of control over the collection and processing of reports and therefore results in an improvement in the quality of the statistics produced. In addition, earlier publication of results may also be possible. The statistical agency is able to adjust the sample size to the funds available. If funds are insufficient for a complete census, sampling may be the only feasible method of conducting the census.

3. Disadvantages of sampling

135. The use of samples introduces additional steps, not present in complete canvasses, which must be carefully and properly executed. These involve some cost and create opportunities for mistakes that do not occur with complete coverage. The sampling system used should be a probability sample, which yields results of measurable quality. However, a disadvantage of probability sampling is that it requires careful control over the selection of the sample units, and the application

---

15/ A detailed discussion of sampling techniques applied to industrial surveys appears in annex I.

of sampling weights. Otherwise, errors can be introduced which would not occur in a complete canvass. The difficulties of administering a sample increase in proportion to the complexity of the sampling plan. If there is a choice between two plans, one more theoretically desirable but also more elaborate, requiring more skilled personnel and detailed records than may be available, the more efficient plan may well be the wrong choice.

136. Another drawback of sampling is that the amount of reliable detail that can be published will inevitably be reduced because the smaller number of observations available for each cell of data results in an increase in the sampling variance. However, where cost considerations are of overriding importance, it may be necessary to forgo some of the industry and location detail and confine the collection effort to a sample.

137. Sampling is not a magic formula for success. The collection, editing and tabulation problems that arise with complete coverage will most probably also occur when a sample survey is conducted. Sampling will not eliminate the problems, but they will be fewer in number because there will be fewer respondents and publication cells.

4. Stratified sampling

138. Where there is a directory that contains information on size and kind of industrial activity, it is possible to select a stratified sample based on these characteristics. In the application of stratification procedures, the totality of sampling units included in the frame is divided into groups or strata, each stratum being sampled separately and independently, so that a specified number of sampling units or a specific fraction of all such units is obtained from each stratum. The usual procedure in industrial censuses is to subdivide the universe into two or more groups defined by size of establishment. For example, a frequently used method has been to distinguish between establishments with less than five persons engaged and those with five or more. To achieve greater homogeneity, the size groups are sometimes further subdivided by industrial classification, but a more common practice is to adopt the same sample rate for all establishments within each size group. This has the great advantage of simplicity. In the example cited, the sample might consist of all cases in the top size class (complete coverage) and a 10 per cent sample of the smaller establishments. Whatever the stratification plan, it is absolutely essential for an efficient probability sample in the industrial field to include all of the very large establishments. The sample might then be selected by arranging the establishment cards or computer record in some systematic way, such as alphabetically, numbering them serially, and then beginning with a random number from 1 to 9, selecting every tenth card for the sample. This method is called random systematic sampling. The selection of a two-group stratified sample of the type described might be appropriate where, first, there is a large number of small establishments and, secondly, they account collectively for a substantial proportion of the universe in terms of output or persons engaged.

139. Sampling based on stratification by size is most appropriate when the enumeration is conducted by mail. If field enumeration is employed, area sampling is nearly always more efficient, the principal reason being the resultant saving in cost of travel. In order for either stratified or area sampling to have validity, it should be a probability sample with a known probability of selection, so that the sampling variance can be calculated. The reliability of data obtained by

sampling procedures depends largely on the number and type of tabulated cells and cross-tabulations to be developed. A general rule to be followed in balancing survey costs against reliability is that the larger the sample in relation to the universe being measured, the smaller the industry category or area for which reasonably accurate data can be tabulated.

5. Area sampling

140. The term "area sampling" refers to the use of areas as sampling units. The areas may be administrative units, naturally bounded areas (such as "segments", "blocks", etc.), or especially demarcated areas. The areas observed may, therefore, be selected by reference to a map (which constitutes the frame) or determined in the course of the operation. Samples of this type are appropriate in industrial censuses where a complete directory of establishments is not available or where there are large numbers of small establishments to be covered by field enumeration. Area samples can yield adequate national totals, but publication of statistics by areas will usually not be possible except perhaps for the areas included in the sample. Area sampling should be employed only to cover small or medium-sized establishments. The largest establishments, including those located in non-sample areas, should be enumerated completely. In the simplest form of area sampling, all of the establishments within each of the sample areas are enumerated. Second-stage sampling, or sampling within areas, is also feasible and can be employed if some of the sample areas contain unusually large numbers of small establishments, or consist of a number of smaller defined areas such as villages. The secondary sampling units could be either villages or establishments, selected by random methods.

E. Pre-testing

141. Pre-testing is the use of statistical methods to obtain facts about reporting or collecting problems; it can be stretched to include all research about this range of problems. In a narrower sense, pre-testing is trying out on a small scale alternative ways of collecting data, one of which will later be used in the statistical project. Similar methods are applicable to internal processing problems. In its broadest terms, pre-testing involves testing and retesting a topic or procedure.

142. Pre-testing can save resources and result in a better census. The pre-testing programme should be carefully planned to ensure that each test has specific objectives and that the results are used to improve the census. It is unwise to invest all pre-testing resources in one large test; it is preferable to undertake a series of pilot tests and identify and correct deficiencies after each test.

143. The principal problems of measurement peculiar to a country's economy, as well as problems associated with the wording of the questionnaires and instructions, may be brought to light in advance of the census by means of small-scale pilot studies or informal pre-tests. Excessively long and complex questionnaires add to the cost of all subsequent stages of processing, and unrealistic inquiries can absorb much energy before they are finally abandoned. Adequate pre-testing of questionnaires will go a long way towards avoiding such problems.

144. No questionnaire, whether intended for mailing to respondents or for use by enumerators in the field, should be printed in final form without first having

been subjected to a brief but thorough test under actual survey conditions. In the case of a census involving field enumeration, the pre-test procedure would be to send trained interviewers into the field to try out the draft questionnaire on a cross-section of establishments. In testing the viability of the questionnaire, the pilot study sheds light on the respondent's ability to complete the inquiries and on the length of time needed for the interview. The results of the pre-test aid in determining personnel requirements and in estimating collection costs. The pre-test also pinpoints areas needing improvement in terms of the wording and format of the questionnaire. If questions of interpretation or definition arise, these should be followed by adequate training of enumerators to deal effectively with those problems in the field. Finally, the field test provides valuable guidelines for the development of a framework for field editing and control of questionnaires. It is desirable to conduct the field pre-test well before the final questionnaires are printed.

145. In determining the suitability of questionnaire items for censuses conducted by mail, the preferable pre-test technique is to send a prototype report form to a representative group of establishments for completion. The response is then analysed by the census planning staff before the content and wording of the inquiries on the items are made final. This type of pre-test can be used with equal effectiveness for both comprehensive censuses and annual or more frequent surveys. Whenever a new item is considered for inclusion in a continuing survey, it should be thoroughly pre-tested before it is printed on the questionnaire.

Chapter V

MANAGEMENT AND BUDGET

A. Introduction

146. If an industrial census is to be successful, meticulous attention must be paid to each phase of the census operation. The planning and technical aspects of a census are important, but organization, budgeting, management and quality control are equally critical factors. The items discussed in this chapter are closely interrelated and decisions on one aspect affect several others. When the publication objectives of the programme are clearly stated, they provide a background against which the right decisions can be made. The following publication goals are suggested for countries participating in the 1983 World Programme of Industrial Statistics:

(a) The preliminary results should be published no later than two years after the end of the reference year to which they apply;

(b) The final results, incorporating both industry and location detail, should be issued within three years of the end of the reference year;

(c) The preliminary statistics for key data items, such as employment, output and value added, should not contain error levels exceeding 5 per cent for national totals at the industry sector level or more than 10 per cent at the provincial or industry-group level.

B. Establishing time-tables

147. An initial step in setting up a realistic work plan is to establish a time-table for the completion of each operation during the census period. Besides a description of the operation, the time-table should show the starting and finishing dates and estimates of cost and personnel requirements during each week or month. Breaking each operation down into suboperations contributes to more realistic estimates and makes it easier to determine the budget for the census. Since the object of the census is to produce data the first dates to be set are those for the publication of results. Once these dates are established, it will be necessary to work backwards through all the operations that are needed to achieve that goal. After the appropriate key dates are determined, it will be necessary to see what needs to be done to meet the schedule. The operations calendar should present, directly or in supporting documents, as much detail as possible on each activity - what organizational unit is responsible, the specific tasks to be accomplished, what skilled and other personnel and equipment resources are needed, when they are needed and where and what allowances should be set aside for contingencies.

148. In order to set a date for the completion of an operation, it is necessary to determine when the operation must be started so as to be able to complete it on time and what other operations need to be finished before the operation in question can be initiated. It can then be determined whether sufficient lead-time has been

provided to avoid delaying subsequent steps in the census. If later on there is some slippage in meeting key dates for critical phases of the census, subsequent operations will need to be re-evaluated and dates changed or resources shifted in order to keep the survey on schedule.

149. A statistical survey can be thought of as a series of interdependent actions leading to the publication of data. Each action stems from and flows into another action. The relationship between these actions can be presented either verbally or diagrammatically. Both methods can be effective.

150. One effective way of visualizing the management of an industrial census or current survey is by listing the major operations and indicating beginning and completion dates for them in a series of columns representing calendar months or quarters of the year. This type of presentation will reveal any overlapping operations, as well as operations that require rescheduling because they will not be completed when needed. Such a presentation can also be the basis for setting work-loads by time-period and, therefore, of indicating personnel and equipment requirements by time-period.

151. Another useful tool for establishing work programmes is the flow chart. Diagrammatic presentation in the form of a flow chart has the advantage of being more concise than narrative material and of making it easier for many people to follow the flow of work. Flow charts also make it easier to pinpoint illogical steps or the omission of necessary steps. A one-page flow chart displays the same amount of information that it would take many pages to describe in writing.

152. Adherence to time-tables may be monitored by utilizing PERT (Programme Evaluation and Review Technique: critical path analysis) methods. PERT is an early-warning tool, which, by displaying graphically the impact on the project completion dates of earlier stages falling behind schedule, offers the programme manager an opportunity to adjust upcoming intermediate steps so that the end dates can still be met. PERT can be helpful in pointing out areas of operation where lost time may be recovered.

153. Several paths may be followed in using good scheduling to stretch available funds and resources. For example, since many operations in a census are sequential, some of the same personnel can be used at different stages of the census. Much of the training and experience gained in one operation is transferable to later operations and results in better and more efficient work. Also, a statistical agency that is engaged in processing a number of different surveys, both demographic and economic, will find that its use of electronic and mechanical equipment will fluctuate widely unless it is carefully planned. Efficient scheduling can take advantage of periods of low usage and result in a general reduction of costs. This is particularly true with respect to the use of computers and peripheral equipment.

C. Budgeting

154. An important feature of census planning is the development of a sound budget. Without a clear, realistic idea of what each step will cost, all the available funds could be spent long before the work is finished. All elements of the industrial census operation - from enumeration through processing to final tabulation and publication - must be considered in determining the budget. The preparation of step-by-step cost estimates, and their consolidation into larger totals, representing intermediate phases of the entire project, is what is known as budgeting.

155. Each organizational unit (division, office etc.) that is responsible for a specific function should prepare a cost estimate, by quarter, covering how much and what type of personnel and equipment will be needed to achieve the stated objectives of the census. In the preparation of these estimates, each unit should divide the major operations for which it is responsible into suboperations for which the time and quality of performance objectives should be indicated. This procedure may reveal potential bottlenecks or excessively costly operations. When all the estimates are aggregated, it is possible to comprehend the over-all picture of cost, personnel and equipment requirements by period.

156. In preparing the cost estimates, each unit should make certain that the budget statements are clear and understandable, that they include all operational steps, and delineate areas of responsibility, and that the cost estimates are realistic. Persons with the administrative or supervisory responsibility for carrying out an operation should participate in preparing the budget estimate for that operation. Although the budget should be prepared in full detail, a summary is usually sufficient for the government officials who will approve the project and release the necessary funds.

157. One of the major budgetary considerations, in terms of the labour costs involved, is the method of enumeration. For population censuses, the field enumeration costs in some countries amount to one half or more of the total budget for the census. While the magnitude of the number of industrial establishments is wholly different from that of the number of households, field costs in any field-canvassed economic census amount to a substantial proportion of the total census budget. It is especially important to estimate carefully how many interviews per person-day can be held, since that constitutes a frequent reason for underestimating resource requirements.

158. A significant element in each cost estimate is the overhead charge. As a rule of thumb, overhead costs in industrial censuses may be estimated at 15 per cent of the direct cost of each operational step. This percentage covers an appropriate portion of the salaries of individuals in the office of the director of the statistical agency, of heads of the agency's divisions and of others who are called upon for occasional work on the census. The overhead charge also includes a proportionate amount of incidental expenses, such as office supplies, communications, transportation, informational services and other miscellaneous expenses not charged as direct costs.

159. Throughout the census period, the budget needs to be re-examined frequently and compared with actual performance. Some functions will be found to have been incorrectly estimated and the budget should be revised accordingly. A carefully prepared budget should set aside between 10 and 20 per cent of the total budget for unexpected costs arising from unforeseen problems and contingent events.

D. Progress reporting system

160. Once the census objectives have been set, a system of progress reporting must be installed. The purpose of that system is to enable management to determine the current status of work and expenditure in any operation at regular intervals during the conduct of the census. The system should provide all the data required by managerial and supervisory personnel to control the progress and flow of work, to control expenditure, and to keep all phases of the project within the time-table and cost estimates.

161. Progress should be measured in two ways - units produced and costs incurred. The measurement of functions such as data entry at keying stations or clerical editing and coding is comparatively easy. Because the activity is fairly uniform for each of the units processed, output and backlog can be stated in terms of such units. Each person engaged in such activities can be judged on the basis of his or her production rate and evaluated by comparison with the production rates of other clerks. Conversely, such functions as preparing editing specifications or writing and testing computer programs do not lend themselves so readily to measurement. However, it is still necessary to monitor them closely, because it is usually not possible to compensate for falling behind schedule by assigning additional staff, since highly specialized personnel are involved.

E. Quality control

162. In addition to keeping operations on schedule, it is very important to maintain acceptable quality. Without control of the quality of the data throughout all the census stages, the entire census effort would be wasted.

163. The concept of quality control is concerned with maintaining quality at an acceptable level. "Quality" refers to the degree to which a specific procedure or service satisfies the objectives of the specific user. "Control" refers to the planned cycle of activities by means of which intended objectives and standards are attained. Quality control and operations control should be closely interwoven to produce total control of the census.

164. In order to achieve this goal, a quality control unit should be established to design a verification system of various operations that will determine that acceptable levels of quality are being maintained. The responsibility for the quality control function should be assigned to personnel trained in statistical methods. The staff of the quality control unit should be independent - that is, it should not be an integral part of the unit it is supposed to be monitoring. Before quality control procedures are put in place they should be reviewed by the processing and planning staffs.

165. In some quality control procedures a sample of cases will be verified. In others, 100 per cent verification is necessary. In specialized activities such as map evaluation or review of questionnaire content, quality depends on the experience and judgement of the professional staff rather than a formal verification procedure.

166. Quality control systems should be designed to check on such activities as the assembly of computer-prepared mailing lists, field operations, clerical coding and editing, table posting and replies to correspondence. Wherever mass clerical operations are conducted, it is important to establish a set of standards against which the production of individual clerks can be measured. These should be reviewed periodically and clerks whose output, in quantity or quality, consistently falls below acceptable levels should be retrained. A comprehensive quality control programme may even include the conduct of a post-census inquiry to check with a sample of establishments the extent to which definitions and other instructions have been complied with and the accuracy of the industrial classification given to the business.

F. Organization of the census

167. The organizational structure of a statistical agency may take different forms in different countries, giving varying emphasis to subject-matter or to function in the divisional arrangement. Listed below is a prototype organization for taking an industrial census, together with a brief description of the responsibilities of each division.

Subject-matter division (manufacturing, mining, electricity, gas and water supply). Maintains liaison with the Government and private data users. Develops census questionnaires. Determines procedures for clerical operations. Resolves data-processing problems. Writes computer programs for processing and tabulating the data (alternatively, this function could be assigned to the computer services division). Plans and develops the census publications.

Field operations division. Directs those segments of the census that are conducted by personal enumeration through data-collection centres in the field.

Industrial directory division. Establishes and maintains a control file of all units within the scope of the industrial census and surveys conducted by the statistical agency; assembles sets of questionnaires and instructions for mailing or for distribution to field offices.

Data preparation division. Performs large-scale clerical and manual operations such as pre-computer editing and coding of returns; clerical review of computer-edited reject records; posting of worksheets and tables for publication.

Computer services division. Operates and manages the electronic digital computer and mechanical tabulating facilities, including the co-ordination and scheduling of computer processing, staging and tape-library services.

Administrative services division. Secures space, supplies and equipment. Arranges for communication, transportation and related services.

Budget and finance division. Responsible for accounting, pay-roll, financial planning and control; co-ordinates budget estimates and their justification.

Personnel division. Formulates personnel policies, recruits personnel, provides guidance in applying personnel regulations, classifies jobs and maintains personnel records.

Publications division. Supplies art, editorial, forms design, copy preparation and printing co-ordination services for census questionnaires and publications.

It is recognized that individual countries will select different organizational structures and the divisional alignment and allocation of responsibilities listed above should only be regarded as indicative.

168. At the very beginning, a census planning group should be organized to set out the basic policies and guidelines to be followed throughout the programme. That group should address itself to making the early decisions needed concerning the reference period and content of the census, legislative and budgetary requirements, organizational responsibility, method of enumeration, method of processing the data, publication plans and so on. The planning group might be comprised of the director of the statistical agency, or his deputy or other designated representative, the industrial census co-ordinator (discussed in para. 169 below) and the chiefs of the subject-matter division and other divisions with major responsibilities in the conduct of the census such as field, data preparation, computer services etc. Once the census operation actually starts, the members of the planning group are expected to be available for consultation on special problems and periodically to review and assess the over-all programme.

169. Regardless of how well the census is planned, the fact that there are so many activities and that responsibility for different phases of the census may be scattered among different divisions results from time to time in bottle-necks in the work flow, the need to resolve conflicting opinions about a course of action, the necessity of shifting personnel, the changing of priorities, the adjusting of budgets etc. To deal with situations such as these, a person with sufficient authority should be appointed as industrial census co-ordinator to co-ordinate the census operation and to anticipate and resolve problems and conflicts. As co-ordinator, he should be authorized to call meetings of personnel representing different divisions and to conduct negotiations aimed at keeping the programme on schedule and within budget.

G. Personnel requirements

170. In selecting and recruiting the personnel to carry out the census operations, the following types of experience, skills and training will normally be found useful:

Educational background in statistics, economics, accounting, business administration and related subjects and training in the operation of industrial censuses in a country offering such assistance;

Experience as a statistician, planner or supervisor in other statistical projects, not necessarily in the field of economics or industry;

Background in setting up classification systems or in applying them to coding operations;

Experience in supervising field collection of data of any kind; or in other field work, such as door-to-door selling;

Experience in budget planning, in purchasing of equipment and supplies and related management functions;

Experience in managing general office and clerical functions;

Experience in computer programming and systems analysis;

Experience in supervising the operation of electronic computers and peripheral equipment.

171. The persons assigned to planning specific phases of the census should later be responsible for carrying them out. This helps to ensure continuity of planning and completion of the planned operations. It permits the training and experience acquired in the planning stages to be used to the fullest possible extent in the later operations of the census.

H. Equipment requirements

172. There are often considerable delays in procuring computers and related equipment, office machines and specialized supplies such as punch cards. It is desirable, therefore, to arrange to purchase, rent or borrow equipment as soon as the kinds and numbers of machines required can be approximately determined. Timetables for the census operations will provide information on when and for how long the equipment will be needed. The number required is more difficult to determine, because for each operation it will be necessary to estimate the total number of physical work units, such as questionnaires, to be processed through the equipment, and the normal capacity of the equipment in terms of number of work units per hour or day. The latter figure should be obtainable from the equipment manufacturer or can be estimated with the assistance of the manufacturer. Allowance should be made for time lost because of a breakdown of the equipment or delays such as those due to failure of the preceding operation to deliver a sufficient number of work units ready for processing. The length of time the equipment will be needed may not be as critical a factor if it is purchased, or rented on a week-to-week basis with the right to return it at any time.

Chapter VI

QUESTIONNAIRES AND OTHER DATA-COLLECTION FORMS

A. Introduction

173. The data-collection forms are the key documents in the industrial census, since they provide the means for recording all the establishment data that are to be tabulated and published. Their format, organization and content will have a significant influence on the quality of the statistical results, the cost of collection, editing and tabulation, and the promptness of publication. In all successful industrial censuses, therefore, considerable time has been devoted to the design of questionnaires and other data-collection forms. Ambiguous phrases, printing errors, and even poor spacing, have caused some countries to abandon important inquiries or to expend scarce resources in an effort to correct the effects of such mistakes. In self-enumeration, when the respondent fills out the questionnaire without the assistance of an interviewer, even a poor selection of type-faces has been known to cause widespread misinterpretation. This situation can arise because a key heading, important for understanding the captions below it, does not stand out boldly and is consequently overlooked by the respondent. Specimen questionnaires for the minimum and complete programmes appear in annex II to the present publication.

174. Pre-testing of the data-collection forms can often bring such deficiencies to light in time to correct them (see paras. 141-145). The early enumeration of one or two districts, for example, may demonstrate the adequacy or inadequacy of collection forms. If alternative types of questionnaires are proposed to cover the same group of respondents, pre-tests can help to indicate which is the preferable type to use.

B. Principles of questionnaire design

175. The physical form of the questionnaire - for example, whether it is to be printed on card-stock or paper, or whether it should be assembled as a continuous sheet or booklet - should be determined after consideration of all the uses to which it will be put. If it is expected that many of the questionnaires will be filled out on typewriters, the continuous sheet or accordion-fold type is preferable to the booklet. The spacing between the lines on such a questionnaire should correspond to typewriter spacing.

176. The size of a questionnaire is determined by the number of questions, the extent of instructions to respondents placed on the form, the size of type, and the amount of space required by the respondents to enter the data. A form should be large enough to accommodate a print size that can be read easily and have sufficient open space for an attractive appearance. At the same time, the form should not be so large that it is clumsy to handle. If several forms are involved in a census, it is desirable to have them of uniform size for ease in handling.

177. The questionnaire should be designed in a manner that makes it convenient to use, both for the respondent (or enumerator) and for the statistical processing

office. For example, the inquiries should appear in a logical order on the form. Related questions should be grouped together (for example, employment and earnings, fixed capital formation, inputs and output - see discussion of content of the Programme in Part One of the present publication). If the information is to be edited, coded and punched directly from the form, the arrangement and spacing of items should also take these operations into account.

178. Each report form should carry a unique form number and letter prefix. The prefix "IC" might be used on all the 1983 Programme questionnaires, the "IC" standing for "industrial census". Forms should be numbered in series or blocks of numbers to identify them as to the specific census (that is, manufacturing, mining etc.) and as to whether the form is long or short (for example, "10" for "manufactures: long" and "100" for "manufactures: short"). Each form should also show, under the form number, the date when the form was sent to the printer.

179. The paper on which the questionnaire is printed should be a colour that is distinctive and easy on the eyes. When several censuses are involved, the use of a different colour for each census facilitates the selection of the appropriate form by the enumerator or respondent.

180. If the information is to be recorded by an enumerator, as in field canvassing for an industrial directory or gathering the census data, a good quality sheet of paper or a large card on which data for many establishments can be recorded may be the most efficient type of questionnaire. The durability of the paper or card-stock used should be based on the amount of handling the questionnaire will receive.

181. In addition to the spaces for respondent entries, it is a common practice to include spaces on questionnaires for statistical office entries, labelled "For office use only". Such entries are of three general types:

(a) Control data such as establishment identification number, identification of the enumerator and others processing the questionnaire, and the dates on which various actions in collection and processing take place;

(b) Code numbers such as those for location, industrial activity, size of establishment and type of ownership;

(c) Calculated figures such as the average number of persons engaged, average earnings and value added. These figures may be intended for tabulation or merely for use in a pre-punch review of the questionnaires for reasonableness.

Some or all of the coding and calculation may be performed by the computer after the data have been punched and put on tape, thus eliminating the need for spaces on the questionnaires.

182. Each data item on the questionnaire should be numbered to facilitate reference to the item in the respondent instructions and correspondence. When more than one type of report form is used, every attempt should be made to retain the same number

for the same item, even if this means that on some forms an item number will have to be included with no data requested but marked "Not applicable to this form". This technique will simplify the writing of editing and clerical procedures and facilitate the writing of tabulating programs.

183. Also, a key code should be assigned to each item of data or other information that will be tabulated or that is necessary for the identification and classification of the establishment. The type of code can vary depending on whether computers or other kinds of equipment are to be used. The proper assignment of key codes can make a significant difference in the cost and quality of the census processing and should be done with a knowledge of how the specific item will be processed and tabulated. For example, if punch-card equipment is to be used, care should be taken to ensure as far as possible that those items of data that will appear in the same table are included on the same punch-card. There should be uniformity with respect to the placement of code boxes, and the kind of type face used.

184. A common problem in the processing of report forms occurs when a respondent fails to make an entry for an item for which the response could possibly be "zero" or "none". The editor is then faced with the dilemma of determining whether the blank entry is correct or whether the respondent inadvertently failed to supply the appropriate data. It is therefore desirable to include an instruction on the first page of the questionnaire that tells the respondent to make an entry of "zero", "none", or "-" whenever such entry is appropriate.

185. Because names and addresses change from time to time, the address box should leave sufficient space for the change to be indicated. In addition, each questionnaire should provide adequate space for the respondent to enter remarks that are necessary to explain any of his or her reported data. It is also helpful to request the name, address and telephone number of the person to contact regarding the report. This information saves time and effort in follow-up correspondence on missing or questionable entries. Finally, good survey practice calls for space to be provided for the signature of an authorized official or representative of the enterprise following the statement on the form that "This report is substantially accurate and covers the period from _____ to _____".

C. Inclusion of self-editing questions on forms

186. A number of countries have found the accuracy of response to be improved by the inclusion in the questionnaire of an inquiry with a number of self-editing questions. The questions are as a rule placed immediately after the last of the data inquiries and immediately before the space for remarks. The respondent is told in the headnote to the item that "The National Statistical Office reviews your report for omissions, inconsistencies and unusual ratios". On the basis of saving the respondent the trouble of later correspondence concerning the report, he or she is asked to make a number of reasonableness checks of the figures after completing the report but before returning it to the enumerator or census field office. A sample self-editing inquiry is included in the specimen questionnaire for the complete programme.

D. Instructions for respondents

187. The instructions for respondents should be clear, concise and complete. Instructions which are long and diffuse are less likely to be read and understood than those which are brief and to the point. Pre-census interviews with potential respondents often provide valuable indications of the points which should be emphasized in the instructions.

188. As many of the instructions as possible should be placed on the questionnaire itself, close to the item to which they refer. However, rather than overburden a form and make it too long, a separate instruction sheet should be used. In such instances, the most critical instructions should be left on the form itself and expanded or supplemented as necessary on the separate instruction sheet.

189. The following is a check-list of points normally covered in instructions on the questionnaire forms or on supplementary instruction sheets:

(a) Legal authority for the census;

(b) Confidentiality of returns;

(c) Types of establishments required to report;

(d) Acceptability of estimates if accounting records do not supply the required data;

(e) Date the completed report is due;

(f) How to return the completed report;

(g) Time period covered by inquiries; where a full year is required, acceptability of periods other than the calendar year;

(h) Definitions of the various items of information requested; rules for valuation in the case of items of data reported in value terms;

(i) How to report individual products or materials on questionnaires when descriptions are not pre-printed;

(j) Special instructions for particular industries when problems affecting many respondents are anticipated; for example, how to report fish-processing operations when the establishment is also engaged in catching fish.

E. Wording of questions

190. Although the instructions are important and should be carefully written, even greater care should be accorded to the wording of the questions. A major factor in securing prompt, complete and reliable replies is the phrasing of questions so that

they are readily understood by respondents and field interviewers. Some respondents will fail to read separate instructions, but all, presumably, will read the questions before entering replies. Every effort should therefore be exerted to keep the wording of the questions clear and concise, and to avoid ambiguities. "Tailoring" the language of certain questions may be desirable or necessary for particular industries.

191. For some items, check-boxes are preferable to direct questions. In a check-box query, all the possible replies are printed out, and the respondent is asked to check the applicable reply. Such devices are sometimes appropriate for information on the various industrial operations performed in the establishment, or on the methods of distribution of products. They may also be employed for obtaining quantitative data if replies in terms of size classes are acceptable. The alternatives must be mutually exclusive and exhaustive; a common way of ensuring that a list is exhaustive is to include an "all other" category. The value of this check-box type of question is enhanced by the fact that pre-assigned codes can be printed next to each box, thus avoiding a clerical coding operation.

192. The layout and design of the questionnaires (as well as the number of different forms) should be the product of a close working relationship between the subject-matter specialist, the data-processing specialist and the form-design specialist. Each of these persons has a role to play in the census. The subject-matter specialist is responsible for gathering meaningful and accurate information. Generally, his or her decisions on the wording and sequence of items should take priority over other considerations in designing the report form.

193. The efficient and economical processing of the data is the responsibility of the processing specialist. His or her views should be respected on the placement of identification items on the form, the method of pre-coding for machine processing, and the optimum arrangement of items on the form with a view to facilitating the clerical and professional review of the data.

194. It should be kept in mind, however, that a questionnaire design which would optimize clerical and analytical operations may not be the most effective vehicle for the collection of the data. The form-design specialist is concerned with obtaining a good response from the respondent as well as with designing a form that lends itself readily to processing. He or she approaches that goal from the standpoint of eye appeal and legibility, so that the form does not appear burdensome. The form-design specialist strives to see that the wording of the items and their sequence carry out the intent of the subject-matter specialist. If all these individuals perform their tasks well and work together smoothly, the chances of the census achieving its objectives are greatly increased.

Chapter VII

COLLECTION OF DATA

A. Introduction

195. This chapter provides detailed recommendations for the data collection phase of the industrial census, with emphasis on the methods considered most appropriate for countries at an early stage of statistical development. The data-collection procedures described here would be essentially similar, whether complete enumeration or sample coverage were planned. When sampling is used, however, an effort should be made to increase the accuracy and reliability of the basic data by training enumerators more thoroughly, and exercising more care in the subsequent processing phases in order to detect and correct response errors. Careful attention to these points can help to offset the errors of estimate introduced by the sampling method.

B. Enumeration plan

196. Basically, two methods of enumeration are used in industrial censuses: (a) the mail-out and mail-back method; and (b) canvassing in the field, that is, direct interviewing by enumerators. In some censuses both methods are used, each for one segment of the universe. In deciding what system should be used to collect the data, it is necessary to analyse the universe to be covered, the types of information to be collected, the availability of funds, qualified personnel and specialized equipment and, most importantly, the presence or absence of an industrial directory of establishments.

197. In this discussion it is assumed that an adequate and reliable industrial directory is available for the conduct of the census (see chap. IV above). It is also stipulated that the collection of data is based on a field canvass of all single-unit enterprises and a mail canvass of establishments of multi-unit enterprises directed to the central or head office of the enterprise.

198. The enumeration of known multi-unit enterprises by the central statistical agency rather than by field interviewers has a number of advantages. Usually, less time will be taken to perform the enumeration. The field interviewer encountering an establishment of a multi-unit enterprise may spend a long time explaining his mission, only to discover that such matters are handled by the company's head office located in another interviewer's district. In multi-unit enterprises, the head office often insists on supervising the filing of questionnaires and transmitting them when completed. This procedure possesses the advantage for the statistical agency that the enterprise's head office assumes responsibility for co-ordinating the completion of returns for all the establishments of that enterprise and reviewing those returns for accuracy and consistency. When such a procedure is followed, however, care should be taken to establish controls that will ensure that reporting is not duplicated. Interviewers should be provided with a list of enterprises and specific establishments that are not to be enumerated (see para. 200 below, item (2)).

199. In some countries, experienced members of the central statistical agency have undertaken to interview officials of multi-unit enterprises in advance of the census. The purpose of such interviews is twofold: (a) to gain the co-operation of the company officials and impress them with the importance of the census; and (b) to

facilitate the actual enumeration by, for industrial directory purposes, updating the list of the company's establishments and the information on the activities conducted at each location. The updated list will serve as the basis for the selection and forwarding of the proper census questionnaires to the enterprise.

200. When the plan of enumeration calls for a mail canvass of establishments operated by multi-unit enterprises and field enumeration of the rest, the appropriate records can be printed out to form a mailing list, for the multi-unit group. At least four copies of this list should be made: (1) to be transferred to the outgoing mailing pieces by labels or other means; (2) to be given to the field supervisors, who will instruct enumerators not to call on the listed establishments, except possibly for follow-up calls at a later date; (3) to be transferred to a set of check-in cards or computer records which will form the central control record of the receipt of completed questionnaires from the respondents; and (4) a master copy to be retained separately in case portions of the other copies are lost or destroyed.

201. The balance of the directory, or the entire directory if no mail canvass is planned, should be sorted out by area and printed out. The supervisor of each area will be given a list of all the establishments in his area to be enumerated. Copies of this list should also be retained in the central office for check-in purposes and for the preparation of reports on the progress of the enumeration.

202. The completeness and accuracy of the directory should not be taken for granted in the field; the enumerators should be instructed to take note of non-listed establishments and to secure reports from them. Provision should be made for creating new directory listings for such added establishments. The field supervisors may also be instructed to compare the directory listings for their areas with information available from local taxation or licensing authorities, classified directories, and other local sources. This procedure may reveal additional establishments or indicate errors in the existing directory listings.

C. Field enumeration

1. Use of maps in coverage control

203. In field enumeration, steps should be taken to ensure that information is secured from every establishment within the scope of the census, other than those canvassed by mail. On the other hand, it is also necessary to avoid duplication, which can occur when two field interviewers make a report for the same establishment.

204. An important aid to field interviewers in achieving these objectives is a set of detailed maps covering each of the enumeration districts. Such maps should clearly define the boundaries of the districts. If the boundary is a street, the allocation of establishments on each side of the street should be indicated. If accurate and detailed maps are available in the central statistical agency, maps of the enumeration districts can be prepared in advance. Otherwise, it may be necessary for the regional or local offices to define the enumeration districts and prepare maps of them for the interviewers.

205. Enumeration districts defined for censuses of population may serve as a starting point for the definition of industrial census enumeration districts. Estimates of the number of industrial establishments in each of the population enumeration districts should be obtained, and combinations made which would tend to produce districts of approximately equal industrial importance.

2. Enumeration district control record

206. In addition to enumeration district maps, systematic canvassing will be aided by a control record on which is entered identifying information concerning each establishment encountered. The control record may be in the form of a set of numbered sheets, each of which provides for entries concerning 20 or more establishments, and the set covering an entire enumeration district. A typical record of this kind might provide for the following items:

(a) Serial number of entry - in some systems, this number is also entered on the corresponding questionnaire;

(b) Name and address of the establishment;

(c) Industrial activity or kind of business;

(d) Is the establishment within the scope of the census?

(e) Is the establishment part of a multi-unit enterprise?

(f) If a questionnaire is left with the respondent, the date it is to be picked up;

(g) Date the completed questionnaire is obtained;

(h) Date the questionnaire is forwarded to the central statistical agency;

(i) Name of person interviewed;

(j) Remarks.

207. Since the control records provide a detailed account of the progress of enumeration and the disposition of the completed questionnaires, they should be preserved for use as reference documents in later stages of processing. Questionnaires can be misfiled or lost, for example, and reference to the control record will indicate whether the questionnaire file for an enumeration district or a local area is complete.

3. Screening of completed questionnaires by field staff

208. For practical reasons, it is desirable to detect and correct errors in the reported data as quickly as possible after the questionnaires have been completed. If there is a delay in re-interviewing or questioning the respondent, the person who

originally supplied the data may no longer be employed by the company, the company may have gone out of business, or the records may have been discarded. On the other hand, it is ordinarily not feasible or desirable to conduct complete editing operations, such as those discussed in the next chapter, in the field offices. The staff should, however, be able to perform a screening operation which involves a limited amount of checking of data. This might consist in seeing that the questionnaire is properly identified and that no key figures are omitted, and similar simple checks. A slightly more complicated but basic check would be to add the reported costs (wages and salaries, materials etc.) to see that they do not exceed the reported value of shipments and sales of services. If they do, the report should be returned to the enumerator or given to another interviewer for verification or correction.

209. Such a field review also provides a quality control over the performance of individual enumerators. When questionnaires are found to be faulty, the enumerator responsible should be required to correct them, if time permits. An excessive number of faulty questionnaires may indicate an inadequate training programme or that the enumerator is incompetent.

4. Check interviews

210. Another means of controlling and improving the quality of the reported data has been found desirable in some countries. This involves the re-interview or check interview of respondents by the local supervisor or by a specially trained enumerator. Check interviews are ordinarily conducted for only a small sample of respondents. Since the re-interview involves an additional demand on the time of these respondents, the interviewer should be skilled at explaining the purpose of the interview, which is basically to increase the accuracy and usefulness of the census. He should then proceed to secure new answers to the questions. The check interview need not cover all the items on the questionnaire, but can be limited to a few of the most important. Check interviews are of value in disclosing unacceptable practices on the part of enumerators and respondents, such as the fabrication of data.

D. Training of enumerators

211. The field enumeration of manufacturing and mining establishments is quite complex and requires, in addition to the setting up of very tight controls, thorough training of field supervisors and enumerators. The statistical units of an industrial census vary greatly in such characteristics as size, organization and industrial activity. One of the first problems encountered by the field interviewer is that of determining which of the establishments are eligible for enumeration. Accordingly, the calibre of personnel used in the field enumeration must be such that they can recognize industrial activities and have good number sense. Time must be taken to train enumerators properly both in the mechanics of enumeration and in the objectives of the survey. They must be made aware of the consequences of the actions they take - or do not take. Time and again it has been found necessary to re-enumerate substantial portions of a census because instructions have been misunderstood. If there are restrictions on the funds

available for the census, it would be better to reduce the scope or coverage of the census (and therefore, the number of enumerators required) rather than the quality of the enumeration.

212. Ideally, it would be desirable to concentrate the task of training in the hands of a few individuals in the central statistical agency. This has rarely been feasible in practice, however, and the usual first step is to train individuals who will teach others. These may be regional office heads and other key personnel, who are given an intensive background course in the central agency. Later, these persons return to their areas to train local office supervisors and these in turn train field interviewers.

213. One country has reported an interesting alternative, in which a team of enumerators, half of whom were permanent staff members of the central statistical agency, moved around the country enumerating area after area until the work was complete. Unlike a population census, the enumeration of industrial establishments can be spread out over a period of several weeks.

214. At all levels, training classes should be small, so that the teacher can give attention to the progress of each individual in the class. If feasible, simulated interviews, in which an experienced person takes the part of the establishment owner or manager, should form part of the training course. Written tests should be given at some time during the course. Some part of the test may take the form of a description of a hypothetical establishment, based on which the trainee fills out a test questionnaire. Each individual approved for field work should demonstrate an ability to cope with at least the most frequent of the complex technical problems that arise during enumeration.

215. Some countries have been faced with training large numbers of field interviewers in a short period and have found it worth while to prepare detailed class-room guides. These are intended for the use of relatively inexperienced teachers, who are expected to instruct the class by following the guide practically word for word. The use of audio-visual materials has proven to be very effective in training enumerators.

216. There are also examples of home-study reference materials, which are given to prospective enumerators in advance of any class-room instruction. After several days allowed for home study, the enumerator is given a written test. If successful in passing the test, he or she is interviewed by the supervisor. During the interview, understanding of the principal questions is tested again, the questions failed on the written test are reviewed, and a supervisor takes the part of the respondent for a practice interview. Following this, the supervisor accompanies the enumerator on at least one real interview. The enumerator's progress is again reviewed closely after a probationary period of several days.

Pocket manuals

217. A condensed manual of instructions for enumerators, intended to be carried on the person and consulted during the work-day, has been used successfully in many

countries. Such a manual may also serve as a basic text-book for the training course. All of the procedural matters affecting enumerators, and the principal problems they may expect to encounter, should be discussed in such a manual. The subjects covered in a typical manual are outlined below:

(a) Rules and regulations: legal obligations of enumerators to provide true returns and treat information given by respondents as confidential; actions forbidden to enumerators, such as soliciting or selling;

(b) Excerpts from the law or executive order authorizing the census;

(c) Procedures: organization of field staff; how to obtain supplies and transmit completed questionnaires; what to do if respondent wants questionnaires left with him or her for completion or for mailing to statistical agency; what to do if respondent refuses to supply information; how to report establishments which changed owners during the census year; how to prepare progress reports;

(d) The "skip-list", that is, a list of establishments the enumerator is not expected to visit, which may include the establishments of multi-unit enterprises which are to be contacted directly by the central statistical agency: how to identify establishments or headquarters offices of multi-unit enterprises;

(e) Scope of the census: definition of an establishment, and types of establishments to be included; types of establishments commonly encountered which are not to be included;

(f) List of questionnaires and other forms used in the census, and a brief description of the content, purpose and scope of each;

(g) Detailed instructions for each question in the questionnaire;

(h) Special instructions for particular types of establishments or industries: how to handle frequently encountered mixed activities (for example, manufacturing and trade, manufacturing and agriculture); how to enumerate household industries, if they are covered in the industrial census;

(i) Alphabetical index to the contents of the manual.

E. Receipt of questionnaires: field collection

218. In field collection, two sets of receipt records are usually maintained: one set in the local or regional offices and another in the central agency. In some systems, the enumeration district control record described above is also used as another check on the receipt of questionnaires.

219. In the following flow description of a typical questionnaire handling and receipt operation, it is assumed that the central agency deals directly with the local offices, that it is desired to maintain a constant flow of completed

questionnaires to the editing and coding operations, and that the central agency is responsible for the collection of data from multi-unit enterprises.

(a) Interviewer collects questionnaires and prepares enumeration district control record;

(b) Interviewer turns in completed questionnaires weekly to local office; when coverage of an enumeration district is completed, interviewer also turns in the control record;

(c) Local office records receipt of questionnaires in check-in file, screens questionnaires for acceptability, returns some to interviewer for correction, and forwards the rest to central agency daily; date of forwarding is recorded on check-in cards;

(d) With the last batch of questionnaires for an enumeration district, the control record is forwarded to the central agency; date of forwarding is recorded;

(e) Central agency records receipt of questionnaires in check-in card file, and forwards them to editing and coding section;

(f) Questionnaires rejected by editing and coding section are returned to check-in section; these questionnaires are returned to the field for verification and correction; date of return is recorded in the check-in card file and also on the enumeration district control record, if one has been received;

(g) As enumeration district control records are received, they are compared with the check-in card file to see that all the required questionnaires (i) have been received in the central agency, and (ii) have not been returned to the local office for correction; if some of the questionnaires are still being processed in the editing and coding section, this comparison is postponed until editing and coding are completed;

(h) A record of incomplete and completed enumeration districts is maintained in the central agency; as enumeration nears completion, the central agency gives local offices a list of incomplete enumeration districts and asks for a report on the status of each.

F. Receipt of questionnaires: mail collection

220. Collection of the census data by mail is feasible when an industrial directory has been established or a pre-canvass of in-scope establishments has been conducted prior to the census and the postal service in the country is considered sufficiently reliable. There are a number of systems that can be used for controlling the mail-out and receipt of questionnaires, and these are described in the following paragraphs.

G. Mail-out and check-in by computer

221. Under the computer check-in system, the computer logs returns as they are received by entering the date of receipt on a magnetic tape file (the "collection control file"), containing the reference numbers (establishment identification numbers) of units in the mail collection. This file should be maintained in establishment number sequence. Addresses for units with returns outstanding are generated by matching this check-in file against a "mailing address file" containing the establishment identification number, name and address, size code based on number of persons engaged (or employees), and report form number for all in-scope establishments. It is suggested that the check-in file should be updated weekly by returns received during that period.

222. In the mailing operation, mailing labels may be prepared by the computer from the master "mailing address file" and affixed to the questionnaires. The inclusion of the industry classification code for the establishment in the computer record would, among other things, make it possible to sort the establishments by the type of questionnaire to be mailed, in addition to providing a substitute for the report form number (although there are advantages to be gained from entering the report form number as well as the industry classification code on the mailing label). The insertion of a code for the location of the establishment (province, county etc.) would facilitate the mail distribution of the questionnaires and, in a field enumeration census, the location code would constitute the primary sort for the questionnaires in distributing them to the field offices.

H. Mail-out and check-in by methods other than computer

223. A frequently used non-computer procedure is the transfer of the mailing-list information to a set of "check-in" cards. Space should be provided on these cards for the entry of the date the questionnaire was mailed and the date it was received. The same card, if desired, can be used for recording the dates of subsequent contacts with the respondent, such as correspondence to obtain corrected figures or to verify the accuracy of the figures reported.

224. Since cards are sometimes lost or misfiled, a master mailing list should also be prepared. Mechanical transfer methods, such as addressograph plates, make it possible to print lists on long strips of paper, which may then be stapled and bound to ensure completeness in subsequent handling. Each strip should also be numbered.

225. As the completed questionnaires are received from respondents, the corresponding control card is removed from the file. The date of receipt is entered on the questionnaire and also on the card, which is then put in a second file representing establishments for which questionnaires have been received. In large-scale receiving operations, the clerks are sometimes instructed to "check in" obviously incomplete questionnaires, including those returned blank, and to treat letters from respondents as the equivalent of questionnaires. The correction of these problem cases, as well as others requiring further contact, is left to a subsequent screening operation.

226. At some time after the due date for filing questionnaires has passed, the original file, which now contains cards only for the establishments which have failed to return questionnaires or other required material, can be used as a mailing list for a follow-up letter reminding respondents of their obligation to report. Subsequently, a second or third wave of follow-up letters or field visits to respondents who are still in default will probably be necessary.

227. Another system would be the use of printed registers of establishments, in which space is provided for recording receipt of returns, dates of dispatch of reminder notices and the like. A simple control system is possible if address plates are used. These can be held as a "register" and, as completed questionnaires are received, the address plates are removed and stored elsewhere. When it is decided to send reminder notices, the address plates representing returns outstanding can be used directly to address the notices.

I. Special treatment of large establishments

228. For the largest establishments in the "default" file, it may be desirable to schedule field visits at an early date. The preparation of reports by such establishments takes more time, and their importance in the statistical results is much greater than is the case with small establishments.

229. In the census planning, resources should be set aside for the field contact of large establishments that are slow to respond.

Chapter VIII

DATA PROCESSING

A. Introduction

230. After the questionnaires have been received by the statistical agency, it is necessary to review them in order to determine whether they are reasonably complete and accurate. It is inevitable that many errors will be found in the reported information, despite careful planning and collection of the data by an efficient and well trained staff. If for no other reason, errors will occur because the concepts behind industrial census inquiries are relatively difficult to grasp and the questions can easily be misinterpreted. In addition, some entries may be treated as erroneous because of processing requirements: values expressed in common fractions, for example, may be quite accurate by ordinary standards but are not acceptable for data entry systems, either key-to-disk or punch card.

231. The basic purposes of editing statistical questionnaires are: (a) to detect errors in the reported figures; and (b) to prepare the questionnaires for data entry into the machine and tabulation. The former is by far the more difficult process, because many entries may appear to be acceptable, but when compared with other information are seen to be questionable or obviously wrong. The review process will uncover a considerable number of problems and potential errors which must be resolved before the report is ready for tabulation. A substantial percentage of respondents will need to be contacted for verification or correction of data (up to 25 per cent or more of all respondents in several countries experienced in industrial census-taking). Since these contacts are very time-consuming, they should be initiated early and pushed vigorously so as not to delay later processing operations. The preparation of questionnaires for machine data entry and tabulation involves operations such as the rounding of numbers, the striking out of excess figures, and the entering of control data.

232. Since it is inefficient and wasteful of resources to treat all deficiencies in questionnaires alike, standards and criteria for editing should be established. In general, the amount of effort expended on detecting and correcting errors and omissions should be in proportion to their probable effect on the published data. There is a certain amount of safety in numbers: if errors are of a random type, they will tend to cancel each other out provided they are individually small. This leads to the conclusion that a high degree of accuracy and rigid standards are unnecessary refinements when dealing with small establishments. The amount of attention paid to individual reports during editing and coding should vary in direct proportion to the size of the establishments represented. The funds and human resources available for these operations will be more efficiently used if this general rule is observed. It follows that the reports from large establishments should normally be more carefully edited than those from small establishments, except where the coverage of small establishments is based on a sample.

B. Methods of data-processing

233. One of the first decisions the statistical agency must make is how it will process the data reported by respondents. That decision will affect the budgeting, scheduling and design of the questionnaires, and the number and types of personnel needed to code, edit and tabulate the results.

234. If electrical accounting machines are used, most of the editing will be accomplished manually, except for some "zero balancing" or "control checks" done in the machine. Data would be transcribed from the questionnaires to punch cards, which would be sorted and tabulated. Clerks would post the applicable tabulated cells to work-sheets, which would then be typed and printed for publication.

235. If electronic data-processing equipment is used, most of the editing and coding and all of the tabulating will be done by computer. Data can be entered conventionally by punching cards and reading them into the computer. Or, countries are now adopting a computer-type data entry system which operates from computer programs or reference tapes containing instructions to the keying machine. The operator enters data on a typewriter-style keyboard in accordance with images flashed on a screen or "scope" at the eye-level of the operator. For example, one of the computer programs would flash the entire first page of the questionnaire in a format for the operator to enter the data onto magnetic tape. The use of this new type of equipment has significanly reduced the number of errors that occur during the key-punching of the data.

236. The procedure followed by the statistical agency in selecting the appropriate processing equipment can be a simple cost consideration or it can be an involved study with long-range implications. A number of considerations are pertinent and these will vary from country to country. In most countries, the technology of data-processing has developed from punch-card accounting machines to electronic computers of great speed and capability. Wherever punch-card equipment must be operated on more than one full shift every day, the cost comparison will nearly always favour the computer over electromechanical equipment. Benefits other than cost influence the choice of the computer. A computer system introduces a new dimension of ideas into an organization - ideas which are not possible without a computer. Also, in an industrial census, the factor of timeliness can be as important as the cost of the census. Census data lose some part of their usefulness every year. If a computer can help to reduce the time from enumeration to publication of the data, this will certainly be a valid benefit, even though there is no way to measure the benefit precisely. These extra dividends from computers that are difficult to measure may well be more important than savings in direct costs if, in fact, there are any such savings.

237. A number of other factors should be taken into account and dealt with when, as is probable in almost all cases, the electronic computer will be the type of equipment chosen by the statistical agency. Some of them are discussed below.

(a) A computer of sufficient capacity is capable of carrying out many more editing processes than are usually feasible in a clerical operation. The computer

will be able to perform several times as many internal-consistency checks, for example. Some census planners, enthusiastic over this extraordinary capability, have specified so many editing checks that practically every questionnaire has been flagged for manual review. As in clerical editing, the usefulness of every individual check should be considered in relation to its cost and also the overall cumulative effect of a large number of checks;

(b) Many of the questionnaires that are flagged by the computer for manual review will require correction. It is costly to correct punch cards and computer tapes. Hence, if a great many questionnaires will eventually require manual handling, repunching of cards etc., computerized editing may result in a much higher overall cost than clerical editing;

(c) Computerized editing is necessarily performed at a later stage than clerical editing. As mentioned in chapter VII above, it is desirable to question respondents as early as possible after the questionnaires are received. With a temporary field staff, computerized editing may entail so much delay in the return of questionnaires to the field that the staff may no longer be available. Thus the timing of requestioning should be considered carefully in relation to the timing of computerized editing;

(d) An important difference between computer processing and the older methods such as hand and mechanical or electromechanical processing is that misinterpretations giving rise to repetitive or systematic errors are often far more difficult to correct. Many costly tabulations have had to be scrapped because the computer programmer did not quite understand what the statistician wanted. Under the older processing methods, such mistakes were usually discovered and rectified before much harm had been done. The tremendous speed of computerized processing, however, means that entire projects are often completed before the systematic errors are discovered. For this reason, sufficient time should be allowed for the thorough testing of computer programs before large amounts of data are processed. A possible solution is to prepare a "test deck" of punch cards similar to those which will eventually be processed, but which contain errors and other material deliberately planned to bring all of the features of the program into operation. The program can be executed on the test deck at relatively little cost, and the results of the test used to correct the program where necessary. In practice, the production of a completely comprehensive "test deck" is often not possible.

(e) Conversion of the subject-matter specialist's specifications to the computer language used by the computer programmer sets up an additional communication barrier. Unless the specialist can read the programmer's language to make sure that the ideas have been correctly interpreted, there will be room for the introduction of erroneous material. Subject-matter staff need to receive sufficient training to enable them to review the programmer's work effectively. It has been held that programming and testing can be safely left to the computer specialists once they have been briefed and fully understand the needs of the program. This millenium will be reached only when communication between planners, programmers and subject-matter specialists is perfect.

238. Under a computer-based processing system, it is still necessary for clerks to screen reports to determine that key data items have been reported, that there are no obviously gross deficiencies in the reports, and that any necessary work that cannot be performed by the computer is accomplished before the data are entered into the computer. In screening the questionnaires, the clerks also carry out other standard clerical operations, such as converting fractions and decimals to whole numbers and verifying that all data entries on the questionnaires are legible and in the proper positions. Also, the report form should be reviewed for any notes or comments entered by the respondent which may require correspondence.

239. Values are frequently reported in terms of the smallest monetary unit, but it is often not desirable to punch or tabulate the data in that manner. Punching the full number would require additional time and lead to a higher number of errors; tabulating would also require more time and space. Similar problems occur when certain quantities are reported down to the units digit. Part of the preparation of the report for tabulation consists of reducing the number of digits to be processed by rounding the number. Training for this operation should be thorough; many people do not seem to be able to grasp the principles involved. It is easier, of course, to use a procedure that calls for striking digits rather than rounding, but the data involved must be examined to make sure that this procedure does not have a significant effect on the resulting statistics.

C. Editing checks

240. Various types of editing checks are employed in reviewing the industrial census returns; the most commonly used types are described below. The number of editing tests performed and the extent of checking the data depend upon the level of accuracy sought and the time and resources available.

Scope. Scope tests are designed to determine whether the primary activity of the establishment comes within the scope of the industrial census. An establishment should not be classified and tabulated in the manufacturing census unless the value added from manufacturing exceeds that from its non-manufacturing activities. Ordinarily, sufficient evidence for this determination is provided by inspecting the components of gross output (see Part One, para. 139, of the present publication). To illustrate: should the receipts from installation or repair exceed the sum of its receipts from products shipped and contract or commission work, then the establishment is outside the scope of manufacturing. However, where reselling or merchandising activity (that is, goods shipped in the same form as they are received) is reported along with the manufacture and sale of its own products - and this mixed activity is being accepted as pertaining to a single establishment - then the preferred scope test calls for the comparison to be made in approximate value-added terms. To be specific, if the value added from reselling (value of goods shipped as received less cost of goods shipped as received) exceeds value added from manufacturing (value of goods produced less cost of materials etc.), the establishment is ruled out of scope. If the schedule does not contain an entry for one of these cost entries, a good rule-

of-thumb is to treat such mixed operations as being within the scope of manufacturing unless value of goods shipped as received is more than twice the value of the products manufactured.

Processability. As a rule, the tests for processability consist of a series of completeness checks to determine whether entries (numbers other than zero) are present for a selected list of critical data items. In computerized editing, the program customarily rejects schedules that do not contain entries for total employment (or total number of persons engaged), total pay-rolls and gross output (value of shipments plus receipts). The same data fields are also commonly deemed essential to further processing of the schedule if visual inspection is used. Indeed, some countries perform presence checks for these items as part of their check-in procedure applicable to questionnaires received from large establishments, and refrain from marking the check-in card "satisfied" unless entries for these fields are present. Additional items deemed critical in some systems are the total cost of materials and at least one detail entry in the products section to support the establishment's industry classification. A lesser number of items may be stipulated for returns from establishments below a certain size.

Omissions. Checks for omissions involve finding items requiring an answer, but where none appears. The procedure may not be simple for items that are not applicable to all types of respondents. It may be necessary to instruct the clerks, for example, not to question the omission of a reply to an inquiry about the cost of purchased repair and maintenance services, because not all establishments purchase such services.

Internal consistency and data relationship checks. These consist of checks to determine whether the reported components of a total are equal to the reported total, whether the answers to the same query in different parts of the questionnaire are the same, and whether certain ratios are reasonable in the light of experience or other criteria. On a questionnaire calling for details on both the input of materials and the output of products, a "presence check" may be possible. As an example, an establishment manufacturing bread should use flour, and the clerk should be instructed to question the bread entry if no entry appears for flour. It may be noted that checks which specify the amount of raw material that should be used lead to complications and should ordinarily be avoided. Judgement as to the reasonableness of ratios, such as the ratio between the number of operatives and total wages paid to operatives, that is, their average annual earnings, may be based on a constant range of values. Ratios above or below these acceptable values would be questioned by the editing clerk and unless an obvious error could be identified and corrected, a query would be sent to the respondent. A list of 16 of these data relationships or ratios appears in paragraph 295 below. The acceptability of a ratio may also be decided after comparing it with the same ratio for other establishments of a similar size in the same industry, or in the same geographical area. In infrequent inquiries, such comparisons are usually not feasible because of the difficulty, at the time of editing, of finding a sufficient number of questionnaires with similar characteristics.

External checks. The reported data and the derived ratios may also be compared with earlier figures for the same industry. Data from other countries can also be used under appropriate circumstances. Such external data, when available, are normally used as guides in setting limits of acceptability.

In the course of editing, the effectiveness of the various checks employed should be examined from time to time. If the application of a given editing check results in the rejection of all questionnaires, for example, the tolerance is almost certainly too narrow. If on the other hand, no questionnaires are rejected, the check should be retested with a reduced range and, if the results remain the same, it should be discontinued.

D. Coding

241. Coding is the transformation of information to numerical form with the object of simplifying tabulation. The information to be coded may be numerical or non-numerical. The assignment of size codes based on the numbers of persons engaged transforms one set of figures to a simpler set for tabulation purposes. Geographical coding, on the other hand, transforms the names of provinces, counties and cities to a compact numerical form. Numerical codes are used also to denote industrial activity, type of ownership and other characteristics.

242. The importance of assigning industrial activity codes to establishments in a careful and consistent manner cannot be over-emphasized from the standpoint of securing internationally comparable results. For national planning purposes it is no less essential, since this phase of the work determines the final nature and structure of the country's industrial economy as it will appear in the published tables. The ability of respondents to provide sufficient information to enable products to be coded at the four-digit ISIC level will be a determining factor in the success of this operation. Equally important is the use of a classification system that is comprehensive and up to date.

243. Establishments with mixed activities - particularly those of significant size where receipts from products classified in two different industry groups are almost equally divided - present special problems. Efforts to effect a subdivision of the records to create two separate statistical units should be made wherever there is a valid basis for securing a distribution of the pay-rolls and gross receipts (all other item totals can be prorated and estimated from those two figures in final editing). The benefits accruing from such efforts are that the resulting statistics are cleaner for the two industries in both their "homogeneity" and "coverage" ratios. The "homogeneity" (or "specialization") ratio is defined as the proportion of product shipments (both primary and secondary) of the industry that is represented by primary products. The "coverage" ratio is the proportion of total shipments of the primary products of an industry (made by establishments classified inside and outside the industry) accounted for by establishments classified in the industry.

244. The assignment of the appropriate industry code to an establishment depends for the most part upon accurate reporting of product shipments by respondents and enumerators. Data entered in the "consumption of materials" section of the inquiry is by no means to be ignored here: quite the contrary, especially in textiles, metalworking, paper-making etc., where industry classifications of establishments can be linked to the nature of the material worked. The use of pre-printed pre-coded questionnaires with product descriptions appearing opposite the code numbers facilitates both reporting in the field and processing in the central office. An alternative that has been employed successfully is to provide reference lists to be used by enumerators and/or respondents in the selection of the appropriate product description and code number. Reference lists are effectively used with short forms as a means of gathering product information. When the advantages of reference lists are outweighed by their bulkiness and other considerations, the respondent (or interviewer) is instructed to write in a description of the products made, materials used and activities carried out by the establishment.

245. The need to enter codes on the returned questionnaires in order to facilitate the punching (or keying entry) and tabulation of the reported information has already been mentioned. Code numbers for some items, however, such as type of ownership, individual products shipped and materials purchased, can be pre-printed on the questionnaire.

246. A common example of the pre-printing of code numbers is found in connexion with the reporting of individual products and services. An inquiry on this subject that has not been pre-printed would provide several blank lines and ask the respondent or interviewer to write in product descriptions and enter the numerical data. A pre-printed and pre-coded inquiry, on the other hand, would provide the code numbers and descriptions of the several products made by the industry and in effect ask the respondent: (a) to select the appropriate line or lines; and (b) to enter the data.

247. The advantages of the pre-printing of code numbers both for materials and products are: (a) that no mass clerical coding operation is necessary; and (b) that punching is facilitated because printed numbers are easier to read than hand-written numbers. There is a drawback in that the respondent may enter data on the wrong line, and it may not be possible to detect the error. This is offset to some extent by the elimination of the possibility of code numbers being incorrectly entered by coding clerks.

E. Computer versus clerical coding

248. The use of the computer has greatly reduced the amount of coding required, especially the amount that needs to be done clerically. In the first place, codes whose primary purpose is to facilitate the sorting of punch cards are not required. The computer can operate about as well on a data item such as the number of employees as on a one-digit code based on the original figure. Moreover, if a one-digit size code is required for other reasons, the computer can readily

assign one. Secondly, clerks are notoriously weak when it comes to assigning codes based on a calculation and/or comparison of relative magnitudes, while the computer is always accurate in this respect if it has been properly programmed. In general, the computer is preferred whenever a code number is to be assigned based on numerical data, particularly if complex calculations are involved. Coding systems may be simplified also by computer processing. Although mechanical tabulation may require a six-digit geographical code in which each position is significant (for example, the first two digits represent the province, the second two the county, and the third two the city), computer operations may be equally efficient using a three-digit number assigned serially to cities while the reference to county and province is kept in computer storage. Many of the practices geared to hand or mechanical processes need to be re-examined if a computer is to be used.

249. The assignment of industrial activity code numbers may be difficult to program for the computer if industries are defined in any other way than by the simple predominance of a product or group of products. Industry codes based only on the respondent's description of activities cannot ordinarily be assigned by the computer. Location codes, likewise, normally require to be assigned by a clerical section, and there may be other items involving similar non-numerical information. Thus, some coding clerks will be required even though a computer is used.

250. With large-capacity computers, several industrialized countries have achieved success in automated area coding. The system enables the computer to assign each establishment on the census mailing list a numerical code based on location reference files in the computer and the address description on the questionnaire.

F. Establishment of work units

251. Whether the processing is computerized or not, assigning work, keeping records and making progress reports will be simplified if the questionnaires are sorted by type (manufacturing, mining etc.) and bundled together in work units of convenient size. Usually not more than 100 questionnaires should be assembled in one work unit. The work units should be numbered serially. A central record should be kept, showing the number of questionnaires in each work unit and a running record of its history, for example, the date it was entered into each phase of processing, the date of completion of each phase, and the number of questionnaires returned to the field for review, corresponded upon or referred to analysts. Work units are usually made up of the same type of questionnaires, and it may be desirable to sort them in other ways. For example, in clerical operations, the questionnaires for small establishments can be assembled in separate work units and edited by less skilled clerks. The instructions given to these clerks may, in effect, provide for the acceptance of errors which would be unacceptable for larger establishments. Conversely, the questionnaires for the largest establishments may be sorted and edited by persons with ability and experience above the average. An intensive edit of this kind would also be desirable for questionnaires obtained by sampling the small establishments. Limits may be set for the magnitude of entries which are to

be questionned: for example, the re-questioning of a respondent may not be permitted if the value of an item is below a certain limit. Instead, a procedure for substituting estimates for incorrect figures may be established.

252. It may be useful to have a separate work unit of cases for the imputation of data for non-respondents and for "failed editing" respondents when no questioning of respondents is planned. If the method of imputation depends on the use of average relationships or average values as determined from satisfactory questionnaires, figures should be imputed late in processing so that the computed average represents as many satisfactory returns as possible. As imputed figures differ in character from data obtained from respondents, the proportion estimated should be recorded in order to measure the effect on the validity of the results.

253. Certain types of checks may be performed only for questionnaires making a very important contribution to the published data. Checking of the data on output of products against related government series, for example, may be limited to establishments whose production of the specified products is above a specified figure.

G. Computerized processing

1. Data entry

254. Generally, the procedure for entering data into the computer is to key-punch the cards and then transfer them to tape. However, many countries are switching to a system by which the information moves directly on to magnetic tapes without an intermediate punched-card stage. Under this procedure, the information is usually keyed initially on to a disc, which functions as a temporary storage medium and avoids tying up a considerable number of computer tapes. Every two hours or so, all the discs (up to 32 or more) are "dumped" (transferred) on to a single magnetic tape. In the meantime, this intermediate disc stage can be used by the operator for instant retrieval of what has been keyed into the machine. This data entry system is capable of editing the data but the editing process can slow the system to a point where it becomes counter-productive. Therefore, it is suggested that a minimum of editing be performed on the equipment; for example, checks might be limited to verifying the presence of specified critical items (number of employees, value of shipments etc.). One such system uses equipment that resembles a computer terminal and is actually a small computer system. The operator uses a typewriter keyboard to enter the information, while a screen displays the format for the data specifying the key codes and relevant data items.

2. Use of "check digit"

255. The entry of reported data on to magnetic tape should be verified as part of an over-all quality control system. In computerized systems, a commonly used technique for the verification of the keying of pre-printed code numbers such as those for products and materials is the use of a "check digit". Mistakes in these numbers, or in the area codes or the establishment identification number, which are also verifiable by the "check digit" procedure, can be especially troublesome.

256. The check digit is a number calculated by the computer from the code number entered by the operator and it becomes part of the code number itself. With a six-digit code number, for example, the check digit becomes the seventh digit of the number. Check digits can be calculated in many ways. 16/

257. In the computer, the check digit is compared with the rest of the code number, which is flagged as incorrect unless the check digit passes this inspection. When the data-keying equipment described above is used instead of conventional punch cards, the comparison with the check digit can be made immediately at the keying station itself. Under that system, the check digit is calculated by the mini-computer at the keying station and when the computed check digit is not identical with the digit keyed by the operator, an error signal is displayed on the screen. The operator then can attempt to rekey the item correctly or flag the field in question by keying a special symbol in the check digit field so that the problem can be corrected at a later date.

3. Primary computer editing

258. After the data have been transferred to magnetic tape, they should be edited for housekeeping and screening purposes. In this process, each establishment record should be checked for basic processability (presence of key data, for example, employment and shipments) if this has not been done at an earlier stage (see paras. 240 and 254 above). This editing might also involve some data manipulation, such as the entry of a missing total as the sum of the reported detail.

4. Secondary (main) computer editing

259. Once the data have been screened to determine that the report is processable, each establishment return is subjected to main computer editing. This phase of the editing may consist of: (a) tests for scope; (b) industry code editing; (c) checks for omissions of data; (d) internal consistency checks; (e) magnitude editing; and (f) data relationship checks. Editing checks (a), (c), (d) and (f) are described in detail in paragraph 240 above. The industry code editing checks the accuracy of the four-digit industry code assigned to the establishment. Magnitude editing tests the establishment record for an appropriate scale of magnitude for selected key items such as wages and salaries and value of shipments. When editing failures are encountered in this test, the computer program triggers

---

16/ In one method, the odd-numbered digits of the six-digit code are summed and multiplied by 3; the sum of the even-numbered digits is added to the product and, finally, the result is subtracted from the next higher number ending in 0. For instance, with the code number 123456 the check digit would be computed as follows:

$$3\ (1 + 3 + 5) + (2 + 4 + 6) = 39;\ 40 - 39 = 1.$$

The complete number to be punched would be 1234561. If it were incorrectly punched as, say, 1324561, it would fail the test. Transposition of digits, as in this example, is a common type of punching error.

a fixed sequence of actions in an attempt to adjust the data, for example, by replacing a total by a sum of detail.

260. If there is an annual industrial survey in the country or an industrial census of sufficiently recent origin, the computer processing for the current census should include the matching of each establishment's current census record with its prior data and a detailed comparison of the two sets of figures (including the imputation of missing current year figures from prior data).

261. As a result of the computer editing operation, a computer establishment record either passes or fails. If the record passes all the checks and editing stages, it is ready for tabulation. If, however, it fails to pass one or more of the editing stages, it is flagged as a reject or as a referral case. As a rule, if the reported information is rejected, it is removed from the basic data tape and stored on another for correction purposes. There are, however, different degrees of rejection. Some computer records have so many failures that the questionnaire has to be recycled, that is, the corrections have to be made and the establishment record sent back through the entire computer editing process. There are also cases in which the computer record has only one failure but the reason for rejection is so critical that the questionnaire, after correction, has to be resubmitted through the entire computer process. In order to handle reject cases, a special reject listing is printed, with all the data for a rejected record plus flags indicating the cause for rejection.

262. Referrals are establishment records referred for review by professionals because the computer has changed a reported figure to a significant degree on the basis of instructions programmed in the computer. These large changes (including sizeable imputations for missing data) are identified and listed out as originally reported and as changed in the computer) for professional review and contact with the respondent, if necessary. Computer rejects and referrals are corrected by field visits, correspondence, telephone calls, or analysts' changes. A computer "diary tabulation" should be programmed to provide summary measures of the impact of different types of computer data adjustments at the publication cell level.

263. In editing for magnitude, consistency and data relationship, parameters of the upper and lower limits of acceptability of the data need to be set. The arithmetic mean of all establishments in the same industry and size group in a prior census or survey can be derived for each of the relationships studied, such as wages and salaries per employee. The upper and lower limits of the acceptable range of reported data for that ratio can be established by applying a standard deviation of two, three or four times the figure representing the mean; a standard deviation of three and one half is often selected. Cases falling outside the acceptable range are subjected to an evaluation system to determine which of the two numbers in the ratio is more likely to be correct. The program determines the "most suspect" component (item) of the failing ratio. It does this by computing an average reliability measure for each component. The component with the lowest average reliability measure is the "most suspect" component. The computer program then instructs the computer to take appropriate action on the data, changing numbers to bring ratio failures within tolerance limits and filling

in unreported data. The same program also provides for flagging of the different types of computer actions (see para. 251 above).

264. An alternative method of arriving at means and tolerance limits for the test ratios is to wait until eight or 10 acceptable records for the particular industry and/or size group have been entered into the computer file and then compute the mean and limits from that set of records. Something on the order of an eight-record moving average is often used but this approach has the drawback that the reports have to be perfect; every item has to pass or the report cannot be used.

H. Clerical processing

265. If a substantial number of questionnaires are to be clerically edited and coded, it is usually preferable to break down the operations into short routines. When this is done, clerks can handle all but a small fraction of the questionnaires after a short learning period. The extra handling involved is usually compensated for by the increased speed of processing and the fact that fewer highly skilled supervisors or industry specialists will be needed. Other procedural rules that have been found useful in practice are:

(a) All entries in the questionnaire should be made in a distinctively coloured pencil or ball pen and the colour chosen should contrast with both respondents' entries and field enumerators' entries;

(b) The colour used by coders should differ from the colour used by editors: some countries also differentiate corrections received from respondents by entering them in a third distinctive colour;

(c) When a correction is made to a respondent's entry or a field interviewer's entry, the original data should not be erased or obliterated; instead, the editing clerk should strike a line through the entry and enter the corrected figure above it; if space is limited, the corrected figure may be entered elsewhere and its proper position indicated by a guide-line;

(d) Space should be provided on the questionnaire for the initials of editing or coding clerks and the date of processing to facilitate later reference as to why changes were or were not made; if space is not provided, this control information should be entered in the margin of the form in a prescribed position.

266. It is the responsibility of the subject-matter staff to prepare specifications which will be used by the processing staff to prepare instructions for the clerks on what tests should be made to determine whether the report is acceptable or defective. Subject-matter personnel should also supply the reference material that will permit the imputation of missing information and the coding schemes that need to be applied to the data. Subject-matter personnel should arrange to check some of the early work units in order to make any necessary modifications in the procedures as soon as possible.

267. The clerical operations lend themselves to a quality control system designed to identify at an early date those clerks who are not following procedures correctly. The system should also be designed to measure production rates in conjunction with quality. Clerks should be notified immediately of their mistakes and those who continue to fall below acceptable standards should be dismissed. Frequently, this kind of check may also reveal an error in the procedures which could affect all editors or coders.

268. A correspondence unit should be established as part of the clerical screening and editing operations. It will be found that many of the answers that are needed to make a questionnaire complete or acceptable can be handled on the basis of a response to a form letter. The editing clerk can indicate the form letter to be sent (sometimes a single form letter can be designed to include a number of different items) and send a referral slip with this information to the correspondence unit, which can then dispatch the letter and set up controls to handle replies.

Chapter IX

TABULATION OF DATA

A. Introduction

269. Once the data have been converted to machine-readable form and the records have been machine-edited, the data are ready for tabulation. One of the first tasks of the census planners is to decide upon the method of tabulation and select the appropriate equipment. As was outlined in Chapter VIII above, the electronic computer can carry out the operations involved in editing, coding and tabulating data more efficiently and accurately than any other equipment.

270. Depending on the size of the computer's memory and the type of peripheral equipment used, the electronic computer can handle vast amounts of data, perform numerous sorts and calculations in response to complicated instructions and print out the results in the desired format. For countries with some experience in industrial censuses and undertaking a canvass covering a sizeable number of establishments with questionnaires of some complexity, a computer system of considerable size and capacity is required. For countries with less developed statistics a smaller, less sophisticated system may be quite adequate. The cost of a small computer system generally is roughly equivalent to a good electric accounting machine system. Many countries will find that the versatility and growth capability of a small computer system will outweigh the performance of the less sophisticated electric accounting machine equipment. Of course, in order to be able to rely upon computer processing, a country must have access to personnel with knowledge of current programming techniques.

271. This emphasis on computers is not intended, however, to suggest that electromechanical equipment, adding or bookkeeping machines or even manual tabulation should never be considered. It is necessary to weigh a variety of factors, including the amount of data to be processed, the editing and coding to be performed, the number and complexity of sorting and tabulating operations, the time constraints and the cost of alternative systems. While introducing computer operations some countries are maintaining other equipment (electromechanical, for example) and using that equipment as a back-up system, even running some parallel editing and tabulation as a verification of the results from the newly installed computers.

B. Choice of the computer system

272. Estimates of work-load, time and costs should enter into the selection of computer equipment. It is essential that the computer chosen should be able to process the census data within the desired time and cost limitations. Timing and cost requirements are especially important if the statistical agency does not own a computer and the facilities of another government agency or an outside organization must be used. In some countries, a central computer office has been established to provide computer services to the various government agencies.

273. The computer system should be analysed in terms of the computer programs to be used in processing the industrial census. The analysis should include the amount of memory, tape drives and discs needed. If the statistical agency has its own computer or has access to a computer, that system should be reviewed to ascertain whether it is adequate. If appropriate equipment does not exist in the Government, the possibility of renting computer time from a university or commercial enterprise should be investigated. Often these organizations are quite willing to provide such time at cost, rather than leave their computer idle. If a computer has to be purchased by the statistical agency, it may take two years or more to select the computer, prepare the site for its installation, install it and perform the necessary tests before accepting it from the manufacturer.

274. The experience of several countries indicates that the type of computer best suited to industrial censuses in countries at an early stage of statistical development is one of the relatively simple systems using magnetic tape. An advantage of taped records is the speed with which they can be processed. If several tabulations are to be made from the same basic records, the availability of taped records permits important gains in processing time.

C. Computer programming

275. In computerized tabulation, trained programmers must be found and they must take time to familiarize themselves with the objectives of the census, particularly with the plans and specifications for editing and tabulation. This may involve intensive and extensive systems design before programming is undertaken. Next, they should plan the broad outlines of the computer program and set down its essential features in the form of a flow chart. At this stage, the programmer and the planning staff should discuss the chart in detail, taking as much time as is necessary to make certain that it is completely understood by all concerned. The object of this review is to bring to light misconceptions and misinterpretations, which, if not detected, may be extremely costly to correct later on. The need for mutual education (that is to say, the programmer has to be aware of the statistician's problems and vice versa) and clear communication, which is the biggest problem in computerization, cannot be over-emphasized. Experience shows that the best results are achieved when the subject-matter specialist has some basic knowledge of systems design and programming. After approval of the general plan, the programmer will begin to write the program itself. The time required depends on the complexity of the program and the type of computer to be used. Simple programs may take only a few days to write; others may take several weeks. After it is written, the program should be tested. This is usually done with a "test deck" of punch cards of the type to be tabulated, which has been designed to bring out the features of the program and to represent the problems anticipated in dealing with the actual data. After testing and correction, the program is ready for use in editing or tabulating the reported data.

276. To design, develop and test computer data editing and tabulation programs for an industrial census, a staff of skilled programmers can be expected to work for several months, often well over a year. Many countries have neither the time nor the staff for such an undertaking. Therefore, early in the planning stage, the availability of generalized package programs for editing and tabulating the census should be explored. The United Nations Statistical Office can be helpful in this regard. Even if enough skilled programmers are available, their time can probably be better employed in developing special programs for which there are no generalized programs.

D. Electromechanical tabulators

277. This term is used to describe the older kind of equipment for tabulating from punch cards. In some respects similar to complex adding machines, these devices have provision for handling about 120 data positions (digits) simultaneously. The statistical results are printed out on a wide strip of paper.

278. Electromechanical tabulators vary in speed and capacity but even the most powerful cannot approach the capabilities of the electronic computer. The speed of the tabulator depends in part on the number of subtotals and totals required, because the cards stop passing through while these sums are being generated and printed. Some models are capable of calculating and printing out minor, intermediate, and major totals for the same data fields - for instance, data for cities and provinces, and for the nation as a whole. Some can produce new punch cards showing summary totals for selected categories; these summary cards may then be used in subsequent tabulations. The machines can count the cards as they pass through and if each card represents an establishment, this count may represent a statistic for publication.

279. With this type of equipment, data from the questionnaires are first transcribed on to punch cards by use of the card-punch machine, and the punching is verified by another operator using a card-verifying machine. Cards are then sorted and collated as needed for the particular tabulation or listing.

1. Design of punch cards

280. Punching and verification will normally be faster and more efficient if the cards are designed so that the figures are punched in the sequence in which they appear on the questionnaire. In electromechanical tabulation, it is desirable to group on one card the items that are to be tabulated together (for example, employment and earnings data). If, to facilitate punching, the items have not been punched in the proper sequence for tabulation, new cards can be created on a reproducing punch machine, but this will add to the costs. With computers, on the other hand, data fields can often be regrouped at little or no cost as they are transferred from cards to magnetic tape.

281. If there are two or more cards per questionnaire, the amount of identification and classification information required on each card is also affected by the method of tabulation. With electromechanical tabulation, it is customary to reproduce on each card the establishment identification number, the industry and geographical codes, and all the other codes to be used for the classification of the data, such as size codes. With computerized tabulation, the full information is needed only on the first card; on subsequent cards, the computer can associate classification codes and data without repeating the full set of codes on each card.

2. Card sorting

282. After punching and verification, the amount of mechanical sorting required depends on the tabulation method. With most types of electromechanical equipment, sorting by classification code number is required before each tabulation is made. With some of the smaller computers, particularly those which do not receive magnetic tape, a similar amount of advance sorting is required. With larger computers, some or all of the sorting required may be done by the computer itself, either while the

records are being transferred to magnetic tape or at a later point. Card-sorting requirements, for this reason, cannot be estimated until the method of tabulation has been determined.

E. Design of machine-output tapes

283. In both electromechanical and electronic systems, the statistical results are usually printed out on paper tape. The location of every digit and every blank space on this machine-output tape has to be planned in advance, and it should be laid out in such a way as to facilitate its subsequent use in the review of data, correction of errors, and posting of tables.

284. In recent years, progress has been made in designing machine-output tapes, which, as they emerge from the tabulating machines, are ready to be photographed and printed by the offset process. This method is successful if the data can be presented in the form of large numbers of similar tables. In most industrial censuses, however, there are many dissimilar tables, and hand-posting and typing is still the common practice. The form of the machine-output tapes is a less critical factor when the latter method of producing final tables is used.

285. The following are some practical rules to be considered in designing machine-output tapes:

(a) The review of the data should be facilitated wherever feasible by the calculation of ratios and other review aids; ratios such as average annual wages and salaries, and cost of materials as a percentage of the value of output, are particularly helpful in this respect;

(b) To avoid revealing confidential information, it is usual to examine publication cells to determine how many establishments or enterprises are represented; on many tabulation machines it is possible to print out these numbers alongside the cell for review purposes; some equipment can also print out the identification numbers and amounts reported by the largest enterprises represented in the publication cell; a less time-consuming and more reliable method of determining whether a disclosure situation exists is to program the computer to identify and flag disclosure cells; however, even here human examination of output tapes will often be necessary, (i) to select the non-disclosure cell to be withheld to protect confidentiality, which will minimize the loss of good information, or (ii) to deal with "disclosure by difference" situations (secondary disclosures);

(c) Space should be provided on the machine tape for the entry of corrected data; if corrections are to be punched on cards and carried through mechanically, consideration might be given to punching cards directly from the machine tape; the reviewers should use a distinctively coloured pencil or pen to make corrections;

(d) Strips of paper showing the column headings should be pasted on machine tapes so that reviewers can readily identify the data shown; in computerized operations, this may not be necessary if the computer can be programmed to print column headings directly on the tape;

(e) If the review specifications call for one figure to be compared with another, the two figures should be close together on the tape; if, owing to conflicting requirements, this cannot be done, strips of pasteboard with rectangular "windows" cut out to reveal the pertinent data may be used.

(f) In tabulations based on samples of establishments, the sampling weights are usually applied to the totals for each sample group at the tabulation or review stage; even where weights are applied in the tabulation process and aggregates ready for publication appear on the machine-output tapes, it is usually desirable to print out the unweighted data for review purposes; if tests indicate that many corrections to the basic data will be required as a result of the tabulation review, it may be best to postpone the weighting of sample data until all corrections have been made.

Chapter X

REVIEW OF TABULATED DATA

A. Introduction

286. The review of tabulated data in the form of statistical aggregates is the final step in census-processing prior to publication. Not until the data are available in aggregate form can they be scrutinized to judge whether they afford a valid picture of the industries and areas being measured. Only at this point can the data be compared against current surveys, past annual inquiries and censuses, or related series published by other countries with a similar economy.

287. The aim of a final review is not to detect insignificant errors or to double-check all the prior editing and coding operations. Instead, the object is to examine the product as it now stands from every angle (historical comparability, consistency with other current data, over-all "reasonableness") and decide whether it is acceptable for publication.

288. This stage represents the last opportunity in the statistical agency to validate the data gathered in the industrial census. Afterwards, the statistics are in print for the users to accept or question and to serve as the basis of judging the performance of the national statistical office. Accordingly, there should be no compromising of the resources (either personnel or funds) allocated to the task. The budget for an adequate review of the tabulated data should be established before the census is taken; thereafter, it must not be drawn upon to rescue other processing steps that run into trouble.

B. Staffing for the final review

289. Unlike the earlier stages of census-processing, final review of data is not a mass operation; it requires the time and judgement of a limited number of highly qualified personnel, principally professionals with training in economics and statistics. Most countries have learned from experience that in the final review operation there are no substitutes for expertise and good judgement. The reviewers need to be very familiar with all the processing operations up to this point, especially the editing performed on the census returns.

290. The census professional should not spend time on housekeeping corrections of the data (correcting routine or semi-routine errors, often of a systematic nature, made in the census-processing). These should be the responsibility of superior clerks or paraprofessionals on the review staff. The paraprofessional is a job classification that has emerged on the scene only in recent years. A paraprofessional is an individual qualified by experience to perform duties more complex than would be assigned to a clerk. Though not possessing a university degree, the paraprofessional has acquired sufficient knowledge of statistical practices and number sense to perform effectively on a level just below that of the junior professional. The paraprofessional works under the general supervision of a professional staff member.

C. Review in terms of data significance

291. The philosophy of the review of tabulated data should be one of a review in terms of data significance. The staff doing the review should avoid becoming engulfed in the arithmetic of the figures. The analyst should examine the cells of aggregated data carefully and ask the question, "What do the data tell me?". With a knowledge of the country's economy and the state of business conditions, the reviewer should consider whether the tabulated results constitute a true representation of industrial activity during the census year.

292. Final review is a time to be looking for sizeable errors remaining in the data. The period for review is too short and the cost of correcting data too great to justify searching for discrepancies other than those materially affecting data cells. Examples of serious errors are the omission or duplication of records for large establishments, reporting in the wrong unit of measure, sizeable entries made in the wrong column or on the wrong line etc. The improper industry coding of a single large manufacturing establishment will distort the industrial employment picture as much as several thousand individual errors in coding persons by industry in a population census.

293. The goals of final review should be to publish on a timely basis results of an acceptable level of accuracy and to avoid incurring excessive review costs by attempting to find and correct every error, including inconsequential ones. In addition to being a misguided concept, the complete elimination of statistical error is impossible in a project the size of an industrial census.

D. Objectives of the final review

294. There are three basic objectives of the final review: to assure internal consistency of the data; to evaluate the comparability of the data with the annual and more frequent surveys and with past censuses; and to detect and describe the limitations of the data.

1. Internal consistency checks

295. The first objective of the review of tabulated data is to ensure that the various items of data bear a reasonable relationship to each other and to corresponding figures for other industries, areas, size classes etc. The following exhibit lists the significant economic relationships that apply to a census of manufacturing activity; with minor modifications, these ratios are also applicable to the mineral, electricity, gas and water industries.

1983 INDUSTRIAL CENSUS DATA RELATIONSHIPS

The following ratios or relationships are to be examined in the final review of tabulations.

1. Number of employees as a percentage of the number of persons engaged
2. Number of operatives as a percentage of the number of employees
3. Wages and salaries per employee
4. Wages per operative
5. Wages per operative hour (or day) worked
6. Wages and salaries as a percentage of value added
7. Value added as a percentage of the value of shipments and receipts
8. Value added per person engaged
9. Value added per employee
10. Annual hours (or days) worked per operative
11. Fixed assets acquired as a percentage of the value of shipments and receipts
12. Fixed assets acquired per employee
13. End-of-year stocks as a percentage of the value of shipments and receipts
14. Value added minus wages and salaries as a percentage of the value of shipments and receipts
15. Value of shipments and receipts per person engaged
16. Value of shipments and receipts per employee

Note: The ratios specified above represent many of the most significant relationships from an "economic inventory" or "structural analysis" point of view. For example, "Wages and salaries as a percentage of value added" provides a measure of labour-intensity; "End-of-year stocks as a percentage of the value of shipments and receipts" is an indicator of the product turnover rate; "Value added minus wages and salaries as a percentage of the value of shipments and receipts" measures the approximate ratio of gross margin.

296. A note of caution is appropriate here. In the rush to meet publication time-tables, too many checks may lead to perfunctory checking. A shorter list of

essential checks would consist of ratios 1, 3, 6, 7, 8, 12, 13, 14 and 15. If a check fails, related ratios should be examined because they may give a clue to which of the figures in the failed ratio is correct. In so far as the minimum programme (Part One, paras. 46-61) is concerned, only checks 1, 3, 15 and 16 are relevant.

2. Comparability with current surveys and other censuses

297. A second objective of the review is to examine the comparability of the industrial census data with current measures of industrial activity for the reference year of the census. Examples of these may be monthly data on employment and earnings, business population statistics from the licensing and regulatory agencies and, in the context of the data on products made and materials used, statistics on exports and imports of commodities. The census product data may also be matched against statistics published by industrial trade associations and trade periodicals. Comparisons should also be made with data from the annual industrial inquiry for the previous year, past industrial censuses (if any) and, in the absence of such data for the particular country, with recent censuses for other countries with roughly the same industrial characteristics.

3. Limitations and qualifications of the data

298. The third function of the review is to develop the necessary knowledge to prepare adequate textual and tabular statements of the limitations and qualifications of the data. In some cases, comparability with past census data cannot be achieved because of changes in classification systems, geographical boundaries or coverage of the census. These come to the analyst's attention in the review operation and should be explained in table headnotes or foot-notes or in the text accompanying the tables.

299. The data shown in the statistical tables are subject to error arising from a variety of sources: response variability, non-response, estimation of missing data, and processing error. Some of these sources of error are more susceptible to measurement than others. For example, the extent of estimation or imputation of data can be quantified, whereas the calculation of processing error is a very difficult undertaking.

300. Where certain data are judged to be unreliable, that fact should be clearly stated in the census publication, together with an indication of the possible size of the error. For example, for figures with low response rates, the extent of coverage on which they are based should be specified; the standard error of estimate will serve to flag figures with large sampling errors. An alternative course of action is to suppress the unreliable data and give the reason for the suppression.

E. Statistical aids used in the final review

301. It is important to have in hand the appropriate analytical tools to perform the review of data at the post-tabulation stage. The most commonly used are described below.

1. Check-lists and worksheets

302. Statistical aids can be prepared in advance in the form of check-lists of items and ratios to be reviewed as well as tabulation review worksheets. These should show, for each line in the stub, ratios similar to those used in testing individual questionnaires. These ratios should be entered on the review worksheets in the form of summary comparisons with both historical data for the same series and comparable current data from external sources.

2. Establishment computer record

303. It is useful to have a computer print-out of the individual establishment records available for the analyst to check if he or she determines that the statistical aggregates are suspect. The computer listing should contain a complete master record of each establishment as it was entered into tabulation. There are two basic arrangments of these establishment records. The one most often used is organized by four-digit ISIC industry, by descending size class within industry, and by identification number within the size class. The other establishment listing is one with the location unit (province or county) as the primary sort, the four-digit ISIC as the secondary sort, employment size class the third and identification number the final sort.

3. Log of cases resolved earlier

304. The final review will be helped greatly if unusual situations relating to large establishments that are verified in the various processing steps are logged and a record of them given to the professional analysts who review the tabulations. These situations might involve the "birth" or "death" of an establishment, change in its industry code, or an exceptionally high or low wage rate, cost of materials or output per employee. It would be wasteful indeed to retrace steps that had already been taken to determine the accuracy of reported data.

F. Correction of tabulations and detailed records

305. Corrections detected in the final review should be made in the detailed computer record, by carrying them to a single computer listing. That listing should contain all the data for the establishment. The technique of using a single correction document helps to ensure the quality of the correction operation. It permits clerks to perform additive checks immediately before entering the corrections into the computer record. For example, if the total value of shipments for the establishment is lowered in the final review, the value of some individual product(s) shipped by the establishment will need to be reduced by an equivalent amount.

306. Verification of the correction operation is aided by the creation of a disposition listing. Wherever a change is made, the listing displays each line of information as it originally appeared in the record and as revised. When an analyst notices a variation between the revised tabulated totals and those he posted as a memorandum item during the course of his review, he can track down the problem by referring to the disposition listing and ascertaining which corrections were carried out improperly.

307. There is often a substantial demand for special tabulations after the completion of the regular census publication. It is important to carry establishment corrections to the detailed computer records to facilitate such additional tabulations. Such a procedure also guarantees that there will be a complete and accurate historical record for comparison with subsequent census data and for industry studies over time.

308. Sometimes pressures at the final review stage lead to the consideration of compromises in data correction procedures. For example, there is a tendency to make mass changes in estimation factors only at the publication cell level. Also, last-minute changes in industry and area totals are all too easily made without altering the establishment records, especially if the revision is extensive, for example, a change in the industry classification of an establishment which affects all the data of the establishment. Such short cuts should not be taken by the national statistical office; they seriously weaken the future use of the census file.

Chapter XI

PUBLICATION

A. Introduction

309. Publication is the final step in the statistical process by which the industrial census results are made available to data users. Within the limits of quality standards and of safeguards taken to prevent the disclosure of confidential information, a statistical agency should publish, at least at the national level, all of the basic data it collects. Some selection among the many conceivable cross-classifications of basic data with establishment characteristics must be made, however, if only because of budgetary limitations. The purpose of this chapter is to offer suggestions on the choice of tables for publication, on the timing of the publication of the results, and on the content of the text which should accompany the tables. In the discussion that follows, the tabulations considered are developed from those listed in the International Recommendations for Industrial Statistics. 17/

310. In chapter III above, it was pointed out that early planning of the content of statistical tables has many advantages. Errors in questionnaire design will be brought to light and corrected and the objectives of editing, coding and tabulation will be more clearly understood. If a set of table outlines is available to illustrate the discussion, the purposes of the census can be explained more simply and effectively. However, at the planning stage, it is not necessary or possible to determine the exact final form of the tables and the accompanying material such as text and foot-notes. Decisions regarding these matters may be postponed until later in the census-processing. The explanatory text, for example, should be based in part on the experience gained in conducting the census and thus cannot be written until processing is completed or at least well advanced.

B. Protection of confidential information

311. In most countries, respondents are much more willing to supply accurate information if they are convinced it will not be used to their disadvantage, for example, by their competitors or by the tax collector. Thus, the confidential treatment of individual reports, specified in the census law, greatly facilitates the collection of data and increases the quality of the results. Confidential treatment extends not only to the questionnaires in the hands of the statistical agency, but also to the published tables. The usual treatment is to suppress cells which, if published, would reveal the activities of a single statistical unit. The unit should be the enterprise, not the establishment, because the enterprise has access to the records of all the establishments it owns and controls. The publication of a cell representing two statistical units is prohibited, because the

17/ Statistical Papers, Series M, No. 48 (United Nations publication, Sales No. E.68.XVII.10).

activity of each unit (in other words, enterprise) might be revealed to the other. Thus, publishable cells must include, as a minimum, data for three statistical units. Some countries also follow a dominance rule for data considered especially sensitive. The rule involves suppressing cells where one or two units account for a very high percentage of the total.

312. In statistical tables with horizontal and/or vertical totals, a further complication occurs because if, to avoid disclosure, one of the components of a total is suppressed, at least one more cell must be suppressed or the "disclosure cell" can be derived by subtraction. Data can be disclosed by comparing one figure with another ("disclosure by difference" or "residual disclosure") in different tables in the same or a different publication and if that happens, appropriate suppressions must be made. In a given table, the application of the disclosure rules may result in so many suppressions that a statistical table is virtually useless. Such tables give a poor impression and should be omitted or combined with other tables unless there are compelling reasons for publishing them. For example, in a series of similar tables, each covering a particular area, two or more areas could be combined to avoid this disclosure problem.

313. Some statistics are considered to be more confidential than others. The amount or value of specific products made and expenditure for fixed assets are examples of data generally considered to be very sensitive. The number of establishments is generally regarded as a non-confidential matter, unless the classification is so detailed as to reveal confidential information. For example, to indicate that there is one manufacturing establishment in a given city could not be regarded as conveying confidential information, but to indicate it as a cotton textile mill, with an annual gross output defined within narrow limits, could well be so considered. Since the number of employees of an enterprise or an establishment is often mentioned in public by industrial executives, it is usually considered to be less confidential than other figures, particularly cost or output values.

C. Copy preparation and duplication

314. A key factor in determining the timing and cost of publication is the method of copy preparation and reproduction that is used. There are three basic methods available. The first consists in hand-posting tables, including stubs, column headings, titles, foot-notes, etc., after which the copy is sent to the printer for type-setting, printing and binding. While this approach will result in an impressive publication, it has the disadvantage of being expensive and time-consuming. The second basic method also calls for the hand-posting of tables, which then, however, are typed exactly as they will be published, with reproduction by the photo-offset method. While this method may not produce quite as good-looking a publication, the product has an attractive appearance and the procedure is less expensive and faster than type-setting. The other method involves the preparation of publication copy on the computer or, more specifically, on the high-speed printer (an auxiliary to the computer). While this is fast and potentially the least expensive of all three methods, it requires computer programming and should be undertaken only when such resources are present in the statistical agency. The computer method lends itself more to tables than to text and to tables that are large enough to warrant the programming effort. A number of countries with well developed statistical systems are using the high-speed printer approach extensively. In this method, the copy is reproduced by the photo-offset method in the printing process. Computer print-outs

of the tabulated data can also be used in place of hand-posting and typing, with significant error reduction, in conjunction with either type-setting or photo-offsetting.

315. Experience has demonstrated the advantages to be gained by adopting a procedure whereby each subject-matter division submits its publication specifications approximately a year before the scheduled issuance of the first preliminary report. A standard publication planning form should be used as the transmittal sheet, accompanied by a "dummy" layout of the publication. This procedure enables the editorial and related work required for publication planning to begin at the earliest possible time.

316. No medium for copy preparation is infallible; all methods require careful review to make sure that errors do not occur during the hand-posting, typing, type-setting or computer operations.

D. Guidelines for tabular presentation

317. Good style in tabular presentation calls for a clear and economical display of the data. Generally acceptable practices should be followed in the layout and design of the tables; the way table titles are written; the use of headnotes and foot-notes; the use of vertical and horizontal rules; and consistency in capitalization, abbreviation, punctuation and indentation.

318. The tables should be planned from the viewpoint of the user. Everything possible should be done to make the table easy to read, understand, and interpret. The table form does not add new meaning to the statistics; it is designed only to make the meaning that is already there easier to find.

E. Description of publication programme

1. Preliminary and final publications

319. It is recommended that the results of the industrial census should be issued in a series of preliminary (advance) reports, to be in print within two years of the end of the reference year, followed by a final report, which would appear during the following year. It is further suggested that efforts should be made to prepare and print a bound volume consisting of all the final reports. There is a minimal amount of work involved, since essentially the operation is merely an assembly of the printer's copy of all the final reports in ISIC and area order, eliminating duplicated text, appendices and other pages. If funds do not permit a regular printing of the bound volume, consideration should be given to the manual assembly of a few sets for official use, to be provided the Heads of Government, and for depository libraries in the particular country, other countries, regional organizations and the United Nations. In addition to its prestige feature, such a volume assures the historical preservation of the census results.

2. Separate series by major divisions

320. It is suggested that three preliminary reports should be issued, one for each of the three major divisions covered by the census: manufacturing; mining; electricity, gas and water. These reports would contain selected key statistics

for industries and geographical areas. The final data for each of these major divisions would be released in a separate series of reports. There would be a pattern to the tables shown for each major division. For example, the manufacturing series would have one set of tables presenting industry data; another set showing geographical information for provinces, countries etc.; and a third, smaller, group of tables providing statistics by other characteristics such as size of establishment and type of ownership. 18/

321. Reports that group tables according to these three categories would also be issued for minerals and for the electricity, gas and water industries. Uniformity in the tabular content of the final reports in each major division benefits the data user. The user will frequently want to refer to all the industry data for in-depth studies of an industry or group of related industries or to all the geographical data for a regional analysis of the economy. Furthermore, this feature of parallelism in the tables aids the statistical agency in the preparation of the census publications.

322. In addition to the separate series for each major division, there should be a summary report bringing together key totals (for example, employment and value added) from all three major divisions.

323. If industrial activity in household units is covered in the industrial census, a publication should be issued with two tables: one showing the household industrial activity according to those three-digit or four-digit ISIC industries in which households are most active; the other presenting the basic household data for those provinces and large counties where such activity is prevalent.

F. Content of tables

324. The published tables should consist of basic data items classified by establishment characteristics. The basic data items are: number of establishments; number of persons engaged; number and wages and salaries of employees; cost of materials; gross output; value added; changes in stocks; and gross additions to fixed assets. Establishment characteristics can be grouped into three categories as follows: branch of industry (ISIC); location; other characteristics, such as establishment size based on employment (or output) and type of ownership.

325. The number of tabulations that could be prepared from an industrial census is so large that in most countries the available funds could not provide for all of them. It is necessary, therefore, to select those that will be of the most general interest. Shown below is a set of tables, together with table titles and brief descriptions of the tables, that might be published in a typical industrial census.

326. The tables suggested are based on those described in the tabulation recommendations for the 1983 World Programme. 19/ The list of tables below is not

---

18/ In this group of tables, establishments might also be shown by size classes of ratios between basic items of data, such as wages and salaries to value of output.

19/ Recommendations for tabulations to be prepared for the minimum and complete programmes as part of the 1983 World Programme of Industrial Statistics, are shown in Part One of the present publication.

necessarily in order of importance, nor is it intended to be comprehensive. For example, except for tables 1 (b) and 2, no provision has been made for the presentation of data according to geographical areas. Such statistics are significant for many classes of data users and countries are urged to make adequate provision for such data. Suggestions for the tabulation by area of the basic data items will be found throughout the discussion of the tabulation plans for the 1983 Programme in Part One of the present publication.

Table 1 (a). Principal indicators of industrial activity, classified by industry

This table would present for the nation as a whole the basic data items (defined above) for the three major divisions: (a) mining; (b) manufacturing; and (c) electricity, gas and water, classified by kind of activity at the four-digit level of ISIC or the national equivalent.

Table 1 (b). Principal indicators of industrial activity, classified by geographical area and industry

This table would present the same basic data items as in table 1 (a), by two-digit and three-digit ISIC detail within geographical areas.

Table 1 (c). Principal indicators of industrial activity, classified by industry and establishment size class

This table would present for the nation as a whole the same basic data items as in table 1 (a), cross-classified by industry at the two-digit level of ISIC or the national equivalent and size classes of establishments.

Table 1 (d). Principal indicators of industrial activity, classified by industry and type of establishment

This table would present for the nation as a whole the same basic data items as in table 1 (a), cross-classified by industry at the two-digit level of ISIC or the national equivalent and form of ownership, legal organization and economic organization (single or multi-unit). If the number of classes for the latter characteristics is large, it would be preferable in practice to split this table into two or three, one for each characteristic.

Table 2. Establishments by employment size class, classified by industry and geographical area

This table would present the number of establishments in each employment size class, by geographical areas with four-digit ISIC detail.

Table 3. Employment and earnings, classified by industry

This table would present data items on employment and earnings, such as number of working proprietors, active partners and unpaid family workers, number of employees, number of operatives, hours (or days) worked by operatives, wages and salaries paid to operatives and other employees, and supplements to wages and salaries. The data would be classified at the four-digit level of ISIC or the national equivalent.

Table 4. Value of stocks at the beginning and end of the year and change during the year, classified by industry

This table would present the value, at the beginning and end of the year as well as the change during the year, of each of the four main categories of stocks, namely, materials, fuels and supplies, work in progress, finished goods and goods for resale. The classification of the data would be at the four-digit level of ISIC or the national equivalent.

Table 5. Fixed capital formation and its composition, classified by industry

This table would present data on gross fixed capital formation and the value of acquisitions of new fixed assets according to type of asset. The classification of the data would be at the four-digit level of ISIC or the national equivalent.

Table 6. Cost of materials and industrial services used, classified by industry

This table would present data on the cost of total input and its components, such as materials and supplies, fuels, electricity purchased, contract and commission work, repair and maintenance, and the cost of goods shipped as received. Data would be classified at the four-digit level of ISIC or the national equivalent.

Table 7. Cost of fuels purchased and quantity and cost of principal fuels, classified by industry

This table would present data on the total cost of fuels purchased and the quantity and cost of individually important fuels. Data would be classified at the four-digit level of ISIC or the national equivalent.

Table 8. Energy equivalent in terajoules of fuels and electricity consumed, classified by industry

This table would present data on the total energy consumed in terms of terajoules, with a breakdown between the energy equivalent of fuels consumed and electricity consumed (electricity purchased and generated less electricity sold). Data would be classified at the four-digit level of ISIC or the national equivalent.

Table 9. Purchases of selected individual materials

This table would present for the nation as a whole the data collected on individual materials purchased, irrespective of the industry in which the consuming establishments are classified. The suggested column headings are material code number, name of material, unit of measurement, quantity and delivered cost.

Table 10. Cost of non-industrial services, classified by industry

This table would present data on the cost of selected non-industrial services rendered by others, classified by industry. In view of the limitations that these data might have (see Part One, para. 136), the table format and level of classification should only be decided after the initial scrutiny of the information gathered.

Table 11. Value of output in producers' prices, classified by industry

This table would present data on total output and its components, such as the value of all products, receipts for industrial work or services, and the value of goods shipped in the same condition as received. The classification of the data would be at the four-digit level of ISIC or the national equivalent.

Table 12. Output of selected individual products

This table would present for the nation as a whole the data collected on individual products, irrespective of the industry in which the producing establishments are classified. The suggested column headings are product code, name of product, unit of measurement, quantity produced and quantity and value of shipments.

Table 13. Receipts for non-industrial services, classified by industry

This table would present data on the receipts for selected non-industrial services rendered to others, classified by industry. In view of the limitations that these data might have (see Part One, para. 159), the table format and level of classification should only be decided after the initial scrutiny of the information gathered.

Table 14. Value added and its composition, classified by industry

This table would present data on value added as a census concept and not as defined in the national accounts (for the latter, see Part One, para. 170, table 15). The suggested items are value added in producers' prices, value of output in producers' prices and cost of materials and industrial services used. The classification of the data would be at the four-digit level of ISIC or the national equivalent.

327. In the presentation of data for the bench-mark industrial census, the major tables should contain some information (for example, total employment and value added) for the previous census or annual inquiry. Such data would be appropriately shown in the summary tables presenting statistics by kind of activity and geographical area. It might also be helpful in such tables to shown percentage changes from the previous census to the current one.

G. Preliminary publication considerations

328. To be of greatest value, particularly in relation to current public or private planning, the results of the census should be published as soon as possible after the end of the census year. Not later than 24 months after the close of the reference year is suggested for the preliminary results; 36 months for the final ones. On the other hand, there are many features of industrial census statistics which are still useful after years have elapsed. The census staff may be faced with urgent demands for early publication of results, but must weigh these against the need for thorough editing of the questionnaires and careful coding and tabulation. Efforts to complete these phases too rapidly may lead to earlier publication but, if the quality of the results is poor, may be self-defeating. If a number of significant errors that require substantial revisions are subsequently discovered, the statistical agency may find itself subjected to a great deal of criticism. It is therefore extremely important that the summary figures selected for advance release should, if at all possible, represent control totals which have been confirmed and will be repeated in the final publications. Where this is not the case but revisions are likely to be held to a tolerance of 5 per cent for industry major divisions at the national level (10 per cent for larger-area or industry-group totals) for the basic data items, then the price being paid for early publication is probably not too high.

329. One commonly used device for satisfying at least some of the demands for early publication is to select certain of the most important basic data items or tables for preliminary release. Tables 1 (a) and 1 (b) above, perhaps condensed to show less than the full detail suggested, could be considered for this purpose. Other schemes are to release individual preliminary reports on particular industries or areas on a flow basis. Such reports would contain summary data selected from the list of final tables; each report might have two or three such summary or abbreviated tables.

330. There are advantages to preliminary publication besides the obvious one of demonstrating the willingness of the national statistical office to make a special effort to satisfy some of the more urgent demands for early release of census data. In the exercise, valuable experience is gained, which can be turned to good account in one or more of the following respects: firming up table formats; settling upon workable disclosure rules; developing a time-table for the release of final publications that can be adhered to; and securing a more realistic projection of funds and staff resources needed to get the census results into print.

H. Contents of the preliminary and final publications

331. The preliminary (advance) reports should be quite functional in order to conserve time and money. Therefore, the reports should consist almost entirely of statistical tables. There should also be a brief text or technical statement that deals with the scope and coverage of the census, the basic items appearing in the tables and methods of establishment classification. Explicitly stated and prominently displayed should be the caution that "the figures are subject to revision in the final reports".

332. The final reports should consist of the following items appearing in the order indicated:

Cover (The reverse side of the cover might have an outline map of the particular country.)

Title page

Back of title page containing acknowledgements

Table of contents

Text

Tabular pages, consisting of statistical tables

Appendices (these might comprise a copy of census legislation as well as copies of questionnaires)

333. To establish the background for the statistical tables, the text might begin with a review of the highlights of the survey findings. This summary could include a few charts to make the statistics more vivid and to point up important aspects of the survey results.

334. The following topics are suggested for inclusion in the text of the final reports and bound volumes:

(a) Reasons for conducting the census and its legal authority; uses of the results; economic significance of the industrial sector;

(b) Brief history of the country's previous industrial censuses, if any, and of its other industrial surveys; if no historical data can be shown in the statistical tables, the reasons might be discussed;

(c) Scope of the census in terms of the types of respondents and industrial activities covered; if any small groups or relevant activities are omitted, the probable effect of their omission on the results both in detail and over-all (objective quality standards require that if more than 5 per cent of employment in an important four-digit industry is underrepresented, a measure of bias should be provided for the key items);

(d) Statistical unit and methods of enumeration; methods of locating and identifying respondents; special procedures, if any, for enumerating multi-unit enterprises; defects of enumeration and steps taken to correct them;

(e) Classification systems employed;

(f) Definitions of the basic data items; valuation methods used for reporting data;

(g) Procedures and definitions for derived data items;

(h) Follow-up methods; treatment of statistical units which failed to report for one reason or another; if estimates are included in the published results, the proportion estimated and the general effect on the validity of the results (objective quality standards require that no more than 10 per cent of employment in four-digit industries should be estimated without call-back for key items);

(i) Editing, coding and tabulation procedures, with emphasis on the effect of the techniques on the published results;

(j) Type of sampling design and reliability estimates, if sampling techniques are used.

Annexes

Annex I

PRACTICAL SAMPLING TECHNIQUES IN INDUSTRIAL CENSUSES

1. Often it is desirable to plan to collect reports from only some rather than all units that are defined as being within the scope of the industrial universe. This method, called sampling, has obvious advantages. It relieves many respondents of the burden of reporting and simultaneously reduces the work-load of the census agency, and, therefore, its costs. Because the lower work-load permits a greater degree of control to be exercised over errors of collection, of reporting, of processing etc., sampling may also improve the quality of the statistics compiled, especially those for small establishments. Additionally, the lower work-load may advance publication of the census results. Another less obvious advantage of sampling is that it provides a flexible mechanism for balancing expenditure against data needs.

2. In some circumstances, sampling may be the only feasible method for obtaining the desired data. This is the case when the number of units is large and either a comprehensive directory is not available or collection costs are high. For example, practical difficulties will ordinarily make it impossible to enumerate completely the manufacturing activity carried on in households. If such activity is sufficiently important to be included within the scope of the census, it will usually be necessary to resort to sampling.

3. Loosely, any partial coverage procedure may be termed sampling. For example, if the reports that were easy to collect were the only ones to be included, they might be called a sample. Such samples, however, would usually have little value, because it would be difficult to draw accurate conclusions about the universe from them, and as a rule it would be impossible to assess the errors of any estimates that might be derived from them. Probability sampling overcomes this difficulty, for then the mathematical laws of probability apply and the relationship between the sample and universe values can de determined, at least approximately.

4. Probability samples can be selected in many ways. All satisfy the condition that the sample elements are chosen by chance, in accordance with positive, ascertainable probabilities assigned to each element in the universe. When these conditions are carefully observed, and the probabilities are properly used in deriving estimates, the sample will yield results of measurable quality and the substantial administrative benefits indicated above.

5. A note of caution is nevertheless in order. Probability sampling methods introduce additional steps that must be carefully executed. These involve some cost, and create opportunities for mistakes that do not occur with complete coverage. Unless the sample is designed to fit the prevailing operating conditions and is satisfactorily controlled, losses rather than gains may result from introducing it. A highly elaborate and theoretically desirable sample design that required more skilled personnel, more records, and more equipment than were available would be worthless or worse.

6. Also, it must be recognized that the chance character of probability sampling implies that many different samples can be selected under the identical sampling plan, and that the estimates will vary from sample to sample. Some of the estimates will be close to the universe total, while some will not. The differences that can occur are the price paid for the cost reductions and other benefits of sampling.

7. How much a sample estimate may deviate from its corresponding universe figure is measured by the standard error of the estimate or briefly, the standard error. It is defined as the square root of the sampling variance, the average of the squares of the deviations of the possible sample values from the corresponding universe value. The relative standard error, the ratio of the standard error to the value being estimated, is often used to describe the quality of the sample estimates. It is especially convenient for comparing industrial statistics that are measured in different units: kilograms, dollars, persons, days etc.

8. Theoretically, the sampling variance can be reduced to any desired level by increasing the sample size sufficiently, but to specify excessive reliability may require enlarging the sample to the point where it approaches a complete census. The effort can be self-defeating, for the quality improvements achieved by expanding the coverage could be illusory. Unless the reporting, processing, and other systematic errors are well controlled, their effect can be nearly as great as the combined effect of the sampling and systematic errors with less coverage. Efforts to control the systematic errors, when too broadly dispersed, can divert attention and resources from more important needs. Conversely, if reducing coverage yields improved control, sampling might yield substantially better quality than a complete census for the principal statistics. A substantial reduction in the systematic errors could more than offset the errors due to sampling, which will often be quite moderate for major classifications. Moreover, when a reasonably comprehensive directory has been compiled, it will provide much of the actually required detail on the distribution of industrial activity by industry and geography. The real needs for the full range of industrial establishment statistics can frequently be satisfied by estimates at only broad industrial and geographical levels.

9. Many factors influence the size of the sampling variance. Among the most important are the variance of the sampling units in the universe, the way in which the sampling units are selected, and the number (and proportion) of units included in the sample. To some extent, these factors can be controlled.

10. The variance in the universe (called the unit variance), for example, will depend upon the definition of the sampling units. They might be entire towns, blocks within towns, or individual establishments, depending upon the conditions for conducting the census. If entire towns are the sampling unit, the total number of persons engaged, total value of production, total cost of materials etc. in each town, and the number of towns will determine the unit variances; if blocks are the sampling units, the totals for each block and the number of blocks will determine the unit variances; if establishments are the sampling units,

their individual values and the number of establishments will determine the unit variances.

11. When the units are selected, they may be considered a single group for sampling purposes, or they may be divided into a number of strata, each of which is sampled independently. Stratification guarantees that every class of units defined by the various strata will be represented in the sample in predetermined proportions. This is not the case when the sample is drawn unrestrictedly from the entire universe. With unrestricted sampling, representation of the various classes is uncontrolled, and more different sample combinations are possible.

12. A small simple example is illustrative. Suppose that a sample of two establishments is to be selected from a list of four establishments; that, on the basis of earlier reports, establishments 1(a) and 1(b) are classified as small and establishments 2(a) and 2(b) are classified as large; and that their actual current values are:

Establishment	Number of persons engaged
1(a)	2
1(b)	13
2(a)	79
2(b)	96
Total	190

With a simple random sampling from the entire universe, all possible pairs of establishments would be equally likely to be selected, so that each of the extreme samples - (1(a), 1(b)) and (2(a), 2(b)) - would be as likely to be selected as any other. The corresponding estimates of the total number of persons engaged would be (2 + 13) x 2 = 30, and (79 + 96) x 2 = 350. If, instead, establishments 1(a) and 1(b) were grouped into one stratum, establishments 2(a) and 2(b) were grouped into a second stratum, and a sample of one establishment was randomly and independently selected from each stratum, the only obtainable results would be:

Establishments	Estimated number of persons engaged
(1(a), 2(a))	162
(1(a), 2(b))	196
(1(b), 2(a))	184
(1(b), 2(b))	218

These four reasonably satisfactory estimates might also be obtained with unrestricted random sampling, but the crucial point is that one third of the time unrestricted random sampling would yield one or the other of the two extreme and poor results. No detailed calculations are needed to see that, by eliminating any possibility of getting both sample units from the same class, stratification would greatly reduce the sampling variance.

13. That the total sample size can be controlled is obvious. Without violating the fundamental principles of probability sampling, any particular unit can be assigned any probability of selection greater than 0, up to 1 (certainty), as desired. In particular, different probabilities or sampling rates can be assigned to different strata. The only restraint that must be observed is that the total sample number over all strata should equal the total budgeted sample size.

14. The assignment of different sampling rates is a simple, and a powerful way in which to compensate for differences in the variability of different strata. If two strata each contain the same number of units, but have unequal variances, more of the total sample should be allocated to the stratum with the greater variance. Likewise, the total number of units in each stratum is a factor in determining the sample allocation. The more units a given stratum contains, the more weight it will have in the sampling variance. Again considering two strata, if their variances are equal, the larger stratum should be allocated more sample units than the smaller one.

15. The best or optimum allocation is obtained when the sample number assigned to each stratum, $\underline{n}_i$, is proportional to the total number of units in the stratum, $\underline{N}_i$, and to the standard deviation of each stratum, σ_i (the square root of the stratum variance). The mathematical expression for the optimum sample number is

$$\underline{n}_i = \left(\frac{\underline{N}_i \sigma_i}{\sum_{i=1}^{M} N_i \sigma_i}\right)\underline{n}, \qquad (1)$$

where

$\sum_{i=1}^{M}$ indicates the sum over all strata,

and

$\underline{n}$ the total size equals the sum of the $\underline{n}_i$.

To find the corresponding optimum sample rate, $\underline{r}_i$, both sides of equation (1) are divided by $\underline{N}_i$, giving the result

$$\underline{r}_i = \frac{\underline{n}_i}{\underline{N}_i} = \frac{\underline{n}}{\sum_{i=1}^{M} N_i \sigma_i}. \qquad (2)$$

The total sample size n is presumed to have been determined in advance by the available budget for the census and estimates of the unit costs of collecting and processing the reports. Presumably, also, the total number of units in each stratum N_i is known, at least approximately. When additionally, on the basis of past experience, reasonably good estimates of the σ_i are available, sampling rates that are approximately optimum for each stratum can be derived.

16. When specific estimates of the σ_i are not available, it becomes necessary to use somewhat cruder approximations in determining appropriate sampling rates. Fortunately the relative standard errors are not highly sensitive to moderate departures from the exact optimum sampling rates, and in sampling for industrial statistics certain simplifying assumptions have been found to give satisfactory results.

17. One of the most important assumptions is that the optimum sampling rates are very nearly the same for all the principal general measures of industrial activity, such as the number of persons engaged, the total value of production, the total cost of materials etc., because these items generally are very highly correlated. (This, however, will not so often be true for estimating individual products, materials, or other unique items.) For estimating general statistics, any convenient one can be used as a basis for assigning sampling rates to strata. The number of persons engaged, which is the size statistic most commonly available in directories, can serve the purpose, but if some other measure is available and preferred it can be used just as well.

18. A second useful simplifying assumption that can be made is that the standard deviation for strata, defined in terms of size classes, is approximately proportional to the mean or midpoint of each class. Stated another way, this assumption says that for all size strata the coefficient of variation (the standard deviation divided by the mean) is the same. Assuming that the coefficient of variation is constant not only over all size classes within given industrial and geographical classifications but is also constant for all such classifications considerably simplifies the problem of determining how to allocate the sample.

19. The assumption leads directly to the following helpful rule for deciding the size above which all establishments should be included with certainty:

> Divide the total number of persons engaged by the budgeted total number of establishments to be included in the census. The result is the size cut-off at and above which all establishments should be included with certainty.

The certainty cut-off value calculated in this way will often be an awkward number, such as 13.7 persons engaged. No serious harm will be done by rounding it to a near, but more standard, size limit. Either 10 or more or 15 or more would be satisfactory for the example given.

20. It is not imperative that every establishment larger than the certainty cut-off be included with certainty. Some misclassifications in both directions are likely to occur and can be tolerated. Misclassification of smaller establishments into the certainty class presents no problem. However, misclassifying very large establishments as being small is potentially damaging. There is no universal rule for defining "very large" that fits all situations, but beyond twice the calculated certainty size, failure to include the establishments with certainty rapidly increases in importance. Every reasonable effort should be made to identify all establishments of about that size or larger before sampling, and to designate them for certainty coverage.

21. The need for complete coverage of all very large establishments is stressed, because each establishment is unique and substantially different from all the others. Usually there are relatively few very large establishments, and they are widely scattered among many industries and geographical areas and thinly distributed over an extremely broad size range. Because of these characteristics, the variances of any group to which they might be assigned would be extremely large. By including them all with certainty the difficulty is avoided, for then they do not contribute any sampling variance.

22. Sometimes stratifying by size below the certainty value is also worthwhile. If the certainty cut-off is as low as five persons engaged, little can be gained by further size stratification. If the certainty cut-off is as high as 50 persons or more, up to a maximum of three size strata might be useful, and for intermediate values of the certainty cut-off two size strata could be helpful.

23. In defining additional size strata, standard size class limits can be used. If a decision must be made on how to combine size classes into strata, adjacent classes should be merged so as to equalize as nearly as possible the aggregate number of persons engaged in each stratum.

24. The sampling rates for each non-certainty stratum, in accordance with the assumption of a constant co-efficient of variation, should be approximately proportional to the mean or midpoint of each such stratum. It can be seen that this rule is consistent with equation (2) for the optimum sampling rates, because the standard deviation for each stratum, σ_i equals the product of the co-efficient of variation and the stratum mean $\underline{X}_i$. Upon substituting for σ_i in equation (2) then, it is found that $\underline{r}_i$ equals a constant times the co-efficient of variation, which is another constant, times $\underline{X}_i$. The product of two constants is also a constant, say $\underline{K}$, so that the result can be expressed as

$$\underline{r}_i = \underline{K}\underline{X}_i. \qquad (3)$$

The proportionality factor $\underline{K}$ can be computed by dividing the total sample number of non-certainty establishments (the budgeted number less the number of certainty establishments) by the total number of persons engaged in all the non-certainty strata combined. The approximately optimum sampling rates for each stratum then can be computed by equation (3).

25. As in the case of determining the certainty cut-off value, the results r_j are likely to be awkward. Here again, not much will be lost and something will be gained in simplicity if the sampling rates are rounded, at least to the nearest integral fraction, such as 1/3, 1/7, 1/26 etc. In fact it will not hurt if rates of less than 1/5 are rounded to the nearest "5", for example, 1/7 to 1/5 and 1/26 to 1/25.

26. Stratifying by size has been emphasized because variation by size dominates the sampling variances. Since, however, estimates will usually be derived in some degree of industrial and geographical detail, as far as possible, these characteristics should also be used as controls. Within size classes, proportional stratified sampling by industrial and geographical classifications will be satisfactory.

27. Sampling systematically from an ordered file is an easy way to select such a proportional stratified sample. For this purpose, the file is grouped by size strata, within each stratum by industry, and within industry by geographical area. Then a sample is simply selected systematically from each size stratum at its specified rate, beginning with a random start. For example, if the rate is 1/5, a random number between 1 and 5 inclusive should be selected. The corresponding establishment and every fifth establishment thereafter would then be designated as the sample from that stratum. If the rate for a second size stratum is 1/20, a random number between 1 and 20 inclusive would be selected as the start, and then every twentieth subsequent establishment would also be included in the sample. This procedure provides reasonable assurance that the sample from each size stratum will be fairly evenly distributed by industries and geographical areas. The unduly heavy concentrations in some industries or areas and light representation of others that might occur with simple random sampling will be avoided.

28. The sampling method described above also permits additional flexibility in designing the sample. Particular industries or particular geographical areas may be considered exceptionally important, and better than standard reliability may be desired for them. In such cases their relatively greater than average importance can be introduced as a weighting factor that is applied to the actual measure of size of each establishment in the given industry or geographical area. For example, if for some special reason industry Y is deemed to be twice as important as any other industry of the same size, the size of every establishment classified in industry Y should be doubled as the very first step in developing the sample. Then industry Y will automatically be given the extra emphasis desired for it.

29. The above procedure for designing the sample requires that the total number of establishments to be included in the census should be decided in advance. The method described does not explicitly involve variances, or indicate how large the relative standard errors might be for various estimates. It is nevertheless approximately optimum in the sense that it roughly minimizes those errors for general statistics estimates. Theoretically it is possible to find the reverse solution, to specify acceptable relative standard errors and then to determine the allocation that reduces to a minimum the total number of establishments. The results would be identical. That is, a given number of establishments leads uniquely to one allocation and one set of relative standard errors; that same set

of errors would lead to the identical allocation and the same total number of establishments. Of the two approaches, beginning with a budgeted number of establishments is more realistic, for it is much easier to reach agreement on what a reasonable budget is than to decide what are reasonable relative sampling errors. In addition, the calculations are very much simpler, since the derivation of the results involves only one pre-specified figure rather than the many that would be involved if individual errors were pre-specified.

30. The question of the reliability of the estimates should not, however, be disregarded. An initially proposed budget might be grossly inappropriate for the purposes of the census. It could be seriously inadequate, so that the estimates would be of such poor quality as to be practically useless. Conversely, a budget chosen in a purely arbitrary manner could result in excessively large samples, and relative sampling errors that were clearly disproportionate to needs.

31. Unless earlier experience can be drawn on, it will usually be possible to derive only crude estimates of the relative sampling errors that might be anticipated. Nevertheless, an effort should be made to determine at least their order of magnitude. The following approximate formula is useful for the purpose:

$$\hat{V}(X'_h) \doteq CV\sqrt{R/n}\sqrt{R_h / P_h} \quad , \qquad (4)$$

where

$V(X'_h)$ is a rough prediction of the relative standard error of the estimated total for category h;

CV is the coefficient of variation, which is assumed to be the same for all size classes, and for every industrial and geographical category, h;

n is the total non-certainty sample size;

R is the proportion of the all-industry grand total accounted for by all the non-certainty establishments in the universe;

R_h is the proportion of the total for category h accounted for by all the non-certainty establishments in category h;

P_h is the proportion of the all-industry grand total accounted for by all establishments in category h.

This formula can be used to determine, roughly, the reliability that can be expected for any particular estimated total X'_h, for all the parameters involved are either known or can be estimated fairly well. The values of n and R, the non-certainty sample size and the proportion of the total value of the universe, are both available as by-products of the certainty cut-off calculation. A satisfactory value for CV can be obtained by computing the mean and standard deviation over the entire size distribution below the certainty cut-off, without regard to any additional size

stratification that may be planned. The values of $\underline{R}_h$ and P_h can be derived by tabulating the data on persons engaged in the directory for category $\underline{h}$, or they might be inferred from previous census or survey data etc.

32. A more important application of formula (4) is to evaluate the quality of estimates for all categories of a particular size, 5 per cent, 2 per cent, 1 per cent etc. of the all-industry total. If these results are judged to be unsatisfactory, then the necessary changes in the sample can also be determined with the aid of the formula. For example, if $\underline{CV} = 1.0$, $\underline{R} = 0.5$, and $\underline{n} = 2{,}500$, the question rises what may reasonably be asserted about the relative sampling errors for categories that are each 1 per cent of the all-industry total? In this case the reference is to a class of estimates, so that it is appropriate to use the average value $\underline{R}$ for $\underline{R}_h$. Substituting in formula (4),

$$\hat{\underline{V}}(\underline{X}') \doteq 1.0 \sqrt{0.5/2{,}500} \sqrt{0.5/0.01} = \sqrt{0.25/25} = \sqrt{0.01} = 0.1,$$

so that in this example, industry or area totals that are about 1 per cent of the all-industry total would each have relative standard errors of roughly 10 per cent.

33. Conceivably, that error level will not be considered good enough, particularly when the full implications of the relative standard errors are considered. The sample estimates will be more or less normally distributed about the universe values. They will almost surely, therefore, fall in the range from 30 per cent (three times the relative standard error) below to 30 per cent above the universe value. Narrower ranges can be defined with the increasing risk that the estimates will lie beyond them. About 5 per cent will be outside the -20 per cent to +20 per cent range, and about one third of the estimates will be outside the -10 to +10 per cent range.

34. It is apparent from formula (4) that, if all other factors remain constant, the relative standard errors vary inversely with $\sqrt{\underline{n}}$. Thus, in the above example, to cut the errors from 10 per cent to 5 per cent would require enlarging the sample to nearly 10,000 non-certainty establishments. Actually 10,000 is an upper boundary, because the optimum certainty cut-off and the corresponding $\underline{R}$ will be lower if an increase in the sample size is contemplated. The easiest way to converge to a definite number that will satisfy the reliability specification is to increase the original budgeted total number by a few different values, then to recompute the certainty cut-off $\underline{R}$, and $\hat{\underline{V}}(\underline{X}'_h)$ for each projected level. In the example, increases of 6,000, 7,000 and 7,500 establishments might be tried.

35. Inferences about the behaviour of sample estimates that are based on the above type of analysis will be reasonably close, except for categories that differ markedly from the average. Particular industries, such as bakeries, in which small establishments account for much more than the average proportion of the industries' total activity will tend to have somewhat larger relative standard errors than those given by formula (4). Other industries, in which nearly all the activity is concentrated among large establishments, are likely to have smaller relative standard errors than those indicated by formula (4). Because of certain

simplifications introduced in deriving the formula, it generally overstates the errors. This positive "bias" is minor except for extreme cases, those industries for which all establishments are included with certainty. a/ The correct relative standard error for them, of course, is zero.

36. The sampling plan described above is contingent upon the availability of a substantially complete list of the industrial establishments in the universe. When the available list is incomplete, simply to ignore its deficiency and to leave a considerable gap in the census coverage would be unsatisfactory, so that it will be necessary to supplement the sample drawn from the list with an area sample.

37. The methods are somewhat different from sampling from the directory, because the sampling units in this case are land areas instead of individual establishments. Measures of size, in terms of the number of establishments omitted from the directory, are imputed to each area, and the sample is selected accordingly. The procedure is carried out in stages, starting at the first stage with large units, cities, towns, and the remainder of each province, b/ and working down to smaller sample areas later. Under this plan the small sample areas finally selected are completely enumerated, and census reports are collected from all the industrial establishments found in them that were not covered by the directory.

38. As a first approximation in assigning measures of size, it might be assumed that the rate of omission from the directory is constant for all areas. The assumption should be modified where there is reason to do so, such as for cities that have submitted "good" lists of their industrial establishments, in contrast to others that may not have submitted any list at all. A lower than average (but positive) rate would be justified for the co-operating cities, and a higher than average rate would be appropriate for the non-co-operating ones. Using the adjusted rates and the number of small and medium establishments in the directory for each area, the corresponding number of omitted establishments should be calculated. Other methods for imputing the number of omitted establishments, such as ratios to population figures, might be used for the computations or to check the reasonableness of the results. The imputed figures will necessarily be very rough approximations of the actual number of omitted establishments, but if the data are sensibly correlated the results will be satisfactory.

---

a/ If such industries account for more than 20 per cent of the entire certainty class, instead of substituting $\underline{R}$ for $\underline{R}_h$ in formula (4), substitute $\underline{R}' = \underline{R} - \underline{p}'(1 - \underline{R})$, where $\underline{p}'$ is the proportion of certainty accounted for by the completely certainty industries.

b/ The term "provinces" is used solely for definiteness to refer to well-defined political or administrative geographical areas of intermediate size. Preferably they should be subdivisions with a population of 20,000-100,000 exclusive of the principal cities.

39. After final numbers of establishments have been assigned to the original areas, an attempt should be made to break down each area into a densely industrial part, and a sparsely industrial part. Then the total number of establishments assigned to the area should be distributed to the two parts in appropriate proportions or numbers. For example, in a city the densely industrial part might be an "industrial zone" that is thought to include about 80 per cent of all the industrial establishments; in a non-urban area, the densely industrial part might be a district containing small quarries, in which there are at a guess three or four of nine imputed omissions for the entire area.

40. Splitting areas in this way will yield substantial gains, so that it should be done whenever feasible. However, it must be physically possible to distinguish between the two parts. Boundaries must be drawn between them that can readily be identified in the field. Ease of identification is vastly more important than being able to distinguish precisely between industrial and non-industrial territory. In any case, it will be necessary to subdivide later by means of clear boundaries. At least for cities and large towns (with a population of 25,000 or more) it will usually pay, at this stage, to draw good boundaries around the areas that more or less correspond to the concept of densely industrial. Smaller towns and non-urban areas should also be split when this can be done cheaply and the resulting imputed counts of omitted establishments are not too small. Individual areas to which fewer than three omitted establishments were imputed after splitting should be merged with adjoining or nearby low-number areas until a total of three or more establishments is obtained.

41. After the splitting (and any necessary recombining) has been completed, the final areas and their imputed numbers of omitted establishments should be listed in a sampling table. In this table, the areas should be grouped by regions that emphasize geographical differences in industrial characteristics. Within regions, list the areas by the classifications described above, urban densely industrial, urban sparsely industrial etc. Then divide the number of establishments recorded for each area by three, and round the result to the nearest whole number. That number will be the sampling measure of size $\underline{M}$ for area. Also compute and record in the same table the successive cumulative totals of the measures of size, $\underline{M}_1$, $(\underline{M}_1 + \underline{M}_2)$, $(\underline{M}_1 + \underline{M}_2 + \underline{M}_3)$ etc.

42. Determining the appropriate sampling rate, which is needed next, requires a more detailed consideration of unit costs than sampling only from the directory. For the directory problem, the total number of units to be covered could be found by dividing the total census budget by a single average unit cost, because all the units were establishments which had to be treated uniformly. For the present problem the available funds must be allocated between the area sample and the units selected from the directory, including the certainty establishments, and the cost and variance properties of the different types of units are quite different.

43. No explicit formula for defining large establishments has been given for the purpose of imputing measures of size to areas, because an intuitive definition will suffice. An objective definition, however, is needed to determine the appropriate sample allocation. The method presented earlier is not applicable, but the following formula can be used:

$$X_c = \frac{\sum X_d}{2(\hat{n} - N_c)} \qquad (5)$$

where

X_c is the certainty cut-off value,

$\sum X_d$ is the total value of all the smaller (non-certainty) establishments in the directory,

$\hat{n}$ is the total number of establishments that would be included in the census if the directory were complete. As before it is the total amount available for collecting and processing reports (excluding fixed charges), divided by the average unit cost for an establishment selected from the directory,

N_c is the total number of certainty establishments.

A solution for X_c can be found most easily by trial and error. For any particular value of X_c, there will be one corresponding value of $\sum X_d$ and one of N_c, both of which can be calculated. By experimenting with a few values of X_c, a result that closely satisfies formula (5) can be found without difficulty.

44. After the number of certainty establishments has been ascertained, the residual budget available for the sample of smaller directory establishments and the area sample can be determined, and the approximately optimum sampling rates for those two samples calculated. The formulas are:

Directory sampling rate: $$\frac{n}{N} = \frac{1}{\sqrt{CD}} \cdot \frac{B}{N\sqrt{CD} + 2.5M\sqrt{CA}} \qquad (5.1)$$

Area sampling rate: $$\frac{m}{M} = \frac{2.5}{\sqrt{CA}} \cdot \frac{B}{N\sqrt{CD} + 2.5M\sqrt{CA}} \qquad (5.2)$$

where

B is the budget available for the directory, small establishment sample and the area sample;

N is the total number of small establishments in the directory;

$\overline{CD}$ is the average unit cost of collecting and processing a census report from a directory establishment;

M is the last cumulative total in the area sampling table (and equals the number of sample areas in the universe);

$\overline{CA}$ is the average unit cost of covering a sample area, including listing checking against the directory list, and collecting and processing census reports for the establishments found that were not in the directory (roughly an average of three per sample area);

2.5 is a factor chosen to reflect the larger aggregate values of sample areas than of single establishments selected from the directory, and the greater variance associated with sampling in clusters.

45. The sample from the directory can now be selected by the methods described in paragraphs 24-28, but the procedure for the area sample is different. Those steps are detailed below:

(a) Compute the reciprocal of the sampling rate given by formula (5.2), and round the result to the nearest integer, I, which then will be used as the sampling interval;

(b) Choose a random number between 1 and R inclusive, then add to it, I, 21 etc.; these are the initial sample selection numbers;

(c) Enter each initial sample selection number in the sampling table against the first cumulative total that exceeds or equals it. The corresponding initial area on that line is selected. Note that an initial area may be selected more than once, namely when two or more initial sample selection numbers are recorded for it; the total number of selections recorded, m', should be within 1 of the expected number, M/I;

(d) On a map divide each selected initial area$_i$ into M_i final subareas, where M_i is the measure of size for area$_i$. Use well-defined geographical boundaries exclusively. Roughly equalize the expected number of industrial establishments in each final subarea if the information for doing so without difficulty is readily available;

(e) Number the subareas of the selected initial areas serially, from 1 to L, proceeding continuously from the first subarea of the first selected initial area to the last subarea of the last selected initial area;

(f) Divide L by m' and round the result to the nearest tenth to find the final sampling interval, I';

(g) Choose a random number, R', between 0.1 and I' inclusive, then successively add I', 2I' etc., to R';

(h) After all the numbers in step (g) have been calculated, raise all fractional numbers to the next highest integer. These are the final sample selection numbers. The subareas with corresponding serial numbers are the selected final sample subareas.

46. Before the sample of final subareas is selected, it may be evident that samples drawn from some initial areas will cost far more to enumerate than others. This may typically be the case for non-urban versus urban initial areas. When such situations occur, the sample can be improved by concentrating more of it in the inexpensive areas and reducing it in others. For this purpose, in assembling the initial areas in the sampling table, they should be stratified according to the two classes, inexpensive and expensive. Then separate cumulative totals should be developed for each stratum. The sampling rates for the directory and the two classes of areas would be

Directory rate:
$$\frac{\underline{n}}{\underline{N}} = \frac{\underline{N}}{\sqrt{\overline{CD}}} \cdot \frac{B}{\underline{N}\sqrt{\overline{CD}} + 2.5(\underline{M}_1\sqrt{\overline{CA}_1} + \underline{M}_2\sqrt{\overline{CA}_2})} \quad (6.1)$$

Inexpensive area rate:
$$\frac{\underline{m}_1}{\underline{M}_1} = \frac{2.5\ \underline{M}_1}{\sqrt{\overline{CA}_1}} \cdot \frac{B}{\underline{N}\sqrt{\overline{CD}} + 2.5(\underline{M}_1\sqrt{\overline{CA}_1} + \underline{M}_2\sqrt{\overline{CA}_2})} \quad (6.2)$$

Expensive area rate:
$$\frac{\underline{m}_2}{\underline{M}_2} = \frac{2.5\ \underline{M}_2}{\sqrt{\overline{CA}_2}} \cdot \frac{B}{\underline{N}\sqrt{\overline{CD}} + 2.5(\underline{M}_1\sqrt{\overline{CA}_1} + \underline{M}_2\sqrt{\overline{CA}_2})} \quad (6.3)$$

The procedure given in paragraph 45 would be applied separately to the two strata to select the sample of final subareas.

47. Stratifying by average unit cost, and adjusting the sampling rates accordingly, can be extended to three or more sets of initial areas, and to the directory as well. The additional gains realized, however, will be small unless the differences in the average unit costs and the number of sample units affected are substantial.

48. When an area sample is to be taken to supplement a listing of concentrated industrial areas, the method used to determine the sampling rates or intervals requires some modification. In this case the listings compiled for the concentrated industrial areas, all combined, replace the directory in formulas (5.1) and (5.2) and (6.1), (6.2) and (6.3), and the total budget is the amount available after allowing for the cost of compiling that listing and collecting reports from the certainty list. With those definitions, the formulas can be applied to determine the rate to use in sampling from the listing and in selecting the supplementary area sample. As an extremely rough guide, the computations and the figures assumed in making them should be carefully checked if the calculated rates are not at least 5 per cent or are more than 30 per cent.

49. Great care must be exercised to associate the correct sampling rate with each census report. Even in the simplest case, when the sample is selected entirely from the directory, it is advisable to enter the sampling rates in advance on each report form, together with the establishment's name and address. Comparably tight devices should be applied to guarantee that reports collected in an area sample are accurately and clearly identified by final sample subarea.

50. Exacting control over the recording of the sampling rates is necessary to ensure valid results. The sampling rates correspond to the probabilities of selection discussed in paragraph 4 above. They are the basis for the weights to use in inflating the sample data to universe levels. Specifically, the weights are the reciprocals of the sampling rates. For certainty establishments, the weights are all 1. For other establishments selected from the directory, the weights are the reciprocals of the final sampling fractions discussed in paragraph 25. For the area sample, the correct weights are less obvious because the sampling is done in stages, but they are simply equal to the initial sampling intervals I_1 and I_2 (merely I when a single interval is used), since the probability of selecting any final subarea is $1/I_1$ or $1/I_2$ depending upon its stratum. For maximum flexibility in estimating totals, the easiest procedure, therefore, is first to multiply all the data for every establishment by its weight; then any desired estimate can be produced by tabulating the weighted figures.

51. To be complete, however, the census publications must include information on the reliability of the estimates. This need can be met by computing and publishing estimates of the relative standard errors, with an explanation of how they are to be interpreted (see para. 33). The reliability of the estimated totals for different industries and different geographical areas is likely to vary considerably, so that relative standard errors should be derived and presented for each separately published classification.

52. Ordinarily it will not be necessary to compute and present specific relative standard errors for every estimated total. For any given industry or area, the relative standard errors of the total number of persons engaged and of value added will usually cover the range of the errors for most general statistics. The important exceptions are apt to be stocks (inventories) and capital investments. Relative standard errors for these four items will as a rule suffice to describe the quality of all the general statistics estimates. The reliability of estimates for individual products and individual materials, however, will also vary greatly and their relative standard errors, too, should be computed.

53. Since the relative standard error is given by the absolute standard error divided by its corresponding estimate, and the standard error is defined as the square root of the sampling variance, it is first necessary to estimate the sampling variances.

54. Expressed in a form convenient for computation, the formula for an unbiased estimate of the variance of an estimated total is very nearly:

$$s^2(X') = \sum_{h=1}^{G} \left(\frac{n_h}{n_h - i}\right)(W_h - 1)W_h \sum_{i=1}^{n_h} (X_{hi} - \bar{X}'_h)^2 , \qquad (7)$$

where

G is the total number of strata,

n_h is the number of sample units selected from stratum h,

W_h is the sampling weight for the units selected from stratum h,

X_{hi} is the value of the ith unit selected from stratum h,

$\bar{X}'_h = \sum_{i=1}^{n_h} X_{hi}/n_h$ is the sample estimate of the mean value for stratum h and is equal to the sum of the sample values divided by the sample number for the stratum.

55. It should be carefully noted that n_h refers to the total number of units selected from the stratum, and not to the number classified in the particular industry, area etc. that is being tabulated. It should be noted further that formula (7) also applies to the area sample, but that the units in that case are the final subareas. The data for the individual sample establishments from the same subareas must be summed to subarea totals in order to obtain the X_{hi} values called for by formula (7).

56. A simplified form for the computation of the variance is

$$\ddot{s}^2(X') = \sum_{j=1}^{n} (W_j - 1)W_j X_j^2 , \qquad (7')$$

where

n is the total number of sampling units over all strata combined,

W_j is the sampling weight of establishment j, and

X_j is the value of establishment j.

The factor n_h/n_{h-1} (which is close to 1 when n_h is large) and the subtracting mean term, $\bar{X}'_h$, have been dropped from the variance given in formula (7). Formula (7'), though convenient computationally, is a relatively crude estimate. A somewhat better formula that retains much of the computational advantage of $\ddot{s}^2(X')$ is:

$$\hat{s}^2(X') = \sum_{j=1}^{n} (W_j - 1)W_j (X_j - \hat{X})^2 \ ; \qquad (7'')$$

where $\hat{X}$ is an imputed mean value for all non-certainty units selected from the same weight stratum as unit j, and is assigned independently of the sample data.

57. Sometimes, values can be imputed independently of the sample data to improve the quality of the estimates of the totals. The essential idea is to adjust the estimated totals derived from the sample of census reports by the differences between imputed universe (or stratum) totals and corresponding sample estimates of

those imputed totals. Adjusted estimates so derived are called difference estimates. They may be written in the form

$$\underline{X}'_{\underline{t}} = \underline{X}'_{\underline{t}} - \underline{R}_{\underline{tu}} (\underline{Z}'_{\underline{u}} - \underline{Z}_{\underline{u}}), \tag{8}$$

where

$\underline{Z}_{\underline{u}}$ is a known universe total for item $\underline{u}$,

$\underline{Z}'_{\underline{u}}$ is the estimate of $\underline{Z}_{\underline{u}}$, derived from the sample of units that was selected to report in the census,

$\underline{X}'_{\underline{t}}$ is the estimate of item $\underline{t}$, derived from the sample of census reports,

$\underline{R}_{\underline{tu}}$ is a factor that expresses, approximately, the ratio of $\underline{X}_{\underline{t}}$ to $\underline{Z}_{\underline{u}}$,

$\underline{X}'_{\underline{t}}$ is the adjusted estimate for item $\underline{t}$.

The difference estimate will often be beneficial in making estimates for the non-certainty universe taken from the directory, but it will seldom be helpful for the area sample.

58. Gains can be derived from the difference estimate when good information on industrial classification is available in the directory. In such cases, $\underline{Z}_{\underline{u}}$ can be establishment counts by industry and area, $\underline{Z}'_{\underline{t}}$ the corresponding sample estimates and $\underline{R}_{\underline{tu}}$ "guessed" average values of small establishments for the respective census items, $\underline{t}$. If data on the persons engaged are also available, they advantageously form the basis for the $\underline{Z}_{\underline{u}}$ and $\underline{Z}'_{\underline{u}}$ values, with a corresponding change in the definition of $\underline{R}_{\underline{tu}}$.

59. When difference estimates are computed, formula (7") is used for estimating the sampling variances, with $\underline{R}_{\underline{tu}}\underline{Z}_{\underline{i}}$ instead of $\underline{X}$.

60. In developing the estimates, and the estimates of the sampling variances, where area sampling is used to supplement a listing of concentrated industrial areas, the sample establishments selected from the lists for the concentrated industrial areas can be processed as though they had been selected from a directory, for the complete listing of those areas is equivalent to a directory. The methods described in paragraphs 57-59, in particular, can be used with great advantage.

*

* *

Note: The above discussion is specifically directed and mainly limited to practical techniques for applying sampling in industrial censuses. It would also be helpful to read Recommendations for the Preparation of Sample Survey Reports c/ which gives some other methods that may be useful and discusses some related issues that are not covered here.

---

c/ Statistical Papers, Series C, No. 1/Rev.2, provisional issue (United Nations publication, Sales No. E.64.XVII.7).

Annex II

SPECIMEN QUESTIONNAIRES

A. Minimum programme

(Penalty for failure to report)	DUE DATE:
FORM NO. (Date) NATIONAL STATISTICAL OFFICE (Name of country) 1983 CENSUS OF INDUSTRY	NOTICE - Response to this inquiry is required by Law. By the same Law, your report to the National Statistical Office is confidential. It may be seen only by sworn employees of the office and may be used only for statistical purposes Identification No. | Industry | Prov. | Cnty. | TEP a/ Size

A separate census report must be filed for each (a) mining and quarrying, including oil and gas wells, (b) manufacturing and (c) electricity, gas and water supply establishment operated at any time during 1983.

If book figures are not available, carefully prepared estimates are acceptable.

ITEM 1. DESCRIPTION OF ESTABLISHMENT

Item 1a. BUSINESS NAME OF ESTABLISHMENT COVERED BY THIS REPORT

_____

Item 1b. PHYSICAL LOCATION OF ESTABLISHMENT (Be sure that the answer relates to the physical location of this establishment, which may be different from the mailing address)

STREET AND NUMBER (If location cannot be described by street and number, give as much information as possible, such as name of road or highway and distance from nearest city, town or village)

_____

CITY, TOWN OR VILLAGE _____

COUNTY (OR DISTRICT) _____ PROVINCE (OR STATE) _____

Item 1c. KIND OF BUSINESS ACTIVITY IN 1983

Describe the major activity of this establishment during 1983:

_____

List the principal products made or services rendered by this estabishment during 1983 (list in descending order of importance):

_____

_____

a/ Total number of persons engaged.

ITEM 2. NUMBER OF PERSONS ENGAGED

Enter the number of different persons engaged in the activities of this establishment during the week ending nearest _____ (a specified date in the last quarter of the reference year such as November 15).
"Working proprietors" include proprietors and partners who are actively engaged in the work of the establishment. "Unpaid family workers" include all persons living in the household of any of the owners and working in the establishment without regular pay for at least one third of the working time normal to the establishment. "Employees" include all persons who received pay for work for the establishment during the specified reference week; include all employees on sick leave, holidays or vacation.

Category of persons engaged	Key	Total number
a. Working proprietors, unpaid family workers	2-1	
b. Employees	2-2	
c. TOTAL (sum of a and b)	2-3	

ITEM 3. WAGES AND SALARIES PAID IN 1983

Enter the total of all payments, whether in cash or in kind, made by the employer during 1983 in connexion with work done, to all persons included in the count of employees (item 2(b)). Include all cash payments, bonuses, cost-of-living allowances and wages paid during vacations and sick leave; taxes, social security contributions, group insurance premiums and the like, payable by the employee but deducted by the employer; and payments in kind. Estimate the cash value of payments in kind according to their cost to the establishment.

Category of persons engaged	Key	Wages and salaries paid during 1983
All employees	3-1	

ITEM 4. VALUE OF SHIPMENTS AND RECEIPTS OF THIS ESTABLISHMENT DURING 1983

Enter the value of all products shipped and receipts for services performed by this establishment during 1983. The value should be the establishment price charged to the customer, net of rebates and discounts, but including taxes and duties which fall on the products as they leave the establishment. Include receipts for work done for others on their materials, receipts for repair and installation work done for others, sales of products bought and resold in the same condition and sales of scrap and refuse. Include transfers to other establishments of your enterprise at estimated market value. Exclude any subsidies received in support of production.

Value of shipments and receipts in 1983: _____	Key 4-1

ITEM 5. CERTIFICATION

This report is substantially accurate and covers the period from _____ to _____

_____ Signature of authorized person

_____ Title

Date _____

B. Complete programme

(Penalty for failure to report)

DUE DATE:

FORM NO.
(Date printed)

NATIONAL STATISTICAL OFFICE
(Name of country)

1983 CENSUS OF INDUSTRY

IMPORTANT - PLEASE READ

A report is to be completed for each establishment, i.e., single physical location, where (a) mining and quarrying (including oil and gas wells), (b) manufacturing (including fabricating, processing and assembling) or (c) the production and distribution of electricity, gas and water is performed.

Ancillary units, such as warehouses, research laboratories etc., at the same location should be included as part of the establishment.

PLEASE READ INSTRUCTIONS ACCOMPANYING EACH ITEM BEFORE COMPLETING.

If book figures are not available, carefully prepared estimates are acceptable.

Complete every item, entering "0" where applicable.

CONFIDENTIAL - You are required by law (cite act or decree) to submit the information called for in this report to authorized representatives of (name of government department). This (act or decree) also makes it unlawful for representatives of the National Statistical Office (or other government department) to disclose any facts (except for name, address or kind of business [omit or modify according to country practice]) from this report. Your report cannot be used for purposes of taxation, regulation or investigation.

In correspondence pertaining to this report, please refer to this file number

Note: A mailing stencil would be imprinted here for inquiries conducted by mail; or for the mail multi-unit enterprise portion of censuses when single-unit enterprises are enumerated in the field. In such cases, the first sentence following "CONFIDENTIAL" above would end with the word "report".

Please correct any errors in name or address. Enter street and number if not shown.

FOR OFFICE USE ONLY	Establishment Identification No.	Industry code	Province	County	TEP SIZE

SECTION I - DESCRIPTION OF ESTABLISHMENT

ITEM 1 (a). BUSINESS NAME OF ESTABLISHMENT (MINE, FACTORY, PLANT OR MILL ETC.) COVERED BY THIS REPORT

ITEM 1 (b). PHYSICAL LOCATION OF ESTABLISHMENT (Be sure your answer relates to the actual PHYSICAL LOCATION of this establishment, which may be different from the mailing address.)

STREET AND NUMBER (If location cannot be described by street and number, give as much information as possible, such as name (and number, if any) of road or highway and distance from nearest city, town or village).

CITY, TOWN OR VILLAGE

COUNTY (OR DISTRICT) | PROVINCE (OR STATE)

ITEM 1 (c). ECONOMIC ORGANIZATION OF THIS ESTABLISHMENT
(Mark (X) one box and enter information as appropriate) /101/

1 ☐ Single establishment - does not own or control another establishment and is not controlled by another establishment

2 ☐ Head office - owns or controls other establishments

3 ☐ This establishment is owned or controlled by another establishment

If box 2 is marked, please provide below the name(s) and address(es) of the establishment(s) controlled by this establishment.

Name of establishment	Address

ITEM 1 (c). (continued)

If box 3 is marked, give name and address of head office of owning enterprise

Name of head office

Address (street and number)

Province (or State)

ITEM 1 (d). FORM OF OWNERSHIP

a. WHAT IS THE FORM OF OWNERSHIP OF THIS ESTABLISHMENT?
(Mark (X) one box based on majority ownership) /102/

1 ☐ Owned by nationals

2 ☐ Foreign ownership

b. If "Owned by nationals" is marked, does the establishment belong to a (Mark (X) one box) /103/

1 ☐ Public enterprise ?

2 ☐ Private enterprise ?

ITEM 1 (e). LEGAL ORGANIZATION

If "private enterprise" is marked above, what is the legal form of organization?
(Mark (X) one box which best describes the legal organization) /104/

1 Unincorporated

☐ a. Individual proprietor

☐ b. Ordinary (unlimited) partnership

2 Incorporated (other than co-operative)

☐ a. Limited liability partnership

☐ b. Joint stock company

☐ c. Corporation

☐ 3 Co-operative

SECTION I - DESCRIPTION OF ESTABLISHMENT (continued)

ITEM 1(f). PERIOD OF OPERATION DURING 1983 AND CHANGE OF OWNERSHIP (IF ANY)

a. Period of operation during 1983

(Mark (X) one box and enter information as appropriate)

105

1 ☐ Operated throughout 1983

2 ☐ Operated for only part of 1983 (enter number of months) ☐

3 ☐ Establishment was idle during all of 1983

b. Change in ownership

107 Was this establishment purchased or sold by you in 1983?

(If Yes, mark (X) one box and enter information as appropriate.)

1 ☐ Purchased by you during (enter month) ☐

2 ☐ Sold by you during (enter month) ☐

NOTE If Box "1" is marked, give former owner's name and address below.
If Box "2" is marked, give purchaser's name and address below.

_____

_____

ITEM 1(g). INDUSTRIAL PROCESSES EMPLOYED IN 1983

Describe briefly, in the space provided below, the principal processes used in producing the goods made by this establishment. For each process shown, be sure to indicate whenever hand processes are used. (Examples of such descriptions would include: stemming and drying cigar leaf tobacco (hand); weaving wool yarn for carpets; upholstering wooden household furniture (hand); bottling carbonated beverages; knitting cotton yarn into gloves and sweaters etc. .)

_____

_____

_____

_____

_____

SECTION II - EMPLOYMENT AND EARNINGS

ITEM 2(a). NUMBER OF PERSONS ENGAGED IN ACTIVITIES OF THIS ESTABLISHMENT DURING PAY PERIOD (OR CALENDAR WEEK) ENDING NEAREST 15 NOVEMBER 1983

PERSONS ENGAGED (exclude homeworkers)

Working proprietors and active partners - Such individuals who are actively engaged in the management of the establishment. Managers and directors of corporations working for pay should be reported with "All other employees".

Unpaid family workers - All persons living in the household of any of the owners and working in the establishment without regular pay for at least one third of the normal working time for the establishment during the period. Non-household members who work without pay may also be included here.

Employees - All persons working in the establishment and receiving pay as well as persons working away from the establishment when paid by and under the control of the establishment (except homeworkers). Include all such persons, whether working full-time or part-time. Include all employees on sick leave, paid vacations or holidays.

Operatives - All employees directly engaged in the production, in contrast to the overhead, activities of the establishment. Include manual workers, clerical personnel and working foremen who fabricate, process, assemble, construct, install etc.. Exclude supervisory employees above the working foreman level. (In some countries, operatives may be defined as "wage-earners" or "hourly-rated employees" or "production workers".)

All other employees - All employees other than operatives as defined above. (In some countries, this category of employees may be termed "salaried personnel", "yearly-rated employees" or "non-production workers".)

Category of person engaged	Key	Total number
a. Working proprietors and active partners	201	
b. Unpaid family workers	202	
c. Operatives	203	
d. All other employees	204	
e. TOTAL (sum of a to d)	209	

NOTE (1). Countries in which the number of homeworkers is significant may wish to add a line for homeworkers as f after TOTAL, which would become "TOTAL, except homeworkers". Homeworkers would include all persons who worked for the establishment in their own homes during the specified period.

NOTE (2). Some countries may wish to provide additional columns to call separately for "Male and female", "Adults and youths", "Nationals and foreigners" etc..

ITEM 2(b). NUMBER OF EMPLOYEES, QUARTERLY, 1983

(Report the number of operatives and all other employees as defined in item (a) during the pay period or calendar week ending nearest the 15th of the months specified below.)

Pay period (or calendar week) ending nearest:	Key	Number of operatives	Number of all other employees	Number of all employees
a. 15 February 1983	211			
b. 15 May 1983	212			
c. 15 August 1983	213			
d. 15 November 1983	214			
FOR OFFICE USE ONLY	219			

ITEM 3. WAGES AND SALARIES PAID TO EMPLOYEES DURING 1983

Report all payments, whether in cash or in kind, to all persons included in the count of employees for work done during 1983. Include all cash payments, bonuses, cost-of-living allowances and wages paid during periods of vacation and sick leave; taxes, social security contributions, group insurance premiums and the like, payable by the employee but deducted by the employer; and payments in kind. Report as payments in kind the estimated net cost to the employer of those goods and services furnished to employees free of charge or at markedly reduced cost which are clearly of benefit to employees as consumers.

Report figures for operatives and all other employees in a manner consistent with the breakdown shown in items 2(a) and 2(b).

Category of employee	Key	Wages and salaries paid during 1983 (in thousands of _____)
a. Operatives	301	
b. All other employees	302	
c. TOTAL (sum of a and b)	309	

Note: Payments to "homeworkers" should be included in item 8d (section V).

ITEM 4. SUPPLEMENTS TO WAGES AND SALARIES DURING 1983 (NOT INCLUDED IN ITEM 3)

NOTE: THIS IS AN OPTIONAL ITEM, ESPECIALLY FOR COUNTRIES WITH DEVELOPING INDUSTRIAL STATISTICS

Include employers' contributions on behalf of employees to social security schemes, private pension plans, unemployment compensation and workmen's accident compensation plans, insurance premiums on hospital and medical plans, premiums on supplementary accident and sickness insurance and life insurance premiums. Include all such programmes, whether legally required, initiated by the employer, or established as a result of a labour-management bargaining contract.

Type of payment	Key	Payments during 1983 (in thousands of _____)
a. Legally required under social security system	401	
b. Payments for other programmes	402	
c. TOTAL (sum of a and b)	409	

ITEM 5. NUMBER OF HOURS (DAYS) WORKED BY OPERATIVES DURING 1983

Include all time actually spent by operatives working for the establishment, including scheduled time, overtime and waiting time. One hour is an hour of time spent at work regardless of how it is counted for purposes of payment. Exclude time paid for but not spent in work for the establishment such as sick leave and paid vacations. Operatives are defined in item 2a.

NOTE: Although hours worked is the preferred concept, days worked is acceptable as an alternative measure for countries that find that hours worked cannot be reliably supplied by a large number of establishments. Where days worked is used, this item is defined as the total number of days actually spent by operatives at work but a distinction should be made between full-time and part-time workers.

Category of employee	Key	Hours (days) worked during 1983
Operatives	501	

SECTION III - STOCKS

ITEM 6. STOCKS OF THIS ESTABLISHMENT

Include the value of all stocks held by, or under the control of, the reporting establishment (at the establishment itself or in ancillary locations such as warehouses). Include materials owned by this establishment but held by others for processing. Exclude materials owned by others but held by this establishment for processing.

Stocks of materials etc. should be valued at current replacement cost in purchasers' prices.* Work in progress should reflect an imputed valuation in producers' prices.* Finished goods and goods for resale should be valued in the producers' prices* at which goods have been shipped immediately prior to the reference date. Alternatively, the book values may be reported. Valuation should be on a comparable basis for both beginning and end of the year.

*Purchasers' and producers' prices are defined under items 8 and 13, respectively.

Type of stock	Key	Value of stocks (in thousands of _____) As at 1 Jan.1983	Key	As at 31 Dec.1983
a. Materials, components, parts, containers, supplies and fuels	601		611	
b. Work in progress	602		612	
c. Finished goods (except those for resale)	603		613	
d. Goods to be sold in same condition as purchased	604		614	
e. TOTAL STOCKS (sum of a to d)	609		619	

SECTION IV - FIXED CAPITAL FORMATION

ITEM 7. GROSS ADDITIONS TO FIXED ASSETS DURING 1983

Include all physical assets expected to last more than one year or charged to capital account. Report the capital goods delivered to the control of the establishment during 1983, including the progress payments applicable to the year for construction work. Include purchase price, taxes and duties, delivery and installation charges.

Include the cost of major additions, alterations and improvements to existing fixed assets that are charged to capital account. Exclude expenditure for current maintenance and repair.

Include the cost of such work done on own account, i.e., the value of new fixed assets and additions and improvements to existing fixed assets made by the establishment's own labour force. Report either the cost of labour, materials consumed and appropriate overheads or the charge made to capital (fixed asset) accounts.

Type of transaction	Key	Total	Land	Buildings and other construction	Machinery and equipment (except transport)	Transport equipment
		(in thousands of _____)				
a. Total cost of new fixed assets purchased (not used before in this country)	701					
b. Total cost of second-hand fixed assets purchased (used before in this country)	702					
c. Total cost of alterations, renovations and improvements purchased from others	703					
d. Total cost of work done on own account	704					
e. Value of sales of fixed assets of this establishment	705					
f. Gross additions to this establishment's fixed assets (sum of a, b, c and d minus e)	709					

ITEM 8. COST OF MATERIALS, SUPPLIES, FUELS, ELECTRICITY, CONTRACT WORK ETC. DURING 1983

Report purchases, at current purchased values, of all types of commodities, fuels, electricity etc. purchased by this establishment, including materials and components used in own-account production of fixed assets or used by other establishments in making products for this establishment under contract.

Include all materials etc. received during 1983, whether purchased from others or transferred from other establishments of the enterprise owning this establishment. Include all raw materials, fabricated parts and components that enter into the products made by the establishment; non-durable tools and repair, maintenance, shop, office and other supplies such as lubricants, water, explosives and polishes; containers and packing materials. Include in item 8 a fuels that enter the product directly or are used to generate electricity.

Report in purchasers' prices, that is, the delivered cost at the establishment, including the purchase price with deduction for rebates and discounts received and additions for transport, insurance and unloading charges paid, as well as all taxes and duties.

Type of cost	Key	Delivered cost at the establishment (in thousands of _____)
a. Materials, parts, components, containers, supplies etc. purchased (same as item 11, line 7)	801	
b. Purchased fuels (same as item 9 g)	802	
c. Purchased electricity (same as item 10 a)	803	
d. Cost of contract and commission work done for you by others on your materials	804	
e. Cost of repair and maintenance work done for you by others	805	
f. Cost of goods bought to be sold in the same condition as purchased	806	
g. TOTAL COST OF MATERIALS ETC. (sum of a to f)	809	

NOTE: SOME COUNTRIES MAY WISH TO SEPARATE PAYMENTS TO HOMEWORKERS FROM 8 d

ITEM 9. FUELS PURCHASED DURING 1983

Follow the instructions for valuation of purchases specified in the heading for item 8. Exclude: (a) fuels produced and consumed in the same establishment; and (b) fuels purchased as raw material to enter into the products of the establishment as well as fuels used to generate electricity

Type of fuel	Key	Unit of measure for quantities	Quantity	Delivered cost (in thousands of _____)
a. Coal	901			
b. Coke (including breeze)	902			
c. Fuel oil	903			
d. Natural gas	904			
e. Liquefied petroleum gases (propane, butane etc.)	905			
f. Other fuels (steam, wood gasoline etc.)	906			
g. TOTAL COST OF FUELS (should equal cost shown in item 8 b)	909			

ITEM 10. ELECTRICITY PURCHASED, GENERATED AND SOLD DURING 1983

Description	Key	Quantity (in thousands of kWh)	Value (in thousands of _____)
a. Purchased (value equal to item 8 c)	1001		
b. Generated, whether for own use or for sale	1002		
c. Sold or transferred to other establishments of the same enterprise	1003		
d. Consumed (sum of a and b minus c)	1009		

ITEM 11. QUANTITY AND COST OF PRINCIPAL MATERIALS PURCHASED IN 1983

Enter in line 7 the total cost of materials, parts, components, containers, supplies etc. purchased during 1983 (same as in item 8a). Then refer to the accompanying materials reference list and use it as a guide to determine the appropriate entries in column (B) below. Enter in column (C) the designated material code for the materials category selected from the reference list and in column (D) the physical unit of measure as specified on the list. Then proceed to complete columns (E) and (F) with the quantity and cost of the material purchased during 1983. Follow the reporting instructions, including the valuation of purchases shown for item 8 above.

Line No. (A)	Description of material (B)	Material code (C)	Unit of measure (D)	Quantity purchased (E)	Delivered cost (in thousands of _____)
1					
2					
3					
4					
5					
6	Other materials, supplies etc. purchased (difference between line 7 and the sum of lines 1-5)				
7	TOTAL COST OF MATERIALS, PARTS, COMPONENTS, CONTAINERS, SUPPLIES ETC. PURCHASED (should be same as item 8 a)				

ITEM 12. COST OF NON-INDUSTRIAL SERVICES DURING 1983 OPTIONAL ITEM

Enter below the following types of costs for which information is available in your records: communication charges (postal, telephone, telegraph); expenses for renting capital goods and non-residential buildings; bank charges other than interest; service charges in respect of casualty insurance (difference between premiums paid and claims received); amounts paid for storage and warehousing; cleaning costs and disposal of waste; printing costs, newspaper and periodical subscriptions; advertising, legal, accountancy, consulting services etc..

Description of service	Key	Value (in thousands of _____)
a. Communication costs	1201	
b. Rental payments for buildings and machinery	1202	
c. Other (specify):	1203	
1		
2		
3		
4		
5		
d. TOTAL VALUE OF NON-INDUSTRIAL COSTS (sum of a, b and c.1 to c.5)	1209	

SECTION VI - SHIPMENTS AND RECEIPTS OF THIS ESTABLISHMENT

ITEM 13. PRODUCTS, CONTRACT WORK AND MISCELLANEOUS RECEIPTS DURING 1983

Enter below the value of shipments and receipts of this establishment from products produced, industrial services and other receipts. Refer to the product reference list to obtain the product code numbers to be entered in column (B) for items a.1 to a.8 and the appropriate unit of measure for column (C). The list also serves as a guide to your description in column (A) of the products produced in this establishment in 1983. Use a separate line for each product produced and enter in columns (D) and (E) the corresponding quantity and value of such products. The value figures shown in column (E) should be in producers' prices, that is, at the establishment price charged to the consumer. Value should be net of discounts and allowances and exclude transport charges to the purchaser or user. Include all duties and taxes which fall on products when they leave the establishment. Exclude any subsidies received by the establishment. Receipts for contract work and miscellaneous receipts should be reported in items b to d. Report resales (products bought and resold without further manufacture or processing in this establishment) in item e. Report transfers of products to other establishments of the same enterprise at estimated market value. Include products made elsewhere for this establishment on a contract basis from materials supplied by this establishment.

Description of item (A)	Product code (for items a.1 to a.8) or key (B)	Unit of measure (C)	Quantity (D)	Value (in thousands of _____) (E)
a. Value of shipments of products made in this establishment:				
1				
2 NOTE: SOME COUNTRIES MAY WISH TO PREPRINT A LIST OF IMPORTANT PRODUCTS AND THEIR CODES				
3				
4				
5				
6				
7				
8				
9 All other products				
10 Total value of shipments (sum of a.1 to a.9)	1301			
b. Receipts for contract work done for others on their materials	1302			
c. Receipts for repair and installation work done for others	1303			
d. Other receipts for industrial services including sales of scrap and refuse	1304			
e. Sales of goods bought and sold in the same condition	1305			
f. Value of own-account investment work (same as item 7 d)	1306			
g. TOTAL VALUE OF SHIPMENTS AND RECEIPTS (sum of a to f)	1309			

ITEM 14. RECEIPTS FOR NON-INDUSTRIAL SERVICES DURING 1983 **OPTIONAL ITEM**

Enter below the following types of revenue recorded in your establishment's accounts: revenue from the rental or lease of industrial equipment and buildings; commissions received; revenue from the operation of employee facilities; receipts for storage of goods, warehousing, cold storage; receipts for transport services rendered to others, other than delivery of own product; other similar items. Do not include the following: dividends, interest or subsidies received; revenue from the sale of used capital goods.

Description of service	Key	Value (in thousands of _____)
a. Receipts for rental or lease of buildings and machinery	1401	
b. Commissions received	1402	
c. Other (specify):	1403	
1		
2		
3		
4		
5		
d. TOTAL VALUE OF NON-INDUSTRIAL RECEIPTS (sum of a, b and c.1 to c.5)	1409	

ITEM 15. CHECKS TO ENSURE A COMPLETE AND ACCURATE REPORT

The National Statistical Office reviews your report for omissions, inconsistencies and unusual ratios. To save you future correspondence about these problems, please make the following checks before returning your report:

(1) Review the report carefully to see that no items are omitted for the year being covered.

(2) Calculate and enter figures for the four "reasonableness" checks below, and review the results.

(3) Correct your report for any errors you find and explain unusual figures in item 16, "Remarks."

For most industrial establishments, the figures computed usually fall within the ranges indicated.

"Reasonableness" checks (A)	Calculations required for each check (B)	Figures for survey year (C)	Usual range* (D)
A. Average annual wage per operative	Item 3 a (Operatives' wages) ÷ item 2 (a) c (Number of operatives)		
B. Wages and salaries as a percentage of value of shipments and receipts	Item 3 c (Wages and salaries) ÷ item 13 g (Value of shipments and receipts)	Per cent	Per cent
C. Cost of materials and services as a percentage of value of shipments and receipts	Item 8 g (Cost of materials and services) ÷ item 13 g (Value of shipments and receipts)	Per cent	Per cent
D. Is the total for Value of shipments and receipts (item 13 (g)) greater than the sum for Wages and salaries (item 3 c) plus Cost of materials and services (item 8 g)?		☐ YES	☐ NO (Explain in "Remarks")

* To be entered on the printed forms by the National Statistical Office. The range should be wide enough to cover almost all possible correct answers, allowing for industry and regional differences within the country but still be set to detect serious errors such as reporting in the wrong unit of measure. (The "usual range" for (B) and (C) above will generally be from 5 per cent to 95 per cent.) Some countries may choose to drop the "usual range" column from this inquiry.

NOTE: Countries, particularly developing ones, may decide to limit this item to two or three questions. In selecting the questions, some countries may decide to emphasize: (a) consistency in totals that occur in more than one place on the form; or (b) the fact that a detailed breakdown should add to the total reported for the item.

ITEM 16. REMARKS

ITEM 17. CONTACT	Name of person to contact regarding this report		Address	Telephone number
ITEM 18. CERTIFICATION	This report is substantially accurate and covers the period		Signature	
	From	To	Title	Date

BIBLIOGRAPHY

Industrial and distributive-trade statistics

United Nations. International recommendations for industrial statistics, 54 p. (Statistical papers, Series M, No. 48) Sales No. E.68.XVII.10.

_____. Recommendations for the 1973 world programme of industrial statistics, Part I: General statistical objectives, 43 p. (Statistical papers, Series M, No. 54 (Part I)) Sales No. E.71.XVII.13.

_____. Recommendations for the 1973 world programme of industrial statistics, Part II: List of selected products and materials, 127 p. (Statistical papers, Series M, No. 54 (Part II)) Sales No. E.71.XVII.16.

_____. Recommendations for the 1973 world programme of industrial statistics, Part III: Organization and conduct of industrial censuses, 111 p. (Statistical papers, Series M, No. 54 (Part III)) Sales No. E.72.XVII.10.

_____. The 1973 world programme of industrial statistics: summary of data from selected countries, 525 p. (Statistical papers, Series P, No. 15) Sales No. E.79.XVII.3.

_____. International recommendations on statistics of the distributive trades and services, 55 p. (Statistical papers, Series M, No. 57) Sales No. E.75.XVII.9.

_____. Organization and conduct of distributive-trade surveys, 165 p. (Studies in Methods, Series F, No. 19) Sales No. E.77.XVII.3.

_____. Bibliography of industrial and distributive-trade statistics, 177 p. (Statistical papers, Series M, No. 36, Rev. 4) Sales No. E.75.XVII.7.

Classification systems

_____. International standard industrial classification of all economic activities, 48 p. (Statistical papers, Series M, No. 4, Rev. 2) Sales No. E.68.XVII.8.

_____. Indexes to the international standard industrial classification of all economic activities, 247 p. (Statistical papers, Series M, No. 4, Rev.2, Add.1) Sales No. E.71.XVII.8.

_____. Classification by broad economic categories (defined in terms of SITC, Rev.2), 53 p. (Statistical papers, Series M, No. 53, Rev.1) Sales No. E.76.XVII.7.

_____. Classification of commodities by industrial origin: links between the standard international trade classification and the international standard industrial classification, 145 p. (Statistical papers, Series M, No. 43, Rev.1) Sales No. E.71.XVII.15.

_____. International standard classification of all goods and services (ICGS) (draft) (E/CN.3/493).
To be issued as a United Nations publication.

_____. Standard international trade classification, revision 2, 117 p. (Statistical papers, Series M, No. 34, Rev.2) Sales No. E.75.XVII.6.

Sampling

_____. A short manual on sampling, vol. I: Elements of sample survey theory, 325 p. (Studies in methods, Series F, No. 9, Rev.1) Sales No. E.72.XVII.5.

_____. A short manual on sampling, vol. II: Computer programmes for sample designs, 77 p. (Studies in methods, Series F, No. 9) Sales No. E.71.XVII.4.

_____. A short manual on sampling, vol. II, Add.1: Computer programmes for sample designs, 46 p. (Studies in methods, Series F, No. 9, vol. II, Add.1) Sales No. E.73.XVII.8.

_____. Recommendations for the preparation of sample survey reports, provisional issue, 16 p. (Statistical papers, Series C, No. 1, Rev.2) Sales No. E.64.XVII.7.

_____. Sample surveys of current interest, thirteenth report, 218 p. (Statistical papers, Series C, No. 14) Sales No. E.79.XVII.2.

Litho in United Nations, New York
16592–October 1981–2,750

Price: $U.S. 11.00

United Nations publication
Sales No. E.81.XVII.12
ST/ESA/STAT/SER.M/71(Part II)