

CONTRIBUTED PAPERS

SEVENTH CONFERENCE OF
**COMMONWEALTH
STATISTICIANS**

November 16-28, 1970

VOLUME III
ITEMS 10-14 OF AGENDA

CENTRAL STATISTICAL
ORGANISATION
NEW DELHI

RESTRICTED

Four documents relate to the work of the Conference :

1. Report of Conference :—A technical report of the proceedings highlighting the various issues discussed at the Conference. Report is printed for circulation to delegates/observers, the national statistical Offices of the Commonwealth Countries and the general public.

2. Contributed Papers :—Submitted in writing to the Conference by various countries :

- A. Papers contributed on Items 2 to 5 of the agenda (Volume I).
- B. Papers contributed on Items 6 to 9 of the agenda and papers contributed on items of the agenda which were not discussed in the Conference (Volume II).
- C. Papers Contributed on Items 10 to 14 of the agenda and papers contributed on items which were not discussed in the Conference (Volume III).

Circulation of these papers is limited to delegates and official observers who attended the Conference, the statistical Offices of all Commonwealth Countries and at their initiative and discretion to the Statistical Associations in their respective Countries.

Issued by the Central Statistical Organisation,
Deptt. of Statistics, Cabinet Secretariat,
NEW DELHI, (INDIA).

SEVENTH CONFERENCE
OF
COMMONWEALTH STATISTICIANS
NEW DELHI, NOVEMBER 16-28, 1970
VOLUME III
Items 10-14 of the Agenda

CONTENTS

		Page
PROGRESS AND PLANS IN DEALING WITH THE REVISED SYSTEM OF NATIONAL ACCOUNTS.		
Australia.	Progress and plans in dealing with the revised system of National Accounts.	728
Canada.	Progress and plans in dealing with the revised system of National Accounts.	740
India.	Progress and plans in dealing with the revised system of National Accounts.	748
Kenya.	Progress and plans in dealing with the revised system of National Accounts.	757
Lesotho.	Progress and plans in dealing with the revised system of National Accounts.	762
New Zealand.	Progress and plans in dealing with the revised system of National Accounts.	764
Tanzania.	Progress and plans in dealing with the revised system of National Accounts.	771
United Kingdom.	Progress and plans in dealing with the revised system of National Accounts.	773
SYSTEMS OF REGIONAL ACCOUNTS.		
Australia.	Systems of regional accounts.	777
Canada.	Approaches to regional economic accounting in Canada.	786
India.	Systems of regional accounts.	792
Tanzania.	Systems of regional accounts.	803
United Kingdom.	Systems of regional accounts.	804
THE GENERAL QUESTION OF CONFIDENTIALITY ESPECIALLY AS INVOLVED IN THE USE OF COMPUTERS.		
Australia.	The general question of confidentiality, especially as involved in the use of computers.	

CONTENTS

Item of Agenda		Page
Canada.	The general question of confidentiality especially as involved in the use of computers.	
New Zealand.	The general question of confidentiality especially as involved in the uses of computers.	
Tanzania.	The general question of confidentiality especially as involved in the use of computers.	
United Kingdom.	Confidentiality of business statistics.	
12. INTERNATIONAL TRADE STATISTICS AND THE S.I.T.C.		
Australia.	International trade statistics and the S.I.T.C.	
Canada.	International trade statistics and the S.I.T.C.	
India.	International trade statistics and the S.I.T.C.	
New Zealand.	Proposed revision of the Standard International Trade Classification - Revised.	
Tanzania.	International trade statistics and the S.I.T.C.	
United Kingdom.	International trade statistics and the United Nations Standard International Trade Classification.	
13. THE NEED FOR INDEPENDENT CROSS-CHECKS OF STATISTICS FROM DIFFERENT SOURCES. THE EVALUATION OF THE ACCURACY OF SURVEY RESULTS.		
Australia.	Need for independent cross-checks of statistics from different sources; the evaluation of the accuracy of survey results.	
Canada.	The evaluation of the accuracy of survey results.	
India.	The need for independent cross-checks of statistics from different sources. The evaluation of the accuracy of survey results.	
Lesotho.	The need for independent cross-checks of statistics from different sources. The evaluation of the accuracy of survey results.	
New Zealand.	The need for independent cross-checks of statistics from different sources. The evaluation of the accuracy of survey results.	
Tanzania.	The need for independent cross-checks of statistics from different sources.	
United Kingdom.	The need for independent cross-checks of statistics from different sources. The evaluation of the accuracy of survey results.	

CONTENTS

Item of Agenda		Page
14(a)	CONSTRUCTION.	
	New Zealand. Statistics of the building and construction industry.	
14(b)	TOURISM	
	Kenya. Statistics of Tourism. A proposed system for collecting migration statistics in East Africa.	
	New Zealand. Statistics of tourism.	
14(d)	POPULATION PROJECTIONS .	
	India. Projections of population for national and sub-national areas, India.	
	Ireland. Population projections in Ireland.	
	New Zealand. Brief history of official population projections in New Zealand.	
14(e)	PERSONNEL TRAINING.	
	New Zealand. Personnel training.	
14(f)	THE PROBLEM OF MAINTAINING HIGH STANDARDS OF SURVEY METHODOLOGY, ANALYSIS AND COMPILATION TECHNIQUES.	
	India. Problems of maintaining high standards of survey methodology, analysis and compilation techniques.	
	New Zealand. Problems of maintaining high standards of survey methodology, analysis and compilation techniques.	
	Nigeria. Proposed all Nigeria Urban household budget survey.	
Prof. P.C. Mahalanobis.	A note on disparities in the level of living. (8(b)-India)	

PROGRESS AND PLANS IN DEALING WITH THE
REVISED SYSTEM OF NATIONAL ACCOUNTS

It is the present intention that Australia should in due course adopt the United Nations Revised System of National Accounts (SNA) as far as possible. However, some marginal variations from SNA principle, definitions and classifications will be made; some tables will be modified or compiled in part only; and some tables will not be compiled at all in the foreseeable future. But the intention will be that the SNA can still be used as a general framework and guide for purposes of exposition and presentation.

At least two stages are involved in the process of working towards the adoption of the SNA. In the short run there is the problem of re-arranging data in the existing system of the Bureau of Census and Statistics in order to report to the U.N. Statistical Office in accordance with the revised SNA. In the longer run there is the problem of adapting the whole system, data base, and estimating methods to the revised SNA.

2. There are several reasons in favour of adopting the SNA for national purposes. First, the Bureau considers that international reporting should be treated seriously, not least because of the requests it receives within Australia for comparable statistics for Australia and other countries. In such cases, U.N. publications are usually consulted and it is unsatisfactory to have to explain differences between Australian national figures and those used for international comparisons. Secondly, it is obvious that reporting internationally must become easier the closer one can bring the national system to the SNA. Thirdly, after several years' experience with the existing system the Bureau sees the need for a number of changes. In making any changes at all it is essential to have them built into a coherent system and the elaboration of any such system is a complicated and time-consuming task. The Bureau has participated fully in the elaboration of the SNA. While not in complete agreement with it, the Bureau would wish to avoid revising and extending its present system in such a way that it would be involved in explaining a separate system and at the same time trying to maintain some sort of compatibility with the SNA by explaining (and in some measure justifying) numerous differences.

3. Another important reason for making maximum use of the SNA is that it seems certain to be in use internationally for a long time. For the next 10 to 20 years it will presumably be influencing both statistical developments and statistical discussions nationally and internationally. It seems likely also to affect academic discussions and the developing techniques of using national accounting and other statistics for forecasting and planning purposes.

International Reporting

4. This is an immediate problem because the Bureau will shortly be asked to report 1968 and 1969 data on both the old and the new SNA bases.

The new questionnaire is currently being examined but it is too early to say how much of it the Bureau will be able to complete. For the present, there is no alternative to adapting the existing data to the new system. This is, of course, what the Bureau has been doing in the past with the old system since the national system was never fully in accord with the old SNA. Where Australian data cannot be completely fitted into the new tables, efforts will be made to obtain as close a fit as possible and any differences will be noted. To do even this thoroughly can be a laborious and time-consuming task because it is only possible to detect differences by a detailed examination of the two systems item by item. It may be easier with the new SNA, since it contains a much fuller description of concepts than the old SNA, and the questionnaire carefully sets out the differences between the old and the new. Because the Bureau expects to give considerable weight to SNA standards in planning changes, it was thought to be desirable to embark on a more detailed comparison than had been done with the old SNA. Even at this early stage, it has become apparent that there were more differences than had been suspected between the Bureau's system and the old SNA. In a few instances reporting was on the wrong basis in the past, but fortunately in some of these cases it has been found that the basis is more in keeping with the new SNA than with the old.

5. There is much in the present reporting system that the Bureau cannot complete at this stage. Although the U.N. reporting system will be taken into consideration when assessing developmental priorities, the Bureau will have to be guided primarily by its assessment of the importance of particular series and tables for Australian purposes. However, one would expect to find considerable common ground between national needs and international reporting requirements.

6. Australian reporting of data in terms of the new SNA for the next few years will depend on the availability of data in the present system. The general policy will be:

- (a) where data can be readily and reliably adapted to the SNA, this will be done;
- (b) where present data cannot be adapted but it is expected that the SNA concepts will be adopted for national purposes, new statistics will be developed progressively as resources permit to enable reporting on the new basis to be undertaken, possibly even in advance of incorporating changes into the Australian system, provided the figures prove to be reliable;
- (c) where it seems likely that the SNA will be accepted in principle but it is also unlikely that its requirements can be met in practice, the Bureau will not report any estimates that it would be unwilling to use in the Australian system; and
- (d) where it is not expected that the SNA will be adopted and it is found that estimates cannot be readily and reliably made as in

10(a) - AUSTRALIA

(a), the Bureau will not undertake special enquiries only to enable it to comply with international reporting requirements.

7. A few examples may perhaps illustrate this policy. For (a), the new SNA changes the classification of personal consumption expenditure, and the Bureau would try to adapt its reporting to this new classification, even though it differs from Australia's. Similarly, the new SNA treats hire purchase charges as being payable by the customer and there would appear to be no reason for not reporting on this basis, even though in the Bureau administrative costs of these companies are treated in the same way as those of banks, that is, as an element in final consumption attributed to financial enterprises rather than to the household sector. A continuing case will be that of capital transfers which is a concept unlikely to be accepted in the Australian system. For (b) the development of adequate accounts for non-profit institutions will involve seven new collections, and as these are developed it may be possible to adjust some of the reported figures as in (a). For (c), examples are not so easy to find, but possible cases include transfers from companies to the personal sector, stocks in transit between Australia and overseas, various insurance transfers, and depreciation of government assets. In these cases (with the possible exception of the last) the difficulties are likely to be practical rather than conceptual. Hence estimates will not be provided until they are of a standard that could be published in the national statistics. For (d), the separation of the rent of land from the rent of buildings or the separation of the sale of land from the sale of buildings provides a good example.

8. The preceding paragraphs refer specifically to the reporting of data from 1968 onwards in terms of the new SNA. There is also the transitional problem of reporting data for 1960-69 in terms of the old SNA but in a revised questionnaire. For the most part the revised questionnaire involves a re-arrangement of items but a substantial number of new dissections are also requested. At the moment, it is not possible to devote a large amount of resources to this part of the questionnaire. Existing sources will be adapted where this can be done economically and reliably, though it may be possible to be somewhat less exacting in this latter regard than is desirable for current figures. However, there will be no attempt to correct cases where the Australian data are found to have been reported incorrectly in the past but are more or less comparable with the new SNA. In the longer run, when the Australian system is changed, the Bureau would expect to take its new figures back beyond 1968 and would report the full run of years on the new SNA basis. This means taking the view that it is more important for international comparisons to provide sound figures on the new SNA for as long a period as possible than to revise those on the old SNA basis, that there would be little interest in comparisons of how the changes have affected the figures of different countries, and that temporal comparisons are better made on each country's own figures than on a formal linking of the two different systems at a particular period.

9. As mentioned earlier, there are cases where it is now recognised that Australia's national accounts were not in accord with the old SNA because of differences of principle. There were also cases where the differences arose simply because the appropriate data were not available. These latter cases include some in which the new SNA has made marginal adjustments to concepts, such as the movement of some non-profit institutions from the enterprise sector to the government sector, which have a low priority rating for investigation and adjustment. In some instances, the Bureau will have to decide for reporting purposes whether to say that the concept but not the estimate has been changed or that the Australian estimates are not comparable with one or other of the concepts.

Revising the Australian System

10. In due course a number of changes will be made in the Australian System of National Accounts. The present system has been in use for several years without changes despite some shortcomings. As has been indicated, changes in the system will be strongly influenced by the new SNA but the Bureau does not intend to make these changes in the immediate future, mainly because a number of important improvements and extensions are currently being made in the statistical data base. For the year 1968-69, a completely new system of conducting economic censuses was introduced with integration of establishment and enterprise data in six broad industry sectors. The detailed results from these censuses should be available next year but a great deal of analysis will then remain to be done to relate them to past estimates and lay down a basis for future estimates making maximum use of all the new data. Use of the census material is closely related to plans to implement parts of the SNA and for this reason some aspects of the census collection and proposed use of the statistics are discussed below. However, it is envisaged that it will be 1973, or even 1974, before it will be possible to make the revisions to, and further elaboration of, the accounts that will follow from this census material. And it is also considered that it would be best to incorporate these changes and also any changes to be made in the light of the SNA at the same time.

Sectors - Enterprises, Establishments, etc.

11. Both the old SNA and the Australian National Accounts (ANA) refer to the sectors as persons, governments, and enterprises. (The ANA distinguishes trading and financial enterprises but this is not particularly important for the present discussion.) In fact, these two systems comprise basically a national production account and a number of highly summarised sector financial accounts. The sectoring arrangement is manifestly inadequate for a complete system of accounts and, in particular, if further development in the production account area is planned, it immediately becomes obvious that interest will be in establishments more than in enterprises. Most countries (including

Anustralia), however, content themselves with showing GDP by industry of origin based on the addition of compensation of employees, classified as far as possible on the basis of establishments, and gross operating surplus, classified on the basis of enterprises.

12. The new SNA now recognises the need for different systems of classification and sectoring for the different types of accounts. Compared with a few more or less passing references to establishments in the old SNA, the new SNA makes it clear that, in principle, all the production accounts of the systems are based on a classification of establishments or establishment-type units and the income and capital finance accounts are based on a classification of enterprises or other appropriate institutional-type units.

13. This has important implications for the whole statistical system. It has been traditional to acknowledge that there are three ways - expenditure product, and income - of deriving national accounting aggregates. In practice, many countries have used the income and expenditure methods, and then in some cases have had difficulty in reconciling these with the production-based estimates of the input-output system. If the new SNA is to be implemented, it would seem that more attention will need to be given to ways and means of applying all three approaches within a single, co-ordinated system. In this respect the Bureau hopes the Integrated Economic Censuses, referred to earlier, will be of great benefit. The censuses included are mining, manufacturing, electricity and gas, wholesale and retail trade, and a number of service industries such as hotels, hair-dressers and laundries. The system provides for returns from all establishments predominantly engaged in any one of these industry sectors, and also from enterprises predominantly engaged in these sectors. For single-establishment enterprises there is one combined return, but for multi-establishment enterprises there are separate returns. The enterprise return summarises the main figures from the establishment returns and obtains separate details for out-of-scope units (all units combined). Ancillary units such as storage or administrative locations are included on establishment or enterprise returns depending on whether they serve one or more establishment. As far as possible details of operational revenue and expenses, capital expenditure, and stocks, etc. are covered in establishment returns and details of other expenses (and of some non-operational income and outlay such as interest and royalties) are included in the enterprise return. Tabulations are planned to provide estimates of GDP for enterprises in broad industry classes and also for establishments by fine industry classes. This latter tabulation necessitates the spreading of the expenses, of which details are collected for enterprises only. Estimates of these expenses will also be required for the "in-scope" establishments of "out-of-scope" enterprises. There will be problems in making appropriate estimates for establishments and enterprises in industries outside the scope of the economic censuses. However, with the census and other data the Bureau expects to go a long way towards completing SNA Table 1 (GDP and Factor Income by Kind of Economic Activity) and Table 17 (Domestic Factor Income According to Kind of Economic Activity and Institutional Origin). Without being certain how

different these two tables will be at corresponding levels of industry detail, there is nevertheless merit in trying to follow the SNA here, since it makes possible a clearer exposition of the concepts than either the old SNA or the present ANA. As to the accounts themselves, the Bureau would expect to develop progressively some of the C accounts (Industry Production Accounts). The A accounts, with their detailed analysis of the destinations of commodities and the need to reconcile commodity flows and statistics of expenditure, will be more difficult, and probably impossible, until an input-output table has been completed on the basis of the new census data.

14. In regard to fixed capital expenditure, Australia has been basing its classification by industry on ownership and not on industry of use as provided in the old SNA and therefore are not affected by the change in the SNA to industry of ownership. However, because sources of information have been based on units which are much closer to enterprises than to establishments; the Bureau has had considerable difficulty for various reasons in reconciling such establishment data as it has with data from its other sources. Here again, it is hoped to use the economic censuses to establish relationships between enterprise statistics (required for the III - 5 Accounts) and establishment statistics (required for Table 7). These problems may be considerably greater than those in respect of product estimates, mainly because of the variety of ways in which the ownership and management of buildings and other capital equipment can be organised, not the least troublesome of which is that all (or the bulk of) capital expenditure may be incurred by a related company which owns all the capital equipment. Although the basic collection unit in the censuses was the enterprise, special arrangements have been made for property-owning companies. The result aimed at is that the capital expenditure of an enterprise should be attributed as far as possible to its component establishments on the basis of use, or intended use, and in the case of property-owning companies to the establishment(s) of the related companies. Possibly the main gap in the system is in respect of capital expenditure of an operating enterprise on equipment intended for transfer or lease to a related enterprise, though even here the Bureau should have enough information to follow up any important cases.

15. It might be argued that the foregoing discussion represents a somewhat uneasy compromise between the use criterion of the old SNA and the ownership criterion of the new. Australia would argue that it will be very close to what would have been obtained had enterprise groups been used as the collection unit (as is envisaged as a possibility by the SNA) and all capital expenditure of the group distributed to the appropriate establishments. Strict comparability with the product tables, however, may necessitate some estimating to remove intra-group payments of rent.

Sectors - Non-profit Institutions

16. At present, the Bureau is following the old SNA in saying that these institutions are in the household sector. In practice, this does not

10(a) - AUSTRALIA

fully reflect the real situation. For example, in large measure payments by persons to non-profit institutions are analysed rather than the expenditure of these institutions; to the extent that governments make payments to such institutions, these are reflected in government expenditure; and the wages and salaries paid by these institutions are included in a table entitled "Private Enterprise Wages and Salaries Paid, by Industry" since there is no provision in the system for wages paid by the household sector.

17. Because it clearly separates production accounts from income and capital finance accounts, the new SNA is much clearer than the old in its treatment of these institutions. The Bureau expects to follow broadly a similar conceptual framework and would hope to develop progressively accounts for non-profit institutions. These accounts may have to be pieced together from a variety of indirect sources, but there are already some direct collections established for hospitals and universities, and the time might be right to consider further collections from private schools. Before accounts for non-profit institutions can be developed, however, certain difficulties over the formulation of concepts in the SNA and the appropriate interpretation of them will have to be resolved.

18. The SNA defines non-profit institutions by reference to whether their charges for services cover their costs. Elsewhere it refers to the possibility that such institutions may be instruments of government social policy and in such cases puts them in the government sector. Australia has difficulties with both aspects of this definition. In Australian usage the term non-profit institutions is normally taken to include an institution in which no-one is entitled to a distribution of profits rather than only those institutions which operate on a strictly non-profit-making basis. Some private schools rely on gifts to help meet running costs, and this clearly meets the SNA definition, but some charge fees sufficient to cover running costs and only rely on gifts for capital purposes, and the like. It is difficult to know where to draw the line and the Bureau would not be disposed to divide schools among all three sectors viz. (i) producers of government services, (ii) producers of private non-profit services to households and, (iii) industries. Difficulty also arises over institutions like public hospitals, medical and hospital insurance funds, and universities. These are all financed in substantial measure by the government and, largely because of this, to some extent controlled. They are closely bound up with government social policies. Again it is difficult to draw a line between the non-profit and the government sectors. It is perhaps worth noting here that this issue may be affected by the need to link production and employment statistics: these are institutions whose employees would not be commonly regarded as government employees even though the institutions apparently meet the financial and social policies criteria in the SNA.

Sectors - Government

19. For some years the Bureau has been working on a substantial elaboration of the public sector statistics which are fairly closely linked with the national accounts. Further development is needed to eliminate some confused treatment in the national accounts, but the ways in which these statistics have been developing are not fully consistent with the new SNA.

20. The present analysis isolates any trading-type activities (i.e. production of goods or services for sale to the public) and classifies them as public enterprises. The operating surplus of such "enterprises" is shown in the trading enterprises income appropriation account but is then passed to the government account and all other financial and capital transactions are recorded in the government sector. But the present analysis deals with activities, which may or may not be co-extensive with those of establishments, and apart from ambivalence about whether these activities are in the enterprise sector or the government sector, it is unsatisfactory to classify activities as enterprises (though it must be added that the old SNA is almost equally confused on the subject of public enterprises and applies terms to them which are properly applicable only to establishments).

21. Subject to the following comments, the Bureau would expect to work out something more acceptable on the general lines of the new SNA, and to indicate more clearly which government establishments are included in the industries sector, which enterprise-type units are treated as part of general government, and which as corporate and quasi-corporate enterprises. A main concern is that the application of the establishment concept to the purpose classification in the SNA is somewhat unsatisfactory, and it is believed that the government sector as a whole is one where conventional ideas of an establishment are very difficult, if not impossible, to apply. Certainly more study of the problem is needed. It is recognised that an establishment of some sort must be found for purposes of industry classification, but Australian public sector statistics have been developed more as a classification of accounting units on the basis of the nature of their transactions than as a classification of physical establishments. For the purposes classification, and perhaps also for the industry classification, the Bureau will for the present have to adapt the present analysis as it is being developed. For the purposes classification, it is believed that this is compatible with the rest of the SNA, and for many uses it may be more appropriate than a strict establishment approach. It also seems less objectionable to refer to an accounting unit as an establishment-type unit than to refer to it (as at present) as an enterprise, and such accounting units may give tolerable results for industry classification even if they fall short of providing a full range of establishment data such as employment and capital expenditure.

Rent, Interest, and Financial Institutions

22. The old SNA treated net rent, interest and dividends as transfers and Australia has adopted the same principle. This decision seems to have been adopted originally because income statistics for enterprises were used at the time and the Bureau took the view that a firm's contribution to gross domestic product should be invariant to its method of financing, whether by equity capital, borrowed capital or by renting of capital equipment. This treatment was, perhaps, reasonable enough in relation to enterprises, but in relation to both government and household sectors it had some undesirable features which might have been aggravated if the logic had been applied more rigorously. For example, logic suggests that net rent paid by persons or governments should have been classified as a transfer: instead the gross rent paid by those sectors was treated as a purchase and gross rent and operating surplus were imputed to owner-occupied dwellings and (in the old SNA) to buildings owned and occupied by governments. In practice, these concepts are extremely difficult to apply to rent payments by enterprises since information has to be collected in terms of gross rent payments which then have to be converted to net payments.

23. In the new SNA, contribution to gross domestic product is looked at from the point of view of establishments rather than of enterprises. The logic of treating interest as a transfer is reinforced by the practical impossibility of attributing interest to the component establishments of enterprises. There is no comparable problem in measuring gross rent paid at the establishment level. There is still a case for maintaining that the product of an establishment should not depend on whether it rents its capital equipment. But if this case is accepted, there are still intractable conceptual and practical problems of splitting the gross rent into its two components (purchases of goods and services and a transfer) and of deciding which industry to credit with the sales of the goods and services, and any factor incomes included therein.

24. There are no ideal solutions to these problems. Except for the distinction between rent of land and other rent, the Bureau will probably change to the new SNA treatment of rent. There are, however, two further problems. The SNA treatment is not really satisfactory for input-output purposes since input co-efficients will be affected by changes in the proportions of owned and rented equipment. Secondly, there is the matter of locating rent receipts by industry or, which amounts to the same thing, identifying the appropriate establishment units which receive rent. In practice, it seems likely that the real estate industry will consist of relatively specialised firms only and much (perhaps most) of the ownership will be a secondary activity in a great variety of other industries. An especially troublesome case could be the life insurance industry: to the extent that it manages the properties it owns and lets, the cost of services should be deducted from the rent received, but in practice it might be hard to prevent it being included, along with administrative expenses, in the imputed charge to the household sector.

25. Closely linked with the rent and interest problem is that of financial institutions. A consistent treatment of rent and interest as transfers results inevitably in negative operating surplus in these institutions. The old SNA principle of imputing a charge over and above the interest received did not appeal to the Bureau and a different solution to this problem was adopted. This was to treat the "excess" expenses of financial enterprises as a component of national expenditure, a form of collective consumption financed out of their interest receipts more or less comparable with other collective consumption financed by the government out of taxation receipts, etc. This seemed an irritant to many of the users of these estimates, partly no doubt because it was not fully understood, though it is perhaps questionable whether the international recommendation would have been better understood or accepted. Since these institutions by and large are operating profitably in a commercial context it seems obvious that their charges are concealed in their existing receipts and should not be imputed in addition. This is roughly the implication of the new SNA treatment. This suggests charging the costs of sales and personal finance companies to the borrowers (mainly households, but the problem of allocating part to industries or capital finance accounts is not discussed), the administrative costs of life insurance and pension funds to households, and the imputed service charge of banks, etc. to a dummy industry so that, without any additional interest imputation, operating surpluses in total are reduced by the operating loss attributed to the dummy industry.

26. It seems likely that the Bureau will adopt this treatment in preference to the present treatment, despite some logical shortcomings even in the new proposal. There are also one or two other minor problems still not answered in SNA such as what to do with costs of unit trusts, investment companies, etc. Probably these should be charged to households, thus necessitating a corresponding addition to the property income paid out by these institutions.

Classifications

27. There is a great variety of classifications in the SNA. Some of these are so closely interwoven with definitions of items that they become to all intents and purposes an integral part of the system. Others are much more peripheral and could be altered substantially without greatly affecting the structure of the system. The classification of final consumption of households, the classification of government purposes, and even in large measure classification by kind of economic activity all fall into this second category. Some of these classifications are at variance with those used in a substantial body of Australian statistics, frequently covering an extended period of years, and it would be impracticable, or at least unwise, to adopt them in the national accounts. In such cases the Bureau will use the national system (i.e. the Australian Standard Industrial Classification and our class-

ification of government expenditure by function) and make the best arrangements practicable for converting to SNA for reporting purposes. In other cases, such as the classification of consumption expenditure of households, the Bureau will probably be able to adopt the SNA classification in principle since the impact on other national statistics is relatively small or some of the developments are at the stage where they can still be readily influenced by the national accounting requirements. This is not to say, however, that there will be no difficulties in practice in aligning Australian statistical sources with the classifications. This is a perennial problem with an inevitable feedback into actual classifications. It is hoped, however, that such variations will prove to be of minor importance.

28. There are elements in some of the proposed SNA classifications about which the Bureau has reservations of principle. Some have already been mentioned in relation to the classification of units between industries, government services, and services of non-profit producers, and in relation to the attempt to distinguish land from buildings. Another case is the distinction between current and capital transfers. Forcing transfers into the capital group if one party so regards them (a principle which seems to be abandoned in the SNA, without any discussion, in the case of insurance claims), must have a distorting effect on the accounts of the other party. The Bureau disagrees with the omission of estate duties from the government income and outlay account when governments clearly regard these as current revenue. It is also reluctant to classify as capital a number of regularly recurring grants by governments when they do not result in the acquisition of any capital asset by the government. A solution may be to allow payments to be recorded as current but receipts to be recorded as capital if they are clearly to finance capital outlay. This will necessitate the inclusion of a reconciliation item in the national capital account to balance the discrepancies between receipts and payments of both current and capital transfers and to record the element of national saving which has not been attributed to any individual sector. Another case that appears to be illogical, is making item 6.7 (unfunded employee welfare contributions imputed) a current transfer back from households to the employer: where the Bureau is able to make such an imputation in the first place it will be treated as borrowing by the employer from the household sector.

Input-output - commodity accounts, etc.

29. Australia's plans in this area are still general and fairly flexible. Up to date, only an extremely small establishment has been approved for input-output and progress on tables for 1962-63 has been slow but work should be completed this year. For the future, it is hoped to compile input-output tables in the years when Integrated Economic Censuses are undertaken, which is expected to be about every five years, beginning with 1968-69. The commodity tables of the system (Tables 2 and 3) should develop out of this input-output work. The annual census of manufacturing will provide much of

the basic data for these tables (in particular Table 2). Whether such tables are prepared on a partial basis annually, and if so, in what detail, will depend on the Bureau's view of the demand for them and the availability of the resources required to complete them from the tabulated basic data.

Capital Finance Accounts

30. Bureau resources have not permitted, as yet, development of these accounts in any overall and systematic way. Flow-of-funds accounts have been prepared for some years in the Reserve Bank of Australia, on a basis broadly consistent with the Australian National Accounts. Changes in the basis of sectoring in the national accounts may affect the Reserve Bank's rather more detailed sectoring and close consultation will be desirable.

PROGRESS AND PLANS IN DEALING WITH THE
REVISED U.N. SYSTEM OF NATIONAL ACCOUNTS

I. Introduction

1. An integrated system of National Accounts is a necessary tool for an integrated approach to the collection and assembly of economic data. The statistical end products have numerous intermediate and ultimate analytical uses, whether these be for micro studies, highly disaggregated macro studies, or macro analysis based on aggregated data. The system of accounts provides a conceptual and data base for price systems and analysis, input-output analysis, productivity studies, etc. It also makes feasible research on highly disaggregated models of the economy involving transactions in goods and services, transfers and financial instruments, as well as saving, lending, stocks of real assets, and levels of financial claims.

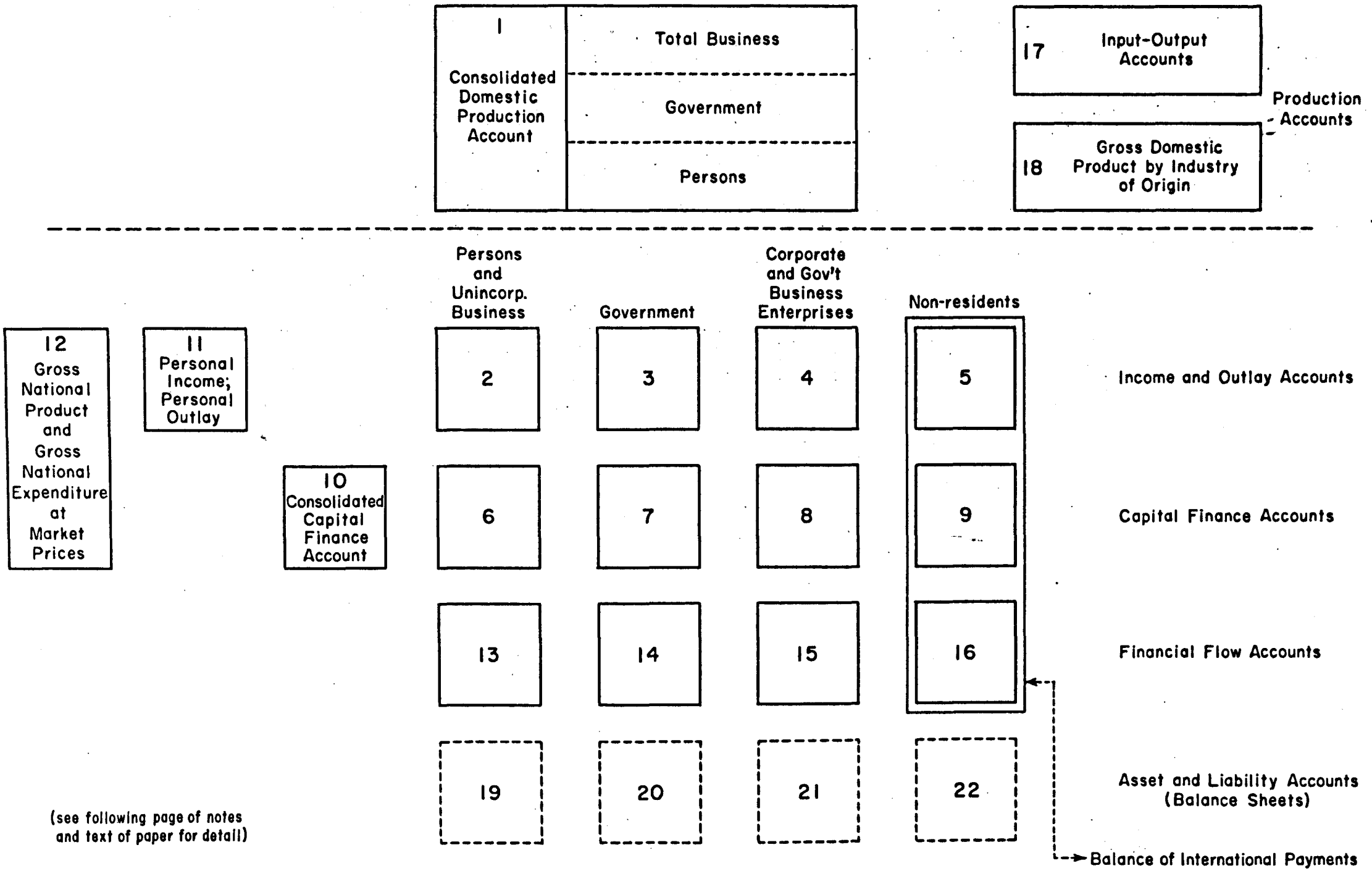
2. The Dominion Bureau of Statistics has just completed a major phase in the development and revision of Canada's system of National Accounts. In 1968, DBS published an historical revision of quarterly and annual Indexes of Real Domestic Product by Industry. In 1969, quarterly and annual Tables of Financial Flows for 1962-1967 were published, and regular quarterly publications of these data now appear. Also in 1969, we published a two volume set of 1961 Input-Output Accounts and analytical tables. By the summer of 1970, DBS will have published an historical revision of the annual National Income and Expenditure Accounts (from 1926) and of the quarterly Accounts (from 1947). We continue, of course, to publish regular quarterly and annual data on the Balance of International Payments.

3. The integrated nature of these accounting systems was recognized from their inception, and given a formal label in 1969. Each of the regular or occasional publications in these areas (containing data, concepts and methods) will now carry a general title, "System of National Accounts". (In Canada, the System of National Accounts includes the Balance of International Payments.)

4. The importance of an integrated set of asset and liability statistics for modern economic analysis is well recognized, and data development for stocks of fixed reproducible capital and for financial assets and liabilities is progressing. These developments are summarized in Section III.

5. A schematic outline of the framework of the Canadian System of National Accounts is given in Diagram 1.

DIAGRAM I. Framework of the Canadian System of National Accounts



(see following page of notes and text of paper for detail)

NOTES TO DIAGRAM ISub-Systems of the Canadian System of National Accounts

6. The annual Income and Expenditure Accounts include accounts and tables #1 - 12; accounts #1 - 10 comprise the annual "sector accounts" of the sub-system. The "main supporting tables" for persons (#11) are described in the last paragraph of Section II of this paper. The "main summary tables" for Gross National Product at Market Prices and Gross National Expenditure at Market Prices (#12) can be viewed as consolidations of income, non-factor primary cost, and final expenditure (current consumption and capital formation) entries in the income and outlay accounts and the capital finance accounts. In the product (or income) table, a sub-total is presented for Net National Income at Factor Cost; this excludes indirect taxes, capital consumption allowances and miscellaneous valuation adjustments, and one-half the residual error of estimate (these three items being part of Gross National Product at Market Prices).

7. The quarterly Income and Expenditure Accounts include #12, #11 and revenue and expenditure tables for government whose content is equivalent to #3 and #7.

8. In the quarterly and annual Accounts, Gross National Expenditure at Market Prices and its components are presented in current and in constant (1961) dollars.

9. In the U.N. SNA, the Capital Finance Accounts include financial flows. In the Canadian system, the balances of the capital finance accounts (#6 - 10) are for net lending or net borrowing - the Financial Flow Accounts (#13 - 16) are presented as a separate sub-system.

10. The following are dealt with in the text of the paper:

<u>Table or Account numbers</u>	<u>Sub-system</u>	<u>Comment</u>
17	Input-Output Accounts	
18	Measures of current and constant dollar Gross Domestic Product by Industry	
(see comment)	Balance of International Payments	Conceptually equivalent to 5 + 9 + 16
19 - 22	Asset and Liability Accounts	Under development

11. Section III of this paper will deal with each of the sub-systems; the extent of the implementation of U.N. recommendations is described in general terms, selected similarities and differences between the U.N. and Canadian systems are pointed out, and important areas of development in Canada are outlined. Section II which follows deals with a few of the broader differences and points out one important problem area that appears to exist in both systems.

II. Major Problem Areas

The General Approach in Canada

12. In the development and integration of the sub-systems of the Canadian Accounts, we have attempted to come as close to the U.N. SNA as our statistical source material, analytical requirements and Canadian customs and usage permit. In the deliberations which led up to the final formulation of the U.N. SNA, we always took the view that countries should structure the conceptual, definitional and classificatory aspect of their accounts such that international comparability can be obtained without too great a cost to the international respondents. However, we have also maintained that the institutional requirements, the social milieu and the past investment of the users of the accounts may necessitate different structural and presentational forms in different countries. Some problems which formerly arose from such differences are now removed by the use of computer automation which facilitates the storage and manipulation of masses of detail, including its rearrangement into a variety of formats.

The Institutional Sectors and their Linkage to the Dissaggregated Production Accounts

13. In recent months, planning has begun for a highly disaggregated medium term simulation model of the Canadian economy. Work on this model will be carried out by several federal government departments and agencies, including the Dominion Bureau of Statistics. The experience gained to this point leads us to believe that in one important area, the U.N. SNA and the existing Canadian system have relatively inflexible structures which do not take full cognizance of the extent of data development and the possibilities of modern coding, programming and computing technologies in advanced countries. For modern analytical requirements, the level of detail for the institutional sectors is not adequate in the Canadian and U.N. systems, particularly in the area of non-financial industries. However, such additional detail would likely have limited analytical possibilities if it were not linked in a detailed way to the production accounts. The aggregation of industry value added into a single account from which institutional sector incomes are distributed (as in the U.N. SNA) appears to be unsatisfactory, even for the existing disaggregation of institutional sectors. It follows that the saving, capital formation, net lending, financial flows and assets and liabilities of the institutional sectors are not adequately interrelated with the establishment-based production accounts.

14. For the Canadian system, we attach considerable importance to the formulation of more detailed and systematic links in this crucial area. For example, it appears feasible in Canada to construct input-output accounts (and to derive analytical tables based on input-output data and models) using establishment-based statistics for the estimation of industry and commodity

production, commodity absorption (including imports), incomes generated, manpower requirements, etc. on an integrated national and regional (or partially regionalized) basis and (through the use of "classification conversion matrices" which are based on a central list of companies and their establishments) to regroup establishment-based industry data into fairly detailed company-based industries for describing: the distribution of profits in the forms of retained earnings, interest, dividends, profit taxes, etc.; transfers; saving; capital formation; net lending; financial flows; and assets and liabilities. Research into the modeling of these areas in an integrated manner with the production statistics of the SNA depends upon this type of data development. Of course, it would not be possible to estimate certain important income components (particularly interest and dividends) in full intersectoral detail, nor would such estimates necessarily be useful in analysis. These transactions could be pooled at a more aggregated sectoral level for redistribution.

The "Standard Accounts" of the U.N. and Canadian Systems

15. The Group II Accounts of the U.N. Standard Accounts are those for Production, Consumption Expenditure, and Capital Formation. These tables provide an aggregation and rearrangement of the type of information contained in the input-output accounts. It now appears to us that the typical user of these accounts would find them unnecessarily detailed and possibly confusing with respect to industries and commodities. Yet with all this detail, he could not adequately relate them to the Standard Accounts for the five institutional sectors (Group III Accounts - Income and Outlay Accounts; Capital Finance Accounts).

16. The Group II and Group III Accounts should furnish a set of useful analytical tools at a level of aggregation between the fully consolidated accounts (Group I Accounts in the U.N. Standard Accounts), on the one hand, and the fully disaggregated set of accounts on the other. They should be relatively simple and straight forward in structure and presentation; at the same time they should furnish certain linkages required by the user.

17. The U.N. Group II Accounts would be much more comprehensible to the user if they were presented as one full matrix containing the aggregated set of input-output accounts, including the requisite detail on secondary commodities produced by each industry group, as well as detail on final expenditure on commodities. However, this would not solve the basic problems which seem to be:

- i) the user cannot directly relate consumption expenditure and capital expenditure on commodities to the producing activities;
- ii) he cannot directly relate the components of gross domestic product by activity in the Group II Accounts to the offsetting entries in the institutional sectors of the Group III Accounts (this is the same problem as that outlined above for the fully disaggregated accounts);
- iii) in any case, it is questionable whether the activity-based industrial detail in the Group II Accounts is a desirable feature in a set of Standard Accounts - perhaps this should be relegated to the Supporting and Supplementary Tables.

18. The U.N. Group II and Group III Standard Accounts are comparable to the Canadian annual "sector accounts" of the income and expenditure accounts, plus our financial flow accounts. Instead of the U.N. Group II Accounts for

activities and commodities, we employ a "consolidated" domestic product account; at present, this account can be fully deconsolidated into total business, government, and persons (see Diagram 1). Although highly aggregated, this production account is an "inter-industry flow" table in which the intermediate "business-to-business" transactions do not appear. Thus there is no problem of linking expenditure on commodities to the producing activities in the Canadian "standard accounts"(1). In addition, the component detail for primary costs of production and for final expenditures are further detailed so as to permit an articulated set of cross entries to the income and outlay accounts and the capital finance accounts of the four institutional sectors:

- i) persons (including private non-profit institutions) and unincorporated business
- ii) government
- iii) corporate and government business enterprises
- iv) non-residents

19. In the U.N. institutional sectoring, non-financial "quasi-corporate" enterprises are combined with non-financial corporate enterprises, while other unincorporated enterprises are included with households. Canada has few such quasi-corporate enterprises in terms of the U.N. criteria as set out in para. 5.52 of the U.N. SNA document. Thus in the Canadian system, virtually all unincorporated businesses are included in a sector with persons. However, the full system of income and expenditure accounts contains two supplementary main tables for persons only:

- i) a table showing "Sources of Personal Income",
- ii) a table showing "The Relation between Net National Income at Factor Cost, Personal Income, Personal Disposable Income, and Personal Saving", in which the detail on the outlay of persons is presented.

Personal income is defined to include the total "net income of non-farm unincorporated business" and the "net income received by farm operators from farm production". Comparable estimates for persons were part of our previous accounts and are considered to be particularly useful in analysis. In addition, the personal income statistics are required for the calculation of equalization payments under taxation agreements between the federal and provincial governments.

III. Comments on the Various Sub-systems

Income and Expenditure Accounts

20. Many aspects of the structure and presentation of these accounts have been dealt with above. From the start, our basic concepts were never very different from those of the original U.N. system and in our recently completed historical revision, they have been brought into closer line with the new U.N. SNA. For instance, our treatment of withholding taxes on income accruing to Canada has been changed, the net imputed rent on government owned fixed assets has been dropped, certain items in consumer expenditure have been brought into closer conformity with the U.N., surpluses of liquor commissions (government monopolies in Canada) are now treated as indirect taxes rather than profits

(1) Of course the problem reappears in a different form - that of relating the consolidated production account to the input-output accounts.

of government business enterprises, etc. However, we felt strongly enough about the established treatment of imputed banking services to retain the existing Canadian (and old U.N.) methods in this area. In the institutional sector accounts, we have not drawn the distinction between financial and non-financial corporations (although the financial flow accounts have made this distinction): We hope to begin work in the near future on a separate private non-profit institutions sector and on an improved data base for unincorporated businesses.

21. It might be noted that for historical reasons, we have continued to feature the "national" concept in our summary main tables (see Diagram I).

Input-Output Accounts

22. The Canadian 1961 input-output accounts are considerably more detailed than is recommended by the U.N. - 187 industries and 644 commodities (goods and services) with the largest published aggregation being 110 industries by 197 commodities. As indicated, the Canadian accounts and analytical tables are rectangular (more commodities than industries) rather than square as in the U.N. SNA. In our published tables, the valuation of commodities (in terms of U.N. SNA terminology) is at approximate basic values, rather than at producers' values as in the U.N. system. In general, we can rearrange our basic data in order to present them in a format which is close to the U.N. system.

23. The 1961 input-output accounts are undergoing a few revisions to integrate them with the new income and expenditure accounts; estimates of final expenditure vectors by detailed category or purpose of expenditure and by commodity are being completed. We are now in the process of updating these tables in current and constant dollars for the years 1962-1967, using current period information where it is readily available.

Gross Domestic Product by Industry of Origin

(a) (Constant Price) Real Domestic Product by Industry

24. In June, 1970, DBS began publication of monthly Real Domestic Product by industry. These measures, as well as the regularly published quarterly and annual measures, include all establishments in the economy classified by industry in the DBS Standard Industrial Classification, regardless of sector; i.e. the Canadian "industry" includes the relevant U.N. "industry" concept, plus establishments (activity units) in private non-profit institutions, in general government, and in households. However, at the total economy level, DBS currently makes a rough split between "Commercial" and "Non-Commercial" by assigning entire SIC industries to one group or the other on the basis of predominance. For example, water systems and other utilities (SIC 576 and 579) although containing commercial establishments, are classified in total to non-commercial. Research is planned to determine the possibility of fragmenting SIC industries following the U.N. sectoring system. In the area of estimating constant dollar outputs and intermediate inputs, DBS is shifting to a greater use of current price values deflated by price indexes and away from the use of physical quantity data. Further detail relating to GDP in current and constant prices by industry of origin is contained in a separate paper to this Conference, "Production Statistics and Indicators for Indices of Production".

(b) Current Dollar Gross Domestic Product by Industry

25. DBS also has a programme for developing annual estimates of current price GDP by Industry. At present, research is underway in an attempt to derive current price GDP for the years 1966 and 1967 in a manner that is consistent with the constant price data using methodology developed in large part for purposes of the 1961 industrial distribution of GDP and the input-output accounts.

Financial Flow Accounts

26. In general, the Canadian system of financial flows follows the U.N. recommendations. In full detail there are 31 sub-sectors including "rest of the world". Two differences in sectoring are worth noting:

- i) In the U.N. system, the Exchange Fund is included with general government. In Canada, the operations and policies of the Fund are so closely linked with those of the central bank that the Exchange Fund is included in the Monetary Authorities sector (this treatment was the earlier recommendation of the U.N.);
- ii) in contrast to the U.N., Canada has set up separate sectors for federal government financial institutions, and for provincial government financial institutions. In Canada, these organizations operate primarily to serve private enterprise (e.g. Central Mortgage and Housing Corporation, provincial industrial development corporations) and have few counterparts in the private sector.

27. Most of the estimates in the financial flow accounts are based on quarter-to-quarter changes in balance sheet totals for financial assets and liabilities, the statistics for which come largely from direct quarterly questionnaires and from data filed for administrative purposes. Emphasis has been placed on the type of instrument; we do not yet have good data on short-term vs. long-term instruments, particularly in the areas of bonds and term deposits. Within the next year or two, the data now being collected will be reviewed and improvements gradually introduced.

Balance of International Payments

28. In the Canadian accounts for institutional sectors, transactions with the rest of the world are not treated as a separate type of account as in the U.N. SNA. Rather, non-residents are treated as a separate sector. Thus the Canadian Balance of International Payments serves, in part, to bring together the transactions with the rest of the world in one set of accounts.

Asset and Liability Accounts (Balance Sheets)(a) Levels of Financial Claims

29. Professor Revell of Cambridge University has prepared for the U.N. a memorandum on financial assets and liabilities; this will serve as a basis for international discussion. A final draft on the subject is expected in late 1972. In Canada, we have taken the position that estimates of balance sheet entries for financial claims should be prepared, but have not yet set a timetable for project. We shall, in general, follow the concepts outlined by the U.N.

PROGRESS AND PLANS IN DEALING WITH THE
REVISED SYSTEM OF NATIONAL ACCOUNTS

1. The fifteenth session of the Statistical Commission adopted the revised System of National Accounts (SNA) for the use of national statistical authorities and in the international reporting of comparable national accounting data. This system provides a comprehensive and detailed framework for the systematic and integrated recording of the flows and stocks of an economy. It brings together data ranging in degree of aggregation from the consolidated accounts of the nation of the old SNA to detailed input-output and flow-of-funds tables into an articulated, coherent system. It has been recognised that it will take a number of countries many years before they can compile all the data of the full system which are needed for the purpose of economic and social analysis. However, in the case of the developing countries, the chapter of the revised SNA which is devoted to discussing the ways in which they might adapt the full system to their own requirements and circumstances, includes a suggested order of priorities in which they might develop the standard accounts and tables of the system.

2. The sequence in which the developing countries might start the compilation of the standard accounts and supporting and supplementary tables of the full system of national accounts is set out in annex 9.2 of the revised SNA. The suggested list of priorities is the result of weighing the difficulties of gathering the basic series needed to estimate national accounting data against the urgency with which the national accounting series are required. The highest priority has been assigned to the compilation of accounts and tables which are urgently needed and which may be prepared from basic data that are, for the most part, already available in the case of a number of developing countries. Priority 1 covers the following accounts:

Standard Accounts

3.

- (i) Consolidated accounts for the nation
- (ii) Final consumption expenditure of government services
- (iii) Final consumption expenditure of private non-profit services to households
- (iv) Final consumption expenditure of households
- (v) Separate production accounts for the following industries:

Agriculture, hunting, forestry and fishing;
mining and quarrying; manufacturing and
electricity, gas and water; construction;
wholesale and retail trade; transport,
storage and communication

- (vi) production account of producers of government services
- (vii) Income and outlay account of non-financial enterprises, corporate and quasi-corporate
- (viii) Income and outlay account of general government
- (ix) Capital finance account of general government, gross accumulation part

Standard Tables

4.

- (i) Supply and disposition of commodities
- (ii) Composition of gross capital formation by type of capital goods
- (iii) National income at market prices
- (iv) External transactions

Consolidated Accounts

5. The consolidated accounts for the nation include the following accounts:-

- (i) Account 1: gross domestic product and expenditure
- (ii) Account 3: national disposable income and its appropriation
- (iii) Account 5: capital finance, and
- (iv) Account 6: All accounts - external transactions.

The account 1 contains the following items of gross domestic product:

- (1) Compensation of employees (2) Operating surplus (3) Consumption of fixed capital (4) Indirect taxes, less (5) Subsidies.

In India, separate information is not compiled for the first two items. It would be possible to compile separate estimates in respect of these items only on an ad-hoc basis. The various

items shown in the expenditure on the gross domestic product include:

- (1) Government final consumption expenditure (2) Private final consumption expenditure (3) Increase in stocks (4) Gross fixed capital formation (5) Exports of goods and services less (6) Imports of goods and services.

Preliminary estimates of gross fixed capital formation and increase in stocks for the years 1960-61 to 1965-66 have been brought out by the Central Statistical Organisation in November, 1969. The estimates of private final consumption expenditure in Account 1 can be obtained only as a residual.

6. The appropriation of disposable income in Account 3 contains the following items:

- (1) Government final consumption expenditure (2) Private final consumption expenditure (3) Saving.

The preliminary estimates of saving for the period 1960-61 to 1965-66 also have been brought out by the Central Statistical Organisation in November, 1969. It should, therefore, be possible to compile the information for these three items provided, the information in respect of "private final consumption expenditure" is obtained as a residual from Account 1. The disposable income contains the following items:

- (1) Compensation of employees (2) Compensation of employees from the rest of the world, net (3) Operating surplus (4) Property and entrepreneurial income from the rest of the world, net (5) Indirect taxes less (6) Subsidies (7) Other current transfers from the rest of the world, net.

Here again separate information is not available in respect of the following items:-

- (i) Compensation of employees
(ii) Operating surplus

7. The capital finance account contains (i) gross accumulation and (ii) finance of gross accumulation. The items of gross accumulation include (1) increase in stock; (2) gross fixed capital formation (3) purchases of intangible assets n.e.c. from the rest of the world, net and (4) net lending to the rest of the world. The information in respect of the first two items is available in the preliminary estimates of capital formation in India, 1960-61 to 1965-66 brought out by the Central Statistical Organisation. Information in respect of the item "Purchases of intangible assets n.e.c. from the rest of the world, net" is not available. However, information in

respect of the item "Net lending to the rest of the world" is available in table No.9: external transactions account of the publication entitled "Estimates of national product (Revised Series)", March, 1969 issued by the Central Statistical Organisation. The items in the finance of gross accumulation include (i) saving (ii) consumption of fixed capital and (iii) capital transfers from the rest of the world, net. Preliminary estimates of saving are available for the period 1960-61 to 1965-66. The estimates of consumption of fixed capital are also available for the same period in the studies by the Central Statistical Organisation. However, estimates of the item "Capital transfers from the rest of the world, net" are not available. The capital finance account in the revised SNA also makes a distinction between (i) net acquisition of financial assets and (ii) net incurrence of liabilities. The capital account in the balance of payments, however, provides separate information for credit and debit items in respect of (1) private (non-banking), (i) long term and (ii) short-term, (2) banking (excluding Reserve Bank of India) and (3) official (including Reserve Bank of India). The Reserve Bank of India have been requested to explore the possibility of making available the details of the three items viz., (1) capital transfers from the rest of the world, net (2) net acquisitions of foreign financial assets and (3) net incurrence of foreign liabilities.

8. The Account 6: "all accounts - external transactions" contains the following items in the current receipts:

- (i) exports of goods and services
- (ii) compensation of employees from the rest of the world
- (iii) property and entrepreneurial income from the rest of the world and
- (iv) other current transfers from the rest of the world.

The disposal of current receipts includes (i) imports of goods and services (ii) compensation of employees to the rest of the world (iii) property and entrepreneurial income to the rest of the world (iv) other current transfers to the rest of the world and (v) surplus of the nation on current transactions. The external transactions account given in table 9 of the Central Statistical Organisation publication does not make a distinction between (i) compensation of employees and (ii) property and entrepreneurial income. A distinction, however, has been made between "investment income" and "other factor services". The other factor services include technical fees, royalties, rentals, technicians and other professional service charges to the extent these could be identified particularly in respect of the period from 1963-64 onwards for payments and from January

onward for receipts. The net diplomatic expenditure is included in the item "government not included elsewhere" in the balance of payments, but separate information about the net compensation of local and foreign employees is not available. Further the net investment income figures relate only to remittances and exclude the undistributed profits accruing to foreigners. However, data in respect of retained earnings of the branches of foreign companies have been obtained from the Reserve Bank of India for the purpose of compiling national accounts.

9. The necessary information in respect of the final consumption expenditure of government services is available. However, separate information in respect of the final consumption expenditure of (i) private non-profit services to households and (ii) households is not available. The aggregate estimates of private final consumption expenditure can be obtained in the case of India as a residual in Account 1. The figures of final consumption expenditure of households can be built up from the National Sample Survey data. An idea of final consumption expenditure of private non-profit institutions serving households can thus be obtained by making use of these two aggregates. However, it will take some time before independent estimates in respect of final consumption expenditure are available in respect of private non-profit institutions serving households.

10. Priority 1 has also been assigned to separate production accounts for certain industries in the case of developing countries. In the revised S.N.A., a distinction has been made between (i) characteristic products of industrial activity and (ii) other products, of the gross output. The items of gross input include: (i) intermediate consumption (ii) compensation of employees (iii) operating surplus (iv) consumption of fixed capital (v) indirect taxes less (vi) subsidies. The operating account of departmental enterprises in the Central Statistical Organisation publication on national product, however, shows two items i.e. (i) sales and (ii) subsidies on the receipts side and the following five items on the expenditure side:-

- (i) purchase of commodities and services
- (ii) compensation of employees
- (iii) interest
- (iv) allowance for depreciation and
- (v) operating surplus.

For the purpose of compiling the production account, it would, therefore, be necessary to compile separate estimates of the two items of the gross input, viz., (i) indirect taxes and (ii) intermediate consumption. The expenditure on interest and rent

which is at present included on the expenditure side of the operating account of departmental enterprises would have to be taken out and included in the income and outlay account of non-financial quasi-corporate enterprises. Studies are also in progress in the Central Statistical Organisation to compile the operating accounts of non-departmental enterprises from their annual accounts. There are, however, certain differences in the scope of various items as compared to the items contained in the revised S.N.A. For instance the available information is inadequate to make a distinction between characteristic products of industrial activity and other products of gross output. Further, the expenditure on fines and penalties, interest, rent and net casualty insurance premiums have been included in the operating account whereas these items should have been included in the income and outlay account of non-financial enterprises, corporate and quasi-corporate according to the revised SNA. It is, however, proposed to explore the possibility of compiling the production accounts of departmental and non-departmental enterprises and other organised industries in the first instance according to the concepts recommended in the revised S.N.A. Production accounts for some of the other sectors recommended in the revised S.N.A. could be prepared only periodically and on an ad-hoc basis.

11. The production account of producers of government services includes the following items of gross input:-

- (i) intermediate consumption
- (ii) compensation of employees
- (iii) consumption of fixed capital, and
- (iv) indirect taxes.

The gross output includes (i) services produced for own use (ii) non-commodity sales and (iii) commodities produced. Information in respect of the consumption of fixed capital of the producers of government services is not available. It is proposed to undertake further studies in the Central Statistical Organisation for compiling the estimates in respect of this item. It is also proposed to explore the possibility of making a distinction between the three items of gross output.

12. It would be difficult to compile the accounts of private non-financial quasi-corporate enterprises due to lack of basic data. However, it is proposed to prepare the income and outlay accounts and capital finance accounts of non-financial corporate enterprises and financial institutions to the extent feasible. The income and outlay and capital finance accounts (gross accumulation part) in respect of general government are available in the Central Statistical Organisation

publication on national product. There are, however, certain differences in the classification of various items of disbursements, receipts, gross accumulation and finance of gross accumulation. Separate information in respect of the following items on the disbursements side is not available in the income and outlay account of public authorities contained in the Central Statistical Organisation publication:-

- (i) rent
- (ii) net casualty insurance premiums
- (iii) social security benefits
- (iv) social assistance grants
- (v) current transfers to private non-profit institutions serving households
- (vi) unfunded employee welfare benefits
- (vii) current transfers n.e.c. to residents.

Similarly, on the receipts side of the income and outlay account, separate information in respect of the following items is not available:-

- (i) withdrawals from entrepreneurial income of quasi-corporate government enterprises
- (ii) property income in the form of
 - (a) dividends and
 - (b) rent
- (iii) casualty insurance claims
- (iv) direct taxes
 - (a) on income
 - (b) not elsewhere classified
- (v) social security contributions
- (vi) unfunded employee welfare contributions imputed and
- (vii) current transfers n.e.c. from
 - (a) residents and
 - (b) the rest of the world.

In the case of the items contained in the gross accumulation part of the capital finance account, separate information is not available in respect of (i) purchases of land, net (ii) purchases of intangible assets n.e.c., net, and (iii) capital transfers, net, from residents. Further, no details are available in respect of the various items contained in the second part of the capital finance account in respect of net acquisition of financial assets and net incurrence of liabilities plus net lending. Studies are in progress to explore the possibility of compiling (i) income and outlay and (ii) capital finance accounts of the various sub-sectors of general government as recommended in the revised SNA.

13. In the standard tables, priority 1 has been assigned to the table entitled "Supply and disposition of commodities simplified version". Necessary data for compiling the estimates of various items presented in this table are not available. However, it is proposed to explore the possibility of compiling the required information to the extent possible, on an ad-hoc basis.

14. Another table which has been assigned priority 1 is entitled "Composition of gross capital formation by type of capital goods". The capital goods in this table have been classified in the following categories:-

(1) residential buildings (2) non-residential buildings (3) other construction except land improvement (4) land improvement and plantations and orchard development, (4.1) land improvement (4.2) plantation, orchard and vine-yard development (5) transport equipment (6) machinery and equipment (6.1) agricultural machinery and equipment (6.2) other (7) breeding stock, draught animals, dairy cattle and the like. The item "increase in stocks" contains the following categories:-

(1) Goods producing industries: (1.1) materials and supplies (1.2) work-in-progress (1.3) livestock except breeding stock, dairy cattle and the like (1.4) finished goods (2) Wholesale and retail trade (3) Other industries (4) Government services.

The preliminary estimates of capital formation in India compiled in the Central Statistical Organisation make a distinction between (i) construction (ii) machinery and equipment and (iii) change in stocks. The estimates of net gross domestic capital formation have been given separately for public sector and private sector and the change in stocks have been estimated separately for (i) agricultural commodities other than food grains (ii) food grains (iii) livestock, (iv) mining (v) manufacturing (vi) electricity (vii) commerce and (viii) public administration. Studies are in progress to compile the estimates of gross capital formation by type of capital goods as recommended in the revised S. N. A.

15. The part of table entitled "National income at market prices" has also been assigned priority 1. The Central Statistical Organisation publication on national product contains the estimates of net national product at market prices. Further it also contains another priority 1 table entitled "External transactions". However, detailed notes explaining the coverage of the individual items of this table as recommended in the revised SNA have been made available to the Reserve Bank of India for further studies in order to explore the possibility of making available the necessary data.

PROGRESS AND PLANS IN DEALING WITH REVISED SYSTEMS
OF NATIONAL ACCOUNTS

1. National accounts in Kenya are compiled for production and capital formation. Only rather broad estimates are available of other variables such as consumption and savings - work on more detailed sub-accounts for these is now underway. The existing accounts were developed following extensive revision of the old series and have been back-dated to 1964. Revision of the old accounts was undertaken for the following reasons:-

1. The development of statistical work since the old accounts were established in 1954 had made available a large amount of new data which enabled production and capital formation accounts for each industrial group to be calculated.
2. The previous accounts lacked sufficient detail and it was felt that for planning purposes particularly a more elaborate breakdown of the production sectors was desirable. The revised accounts are based on the 3-digit ISIC, with minor modifications to suit local conditions. The public sector has been broken down into its main components, each being distinctly identified in the accounts - Central Government, Local Authorities, Statutory Corporations and the East African Community. Each of these has been further sub-divided into "general government" and "enterprise" sectors.
3. Advances in methodology made the previous series appear crude and obsolete. The revised accounts are based on the new S.N.A.

2. The rest of this paper will be devoted primarily to a discussion of some of the methodological problems which have risen, and the manner in which solutions to these have been attempted.

3. Treatment of the East African Community in the accounts raises some interesting issues. The Community is a supra-government organization responsible for rail, water and air transport and communications operations in the three East African territories of Kenya, Uganda and Tanzania. These operations are run by corporations set up by the Community. In the revised production and capital formation accounts these corporations have been both functionally and territorially disaggregated. As far as the revised S.N.A. is concerned, the only guidance it offers on the treatment of such institutions is contained in paragraph 5. 109. It reads:

"Problems of definition of residents also arise

where an enterprise which operates aircrafts, ships or fishing fleets entirely or primarily in international commerce is jointly organized and owned by a number of companies or by enterprises resident, in a number of countries. In this case the sales and costs of production, the other current receipts and dis-bursements and the capital transactions of the entity, should in principle, be allocated to a resident corporate enterprise of each country in principle to the share of its company, or its resident enterprise, in the financial capital of the entity".

4. This solution presumes that the bulk of operations take place in international waters or air space. However, in the case of corporations run by the East African Community, by and large, the major part of their operations take place within the territorial limits of the three partner states of the East African Community. The revised S.N.A. does not offer a realistic solution to the problem posed by inter-territorial public corporations like those run by the East African Community. What has therefore been attempted in Kenya is allocation of territorial shares for these corporations on the basis of location of activity rather than ownership. These corporations have interlinked operations in all the participating territories, as well as (in the case of air transport and shipping) outside the borders of these territories. In order to break operations down by location, a lot of complex data have had to be obtained. In addition to making a territorial breakdown of the operations of the corporations have also been classified into industrial functions according to the ISIC. In doing this, the intra-functional flows of each corporation have had to be determined along with the proportion of output of each function which is either capitalized or exported.

5. Although, by and large, the methodology adopted conforms to the Revised System of National Accounts, there are some minor differences. The first set of these concerns "subsistence activities" in these accounts. Calculation of "subsistence activities" in the production accounts of Kenya has been undertaken within the context of the report by a "Working Group on the Treatment of Non-monetary (Subsistence) Transactions within the frame-work of National Accounts". This "working group" was convened by the United Nations Economic Commission for Africa and met in Addis Ababa in 1960. The report of this "working group" has been used, however, as a broad point of reference only - since there are certain differences between the definition and methods adopted for actual calculations in Kenya and the main recommendations of the "working group" - which, to a certain extent, are also incorporated in the revised systems of National Accounts.

6. To begin with the term "non-monetary" rather than

10(a) - Kenya

"subsistence" is preferred in Kenya. This is due to the extent of production covered by this sector. It includes the production of all goods and services which are produced within the rural households whether for "non-monetary" exchange or for own consumption, and which are comparable with those normally produced by enterprises in the monetary sector, has therefore been classified into the following categories:

- (1) Agriculture
- (2) Forestry
- (3) Fishing
- (4) Construction and building
- (5) Ownership and dwellings
- (6) Water (collection)

7. Coverage extends to the production activities of rural households as opposed to industrial/commercial enterprises. The households covered are engaged primarily in farming, forestry and fishing, outside the "modern sector" which also embraces some enterprises in the rural areas, e.g., large scale farms. However, contrary to the views of the working group and the S.N.A., the product contribution of "ownership of traditional dwellings" is also included in the revised Kenya Accounts. This follows the classification of construction of these dwellings as capital formation by end-use in the first place. To make the product estimates consistent with capital formation estimates, therefore, production accounts must reflect the production of the capital assets thus created. This is the second point of difference with the main recommendations of the working group whose contrary view on this point states: "The working group does not consider the classification of such production (traditional dwellings) as capital formation necessarily involves the imputation of a rental income and has not provided for this in its recommendations".

8. For categories 1 and 3 noted above, valuation of the commodities which have marketable counterparts has been undertaken at producer prices. Average prices have been used - calculated as a simple average for each region and for each commodity in a given year, prevailing in markets near the main producing centres. In those areas where such prices are not obtainable they are instead derived from prices in the neighbouring main marketing centres.

9. For activities which have no comparable market sectors, other methods of valuation have been applied. As in similar cases in the monetary sector, where there is no market, e.g. for the government sector, valuation at cost has been adopted and is considered as the next best alterna-

tive and the opportunity cost of labour is considered to be the best indicator of this. Output therefore is valued by using wage rates which are offered in alternative employment to the type of labour similar to that engaged in the production of these goods and services i.e., the wage rates which this labour could expect to earn in the vicinity of production if it were to opt for wage employment instead of production for own use. The wage rates used are calculated on an hourly basis and applied to very rough estimates of the time spent in production of these goods and services. Finally, the output of ownership of dwellings is an evaluation of the imputed rental value of these dwellings based on their depreciable lifetime.

10. For the monetary sector also there are some differences between the revised S.N.A. and the practices adopted in the current calculations in Kenya. In forestry, an imputation has been made for estimated output of planted forests, on the basis of the annual growth of timber in these forests. In mining, "negative" product for oil exploration is avoided by "exporting" the loss on such oil prospecting to the country from which the company undertaking oil prospecting originates. The company in question is treated as an agency rather than a locally registered firm. In insurance it has not yet been possible to undertake product calculations separately for casualty and life insurance business. However, it is hoped that the necessary data would be forthcoming from a new statistical inquiry of insurance companies. There are also differences in the treatment of bank service charges on households and depreciation charges on government machinery and equipment. Bank services charges under the revised S.N.A. need no longer be imputed to household expenditures. This is a difficult change to make in the current calculations since the basic data to make this adjustment cannot be obtained from existing banking statistics. Also, consumption of government machinery and equipment is not included in product calculations of the government sector since here too the basic data required to make this adjustment are not yet available. However, these modifications will be introduced in due course.

11. These in brief are some of the main features of the national accounts as they exist at the moment in Kenya. Currently, a fairly detailed input/output analysis is being undertaken for 1967. One of the objectives of this exercise is to test the internal consistency of the revised production accounts of the monetary sector. Some results of this exercise have already revealed some discrepancies in the estimates for the monetary sector. Efforts are also being made to improve the coverage and estimation of production in the non-monetary sector through rural household budget and agricultural surveys.

12. Apart from refining the quality of existing national accounts, consideration is also being given to extension of

10(a) - Kenya

se accounts beyond production. It is hoped to make more exact estimates of consumption and savings on basis of the results of Rural and Urban Household Budget Surveys which are underway right now. This will enable consumption and savings - the only elements in estimates of use of resources hereto arrived at as residuals - to be directly estimated. In addition, efforts are being made to establish income distribution statistics. Attempts are also being made to improve the quality of data for constant price calculations in order to make these more refined.

Progress and plans in dealing with the revised
system of National Accounts

1. The first National Accounts report for Lesotho was produced for the years 1964/5 and 1965/6. In constructing the tables in this report, it was mainly the income approach that was adopted and, taking into account the lack of detailed data from some sectors which are required for the successful use of the product approach, the income approach will be used until after the 1968/69 report has been produced. The Bureau of Statistics has always been aware of the serious shortcomings in its statistical series and, recognising the need to adopt the revised system of National Accounts, sought the expert guidance of the Economic Commission for Africa to pave the way for a smooth transition to the new system. The regional adviser in National Accounts arrived here in November 1969 and spent about two weeks on:

- (a) making a general review of available statistics as well as suggestions for their improvement;
- (b) examining the national accounts work so far done and making recommendations on the methods of estimation and compilation and presentation of the accounts.

Frames of the four consolidated accounts of the nation (recommended in the revised S.N.A.) as well as ten supporting tables were drawn.

2. Unfortunately, at the time of the adviser's visit, questionnaires for collecting national accounts data had already been sent out for the fiscal year 1968/69 and the suggestions made regarding improvement in the statistics collected will be implemented in about August, 1970 when these suggestions will be incorporated in the revised questionnaires for 1969/70.

3. With regard to the Agricultural Census which was in progress when the adviser arrived, she recommended that information on some items which were not included in the census questionnaires could be obtained from the proposed annual survey of current agricultural statistics scheduled to start after the main census. These include:

- (i) Own consumption in quantity: human consumption, animal feed, waste and stocks;
- (ii) Commercialization in quantity and value;
- (iii) Prices at nearest market (considered as farm prices);
- (iv) Credit received by farmers during crop year (from banks and co-operatives), the duration of the credit and the interest on the credit;
- (v) Production of vegetables and fruits: Own consumption and sales and prices received by the farmers.

The following additional information on livestock could also be

included in the annual survey:

- (i) Cost of rearing livestock - animal feed, chemicals etc;
- (ii) Prices received by farmers for each type of livestock;
- (iii) Capital formation - equipment, sheds etc. during the year concerned;
- (iv) Credit received by livestock farmers (from banks and co-operatives), duration of credit and interest thereon.

4. Improvements in the questionnaires for collecting data on the following were also suggested.

- (a) Manufacturing,
- (b) Building and Construction,
- (c) Imports and Exports,
- (d) Services.

5. In conclusion, it might be stated that a foundation has been laid for a gradual change over from the old to the new S.N.A. However, it is the rate at which the implementation of the recommendations made will take place that will dictate the pace at which the change over will take place. It has to be appreciated that for a country like Lesotho, with a very young Bureau of Statistics, the fundamental problems of making our respondents statistics conscious and educating them to maintain good records of their business activities have not as yet been overcome. The Bureau has since 1965 been encouraging businesses to keep proper books of accounts and this effort on its part has shown some encouraging results. The new S.N.A. necessarily requires elaborate details and a complete change over to the system can only be attained after many years of relentless work.

PROGRESS AND PLANS IN DEALING WITH
THE REVISED U.N. SYSTEM OF NATIONAL ACCOUNTS

1. The catalyst for the radical revision of statistical collections and methods now taking place in New Zealand was the National Development Conference. This was a body charged with examining indicative planning and the implementation of targets.
2. The inadequacies of New Zealand's existing national accounts were highlighted during the Plenary Session of the N.D.C. by the publication of the U.N. "A System of National Accounts". Impetus was given to the early development of a national accounting system based on that proposed by the United Nations.
3. Much of the work of the National Development Conference had to be done on the basis of estimation as a result of unsatisfactory statistical information, particular in the National Accounts.
4. A Technical Committee on Statistics was subsequently set up by the National Development Council and made proposals for considerable changes in the nature and extent of statistical collections. Nearly all of these changes are based on the concepts and structure of the SNA. Included in this Committee's proposals were specific recommendations for the preparation, by 1972, of the major part of the U.N.-S.N.A. These proposals involve the preparation of annual accounts within six months of the end of each financial year and less detailed quarterly accounts. It is proposed to complete those parts of the accounts required for indicative planning purposes and they are sufficiently comprehensive to meet most general needs. The industrial and institutional sector classification of the U.N.-S.N.A. are being similarly adapted to meet New Zealand's planning needs whilst at the same time fulfilling requirements for international comparability.
5. Looking beyond 1972 the same Committee has made proposals for the implementation of other parts of the United Nations System. The proposals have been accepted as desirable by the National Development Council and the Department of Statistics.
6. Full details of the accounts and tables to be included in the New Zealand system, along with details of their relative priority are available on request from the Department of Statistics.
7. Progress to date with these proposals has been mainly exploratory with the emphasis on training staff and making prospective users of statistics sufficiently familiar with the new system to be able to contribute to its development for New Zealand. Formal and informal discussions have been held with Government departments, members of the National Development

Councils and their staff, representatives of private enterprise, and the Department of Statistics.

8. A survey of electricity usage is also to be carried out by the Department of Statistics. This is regarded in part as a pilot study and a means of building up the nucleus of a staff of statisticians and interviewers necessary for a full sample survey of household income, expenditure and savings. A full survey will be a 'first' in New Zealand and is considered to be essential to the compilation of the more detailed national accounts called for in the United Nations proposals.

9. The enterprise census, as a means of collecting production and related statistics, will be a further major development in New Zealand. A three-man team is at present examining alternative methods and their feasibility in the New Zealand context, prior to a firm decision on this method being taken. New Zealand statisticians regard the need for some such method of integrating statistical collections as a vital prerequisite for accurate national accounts and supporting series.

10. Other changes in basic collections are less dramatic. They have been made and will be made as the development of the national accounts reveals a need for additional information.

11. These long term plans will take several years to implement. During this intervening period it is intended to prepare the accounts as best we can using existing series, particularly the inter-industry study of 1965-66 and some ad hoc collections.

12. Some of the basic classifications necessary for the accounts have been prepared in the first six months of development work. The preparation of a five digit industry classification based on the revised United Nations I.S.I.C. (M4 Rev.2) is well in hand and should be completed by July 1970. The broad industrial groups for which it is intended to prepare production accounts in the N.Z.-SNA have been finalised. To meet the planning needs and the need of more general users, it has been found necessary to have 27 classes which can be combined either in the more usual way following the Divisions of the I.S.I.C. or by grouping together vertically related industry classes. This latter grouping is required to provide relevant information on the performance of the industries for which the various (industrial) Sector Councils of New Zealand's planning body are responsible. The 27 classes and their place in the indicative planning system are shown in Appendix A. Appendix B. sets out the institutional sector classification for New Zealand.

13. Some additions have been made to the United Nations classification. The additional sub-classes in the financial institutions sector identify Government data necessary to enable the full place of Government in the economy to be seen.

14. The class for 1(b) meets a particular New Zealand need. Producer Boards are largely statutory bodies generally responsible for marketing an agricultural product. They have considerable powers granted by Government but are independent of Government in many respects. A separate sub-class seemed the only reasonable way of treating these organisations.

15. One important point which is not apparent from the classification is New Zealand's definition of a "quasi-corporate non-financial enterprise". In New Zealand the possible legal forms of enterprise include the public company, usually large; the private company which ranges in size from an organisation with a multi-million dollar turnover to a small family or even one man concern and the unincorporated individuals and partnerships. The unincorporated businesses can be, and often are larger, than the small incorporated businesses. Both keep similar accounts for taxation purposes and in practice both "mingle and manage" the financial assets and liabilities of the enterprise and the household together.

16. In this situation New Zealand has to make a meaningful and practical distinction between corporate and quasi-corporate enterprises on the one hand and unincorporated enterprises on the other. A distinction between the small company and the normal one-man enterprise is meaningless but to identify the "small" private company from the "large" private company is impractical. The conclusion is that the only feasible and the most useful grouping is to regard all businesses that keep taxation records as quasi-corporate. As a result it seems probable that New Zealand will include some enterprises in Sector 1 which in other countries would be included in Sector 5.

17. We have also had some trouble in preparing a usable definition and classification of capital formation by type. The distinction between "fixed capital" livestock and "stock" livestock whilst alright in concept is not really practical especially when one tries to apply it to the 40 million sheep born in New Zealand each year. Whether they are to grow wool or go for slaughter is a week by week decision dependent on relative prices of wool and meat and on the availability of feed. New Zealand has kept the United Nations distinction but the measure of it will necessarily be an arbitrary one in practice.

18. In the case of expenditure on trees grown for timber we feel that the U.N.-SNA has not followed the related transactions through the accounting system. The present U.N. proposals would require sale of timber trees to be a sale of capital which is unrealistic. New Zealand's present proposal is to treat timber as work in progress.

19. Except for the points raised above the classification of capital proposed by the U.N. has been accepted by New Zealand. Whilst some finer points are still being argued it seems probable that the classification of capital formation by type will be finalised by the middle of 1970.

20. Other classifications which are being prepared are the classification of commodities by characteristic industry of production, of types of financial assets and the purpose classification of Government activity.

21. Some attempts have been made to prepare production accounts using existing sources of information. Whilst this has been mainly a training exercise it has shown up many areas where additional information is required. In some cases the economic transactions required by the system will never be available from financial accounts of enterprises or be measurable except in some arbitrary fashion.

22. Particular difficulty has been found with the treatment of rent. The conceptual distinction between rent for non-reproducible assets and that for reproducible assets may be a useful one. However, there is considerable difficulty in identifying the two elements and this may outweigh any advantages in having a measure of operating surplus which is independent of the ownership of non-reproducible resources used in production.

23. Other problems of measurement are apparent with depreciation. Here it seems probable that we shall be forced into measuring the value of capital stock, something our present statistical collection system is not geared for, before consumption of fixed capital can be estimated with reasonable accuracy.

24. Similar problems arise with the measure of depreciation, the practical distinction between capital and current repairs, and the practical treatment of bad debts, discounts and exports sold on consignment.

25. New Zealand is only on the brink of developing a system of national accounts in the sophisticated form proposed in the 1968 System of National Accounts. We can see many problems and these are probably only the tip of the iceberg. There is a definite need for the implementation of the system and we are hopeful that resources will continue to be available to complete the programme that has been outlined.

Attached: Appendix A
Appendix B

APPENDIX "A"

<u>Industry groups to be used in the N.Z.-SNA.</u>	<u>Arrangement of N.Z.-SNA industry groups to provide accounts suitable for planning</u>
1. Agriculture and hunting	<u>Agricultural Production Council</u>
2. Forestry and logging	a. Agriculture and hunting
3. Fishing	b. Manufacturing Industries Processing Agricultural Products
4. Mineral fuel extraction	<u>Fisheries Council</u>
5. Other mining & quarrying	a. Fisheries
6. Manufacturing n.e.i.	b. Manufacturing Industries Processing Fisheries Products
7. Manufacturing Industries Processing Agricultural Products	<u>Forestry Council</u>
8. Manufacturing Industries Processing Fisheries Products	a. Forestry and Logging
9. Manufacturing Industries Processing Forestry Products	b. Manufacturing Industries Processing Forestry Products
10. Manufacturing Industries Processing Mineral Fuel Products	<u>Minerals Council</u>
11. Manufacturing Industries Processing Mineral Products	a. Mineral Fuel Extraction
12. Fuel and Power	b. Other Mining and Quarrying
13. Water	<u>Manufacturing Council</u>
14. Construction	a. Manufacturing n.e.i.
15. Distribution	b. Manufacturing Industries Processing Mineral Products
16. Hotels, Restaurants	c. Manufacturing Industries Processing Fuel Products
17. Transport	<u>Fuel and Power Council</u>
18. Communications	Fuel and Power
19. Finance and Insurance	<u>Building Industry Advisory Council</u>
20. Real Estate	Construction
	<u>Distribution Council</u>
	Distribution
	<u>Transport Council</u>
	Transport

APPENDIX "A" (cont'd)

<u>Industry groups to be used in the N.Z.-SNA.</u>	<u>Arrangement of N.Z.-SNA industry groups to provide accounts suitable for planning</u>
21. Owner Occupied Dwellings	<u>Government Programming (Treasury)</u>
22. Other rental and Leasing	a. Producers of Central Government Services
23. Services	b. Producers of Local Government Services
24. Producers of Central Government Services	<u>Industries excluded from N.D. Council purview</u>
25. Producers of Local Government Services	a. Water
26. Producers of Private Non-Profit Services to Households	b. Hotels and Restaurants
27. Domestic Services to Households	c. Finance and Insurance
	d. Communications
	e. Real Estate
	f. Owner-occupied dwellings
	g. Other Rental and Leasing Services
	h. Services
	i. Producers of Non-Profit Services to Households
	j. Domestic Services to Households

APPENDIX "B"INSTITUTIONAL SECTOR CLASSIFICATION FOR NEW ZEALAND

1. NON-FINANCIAL ENTERPRISES (CORPORATE AND QUASI-CORPORATE)
 - (a) Private (including private non-profit enterprises serving business)
 - (b) Producer Boards
 - (c) Government
 - (i) Central
 - (ii) Local
2. FINANCIAL INSTITUTIONS
 - (a) Reserve Bank of New Zealand
 - (b) Trading Banks
 - (i) Government owned
 - (ii) Other
 - (c) Insurance and Pension Funds
 - (i) Central Government owned
 - (ii) Local Government owned
 - (iii) Privately owned
 - (d) Other Financial Institutions
 - (i) Central Government owned
 - (ii) Local Government owned
 - (iii) Privately Owned
3. GENERAL GOVERNMENT
 - (a) Central Government
 - (b) Local Government
4. PRIVATE NON-PROFIT INSTITUTIONS SERVING HOUSEHOLDS
5. HOUSEHOLDS
6. BEST OF WORLD

Progress and plans in dealing with the
revised system of national accounts

1. For Tanzania (Mainland) official national income estimates date as far back as 1954. Till last year, the Bureau of Statistics was compiling estimates of gross domestic product and capital formation of Tanzania (Mainland) broadly in conformity with the United Nations recommendations on the subject contained in the old 'System of National Accounts and Supporting Tables'. The methodology adopted in practice was, however, very crude in many respects and was largely dictated by the availability of data obtaining in the early sixties. Since then a considerable amount of statistical data had become available through the population census, survey of manufacturing industries, and other ad-hoc surveys and enquiries. Furthermore, many defects or inadequacies of this series for economic planning purposes had also assumed considerable importance. Realising all these, the Bureau of Statistics embarked upon a detailed and comprehensive revision of the national accounts of this country and this opportunity was taken to effect the changeover to the revised system of national accounts.
2. In February this year, the Bureau of Statistics brought out a report entitled 'National Accounts of Tanzania, 1966 to 1968' which presented for the first time the consolidated accounts of the nation along with a number of detailed tables on the industrial origin of domestic product, breakdown of capital formation, analysis of the accounts of various public authorities and the account of public corporations for the three years 1966 to 1968. These estimates have been built up in conformity with the latest recommendations of the United Nations on the subject contained in 'A System of National Accounts' - U.N. Series F.2 Rev.3. Three points of divergence may, however, be mentioned here. Firstly, it has not been possible to identify and accord separate treatment to 'Machinery and equipment rental and leasing'. This economic activity, however, is of no major importance in the Tanzanian economy and is, in the new series included as part of the operating surplus of the respective industry of use of the equipment. Secondly, automobile repair services have been included under manufacturing industry contrary to the U.N. recommendations to treat the same as personal services. This is so because in Tanzania by far the larger proportion of automobile servicing activity relates to business vehicles and cannot be considered as personal or household services. Thirdly, motor vehicle duties have been classified entirely as indirect taxes as the major portions of the receipts under this duty in Tanzania relate to business vehicles and further no separation of the vehicle duties into that paid by household and that paid by nonhouseholds has been possible. Similarly no split-up has been attempted in respect of some of the other government receipts such as airport charges, passport duties etc. as well.
3. Further, in conformity with the U.N. recommendations, three types of own-account activities of households have been treated as industries in the revised series. They relate to (1) the dwellings which households own and use on their own account (owner-occupied dwellings), (2) the own-account construction of dwellings in rural areas and (3) the own-account production by households of primary commodities (subsistence production). The own-account construction of non-residential buildings in rural areas have, however, been ignored for want of any relevant data. Again, in conformity with the U.N. recommendations a nominal finance industry has been recognised and its negative operating surplus taken as equal to the

imputed service charge of financial intermediaries to avoid double counting. Thus there is no imputation of bank services charged to households. Also, no imputation has been made for net rents in respect of buildings and structures owned and occupied by government.

4. The revised international standard industrial classification of all economic activities (with two minor amendments) has been adopted for the industrial origin classification of domestic product.

5. Analysis of expenditure of public authorities by purpose as recommended in the revised system of national accounts has also been undertaken for the years 1965-66 to 1969-70.

6. The revised series of national income estimates of Tanzania (Mainland) has, presently been carried forward to 1969. It is proposed to build up later in the year comparable estimates for 1964 and 1965 as well.

7. Independent estimates of private consumption expenditure based on commodity flow approach as well as household budget survey data remain to be built up. These can, however, be taken up only next year after the results of the nationwide household budget survey become available.

PROGRESS AND PLANS IN DEALING WITH THE
REVISED SYSTEM OF NATIONAL ACCOUNTS

1. The system used by the United Kingdom for our own national accounts is described in the book published in 1968 "National Income Statistics: Sources and Methods". The system does not follow precisely the former SNA but is reasonably close to it. There are differences in the treatment of a few relatively small items and in some of the classifications. However, in most cases the United Kingdom figures could be adjusted to SNA concepts for the purpose of the United Nations annual questionnaires without too much difficulty, although some differences in classification remained.

2. With the arrival of the revised SNA we are faced with two sets of problems - first how far it will be possible for us to complete the new United Nations questionnaire from data already available and how far new data will have to be collected or special estimates made; second, how far, if at all, should we modify our own national accounts to bring them more in line with the concepts of the revised SNA. At the time of writing it is possible to provide the answer to most of the problems in the first group, but only to some of those in the second. Our decisions on possible changes in our own accounts have not yet been made, but they are likely to reflect the extent to which we consider the revised SNA is an improvement on its predecessors for national purposes as distinct from international use.

The structure of the system

3. The main changes in the structure of the revised SNA are the introduction of input/output tables and the setting up of institutional sectors, each with an income and outlay account and a capital account leading to tables on the acquisition of financial assets and the incurrence of financial liabilities, and, eventually, to balance sheet data. These extensions do not, in principle, present any difficulties for the United Kingdom. We already link our input/output tables with our national accounts, although, unlike some countries, we do not use the input/output tables to construct the national accounts but rather reconcile the two sets of data. We have for many years had income and outlay account for institutional sectors and several years ago these were extended to include data in financial transactions. The further development of the financial accounts to include balance sheet data is already under consideration. Nevertheless, on a number of points of detail, we cannot at present follow the revised SNA completely. For example, our input/output tables are based on our own Standard Industrial Classification, which differs in some points from ISIC and the tables are not as detailed in respect of the service industries as those proposed for SNA. The institutional sectors we already use are very much the same as those proposed in SNA except that, in common with most other countries, it is likely to be some time before we are able to establish a sector for private non-profit institutions separate from the rest of the household sector.

The main changes in concepts

4. In the revised SNA, valuation at factor cost has almost disappeared - even national income is now valued at market prices. Where factor incomes are still used, eg in Table 5 of the new United Nations questionnaire,

operating surplus is to be recorded net of capital consumption for each industry group. We shall have no practical difficulty in adjusting our figures to these new concepts so far as aggregates are concerned, but there will be problems in allocating both indirect taxes and capital consumption by industry groups. At present in the United Kingdom accounts the analysis by industry relates to gross domestic product at factor cost. Some specific taxes, eg those on tobacco, alcoholic drink or petroleum products, can be allocated readily to the appropriate industries but for others - motor vehicle licence duties, local rates, purchase tax etc. - we have not at present the data necessary to make a precise allocation, and will have to rely on estimation, at least in part. Moreover, the allocation in principle of some of these taxes, under the United Kingdom system, is not always clear. ×

5. At present, estimates of capital consumption are published regularly for institutional sectors and for the main types of fixed asset, but not for industry groups. The basic data is available for an allocation by industry, since we already publish an industry allocation of the total capital stock, but the production of industry estimates will involve a considerable amount of work and is likely to take some time. However, preparations are in hand for transferring most of the work on capital stock and capital consumption to a computer and when this has been done - we hope later this year - the work of producing an industry analysis of capital consumption should be simplified.

6. It is convenient to mention at this point an allied problem which arises in the analysis of government expenditure by purpose. Here, we have hitherto followed the former SNA in including an estimate of the imputed rent of government owned buildings. The revised SNA omits this imputation but substitutes capital consumption. It will therefore be necessary for us to estimate the detailed allocation by purpose of the expenditure of the estimates already available in total for capital consumption in the government sector. Apart from this difficulty, we welcome the new table in SNA which provides for the analysis by purpose not only of current expenditure on goods and services but also of transfer payments, capital formation, etc. We have had a table on these lines in our own accounts for many years. It gives a more detailed analysis by purpose than is called for in the revised SNA but can be abridged to the latter fairly closely.

Classifications

7. In the United Kingdom, we have our own Standard Industrial Classification, which is broadly similar to, though not identical with, International Standard Industrial Classification. The United Kingdom classification was revised in 1968 at about the same time as the revision of ISIC and changes in both have brought them closer together than they were before. For example, the change in ISIC whereby establishments specialising in the repair of consumer goods (footwear, watches, motor vehicles, radio and television etc) are now classified in service industries rather than with the corresponding manufacturing group is in line with United Kingdom practice while the United Kingdom classification now follows ISIC in treating tea blending, coffee blending, roasting and grinding and the heat treatment of milk as part of food manufacturing rather than as part of the distributive process. We are in the process of re-classifying our own national accounts (including the input/output tables) on the basis of our 1968 Classification, carrying the re-classification as far back in time as data permit. When this

has been completed, we shall be able to provide data for the new United Nations questionnaires which will only differ from ISIC on a few minor points.

8. The United Kingdom classification of consumers' expenditure is fairly detailed and can be re-arranged to match that in SNA for most categories. There will, however, be differences in some cases. Our own classification provides for three categories of durable goods. Two of these, motor vehicles and furniture and floor coverings, correspond to SNA groups but the third covers not only household durables (cookers, refrigerators, washing machines etc) but also radio and TV sets and other similar items which are classified in different SNA groups. It is unlikely that we shall be able to breakdown our figures in order to follow SNA. Also, as a result of our mainly free National Health Service we shall have little to report under the heading "Medical care and health expenses". Some items which would be appropriate here - the medical and pharmaceutical products in SNA heading 5.1, to the extent that they are purchased outside the National Health Service - are included in the United Kingdom figures with toilet preparations etc. under a single heading "Chemists' goods". Here again, it is unlikely that we shall be able to adjust our figures to fit the SNA classification.

9. The analysis we have available for capital formation is less detailed than that recommended in the new SNA. For example, in the case of stocks, the analysis between materials, work in progress and finished goods, is available only for manufacturing industries. In the asset analysis of fixed capital, we distinguish dwellings, other buildings and works, ships, vehicles and aircraft, and plant and machinery. It is unlikely that we shall extend our figures for either stocks or fixed capital formation to show separately the other categories recommended in SNA, since the additional work which this would impose on respondents to statistical enquiries could not be justified by the value of the further figures.

10. Our classification of exports and imports of goods and services is based on the IMF Balance of Payments Manual and it is unlikely that we shall change it unless and until the Manual is revised. There will therefore be some minor differences between what we can provide and the recommendations in the revised SNA. Some of these recommendations, while logical in the context of the system, would be difficult to apply in practice without asking respondents for additional data which may well not be readily available. For example, the expenditure of the crews of ships and aircraft in overseas countries is not at present distinguishable from the other overseas expenditure of the shipping companies and airlines concerned. Similarly, in the case of individual travellers overseas, while data are available which enable us to distinguish between holiday travel and business travel, it is not possible to breakdown the latter between the outlays reimbursible by the employer and those made by the traveller on his own account.

11. As mentioned above, our own classification by purpose of government expenditure fits in well with the revised SNA. This is also the case with the classification of financial transactions.

Minor points

12. There are a number of detailed points on which our present practice differs from that recommended in the revised SNA. For some of these we

shall be able to make adjustments and may even embody them in our own accounts, but there are others for which the necessary data are not at present available and are not likely to become available because we believe that the attempt to collect this information could not be justified in the United Kingdom. In general, however, this will have little effect on the main aggregates. One example is the proposal to treat purchases and sales of land as separate entries in the capital account, apart from fixed capital formation. It will not in general, however, be practicable to do this. The effect on the aggregates will, of course, be nil since inter-sector transactions in land will cancel out, apart from the transfer costs incurred. Similarly, it would be extremely difficult to treat the rent of land separately from other rents. We shall not follow the proposal to treat certain large unincorporated businesses as "quasi-corporate" and include them with corporations. Even if it were possible to identify such businesses for some purposes, it would not be feasible to ensure that they were treated consistently throughout the accounts. However, such businesses are not important in the United Kingdom. We shall probably amend our own treatment of life assurance and pension funds to bring it more into line with the revised SNA, but it may be difficult to do this in the case of casualty insurance, again because of lack of data on which to allocate premiums and claims between different sectors and different industries. A number of other examples could be cited, but none of them is of great importance.

SYSTEMS OF REGIONAL ACCOUNTS

Introduction

Australian accounts have so far been compiled mainly on a national basis, although some estimates of personal income and expenditure and farm product have been compiled and published for each State. The basis has been laid in the new Integrated Economic Censuses for the development of estimates of gross product by industry for each State. This paper reviews in a general way some of the problems which have been recognised in the compilation and use of regional accounts, together with some indication of the type of work which it is hoped will be undertaken in this field in the near future.

Concepts and Uses

2. It is difficult to assess the demand for regional accounts in Australia. Some of the conditions which result in strong pressures for their development do not exist here to the same extent as in some other Commonwealth countries. For instance, variations in income per head from one State to another are not significant enough to be a matter of much policy concern nor are there serious problems of depressed areas with chronically high unemployment rates. Perhaps the greatest regional disparity is between the populations of metropolitan areas and the remainder of the country, evidenced by the exceptionally high and growing concentration of population in the capital cities and major urban areas. This in itself could present an increasingly important problem which might warrant a rather different structuring of regional accounts than is contemplated at present. Interest in Australia therefore tends to centre on urban-rural problems and equity questions arising from Commonwealth/State financial relationships, but there is some interest in regional data as an aid in considering the allocation of regional resources involved in national development proposals. From time to time, the Bureau of Census and Statistics receives requests for information on gross product by States, for the product of particular industries by States and for product and expenditure within a particular area or State. Requests have also been received from those interested in town-planning and marketing for information on personal income in relatively small geographical regions. For Western Australia in particular efforts have been made at the University of Western Australia to compile a reasonably full set of regional accounts, including a balance of payments, etc. and an input-output study of the Perth metropolitan region.

3. The requests that are received and the uses to which the information is to be put are frequently not clearly formulated. In some instances these requests appear to be based on a view that since national

accounts are essential statistical equipment at the national level so by analogy regional accounts are required at the lower level. Sometimes these requests are market oriented and are expressed as a demand for regional accounts when in fact estimates of regional product, or even personal income, would appear to be more appropriate. It is evident, however, that there are many problems which could be handled much more effectively using statistics tested for consistency within an accounting framework. Although the regional analyst has lacked a general purpose framework and has had to rely on whatever partial and often unco-ordinated statistical series he could find for the job on hand, there has not been great pressure for the development of a body of regional statistical data into a working continuous system in which the analyst might with confidence jointly employ different but logically related series.

4. Australian national accounts were developed mainly as a tool for use in connection with stabilisation policies during the last war and in the relatively full employment situation which has existed since. This particular use would seem to have a limited importance in relation to regional accounts.

5. More recently, increasing attention has been given to problems of economic growth and the attendant problems of achieving an optimum allocation of resources. This aspect of national accounts is perhaps of major importance in relation to regional accounts. Many of the potential users of regional accounts appear to be interested in relative growth rates in particular regions, in potential developmental projects in particular regions, in developments in productivity and costs between different regions, and related matters. It seems doubtful, however, that a full system of accounts, parallel with the national accounts, is necessary for studying this type of problem. For example, problems associated with the balance of payments, overseas investment and capital financing are generally much more important at the national than the regional level, where the transfer of financial resources from one area to another presents few problems. However, even the compilation of regional production accounts requires information at least on the goods and services components of the inter-regional transactions and the regional dissection of external transactions. These are obviously important in any consideration of the allocation of resources between particular regions, but problems of obtaining sufficient and reliable data, particularly of inter-regional transactions are considerable even for a comparatively small number of States or regions - as is the case in Australia. At present three States compile somewhat limited statistics of interstate trade, but information for the other three states is incomplete. To complete the trade figures for the remaining States and to make estimates for interstate service transactions would be a formidable task, requiring something in the way of tables for each State

10(b) - AUSTRALIA

showing the supply and disposition of commodities. As mentioned earlier, academic researchers have done some pioneering work of this kind for Western Australia but the Bureau has no plans at present for undertaking such work for all States. Action by the Bureau may well depend on demand for State Statisticians to produce this sort of information in particular States.

6. Another use of the national accounts has been in connection with income distribution policies. Here also regional accounts may be of considerable importance, but again it may be that only selected parts of the accounts are really needed rather than a fully developed system. For example, for many years the Commonwealth Grants Commission has been concerned with the problem of comparative taxable capacity among States and the need to ensure that the various claimant States did not fall below minimum national standards in respect of certain social services. These considerations involved a detailed analysis of government accounts in the States. The importance of the Grants Commission's recommendations has declined in recent years - they now concern only one State - but equalisation problems still remain and the statistical techniques and published information have not always kept pace with the changes brought about by the increasing role of the Commonwealth through the medium of "special purpose" grants to all States.

7. It would seem, therefore, that there is a need to develop a regional accounting framework. Academic studies, however rudimentary in the first instance, would be useful aids to understanding and would offer some needed guidance and stimulus for the development of regional statistics. But at least until international standards are established in this area there is no real case for the Bureau of Census and Statistics to attempt to produce a set of continuing balanced accounts. Rather there appears to be a need to produce (within a "shadow" accounting framework) estimates for the immediately useful components for which reasonably sound estimates are feasible.

8. In the absence of a clearly defined demand for a full integrated set of accounts for particular States or regions within States, development in the Bureau will proceed relatively slowly in a step-by-step manner rather than with some priority programme to develop these systems of accounts. As the first step, it is hoped to develop estimates of gross domestic product on a State basis (hereafter referred to simply as "gross State product".) ⁽¹⁾ to supplement the existing systems of personal income

(1) In the Australian National Accounts product and production are regarded as essentially geographic concepts which cannot properly include income received out of other countries' products. Consequently, the term gross national product, is used instead of gross domestic product, and gross State product is an exact equivalent. Similarly, gross product by industry needs no "domestic" qualification since it is "domestic" by definition.

by States. The Bureau would also expect to compile estimates of other components of expenditure in addition to the estimates of personal consumption already published, but conceptual difficulties may preclude complete estimates in this field. The problems of providing income estimates for smaller regions within States present few conceptual difficulties, but there are considerable practical difficulties, and progress in this field is still somewhat uncertain. Plans and possibilities in these three fields are discussed in more detail later in this paper.

9. Progress with the regional product estimates in Australia is likely to depend in part on progress in utilising the Integrated Economic Census data within the national accounts and input-output framework. The problem is first to align the income tax tabulations with the integrated census tabulations for enterprises in the scope of the censuses. To do this positively enough for regional product estimates, it would be necessary to match the companies in the two data sets at the individual record level so as to ensure that duplications or gaps do not occur when income tax and census tabulations are used together. The problem arises from the fact that the income tax records as they are received in the Bureau have no identification other than the (not always reliable) industry description as supplied by the taxpayer. Currently, attempts are being made to devise procedures acceptable to the Taxation Commissioner and which overcome this problem but the indications are that legislative changes would be necessary and perhaps difficult to achieve.

Gross State Product (i.e. gross domestic product by State)

10. In planning our integrated economic censuses for 1968-69, a co-ordinated set of returns for both enterprises and establishments was introduced. The censuses are those of mining, manufacturing, gas and electricity, wholesale and retail trade (including hotels and restaurants and certain services such as laundries, hairdressers and garages). Returns were required for all establishments whose operations are predominantly in these industries and enterprise returns were required for all enterprises which were mainly engaged in any of these industries. The returns sought details of the operations of separately located head offices and ancillaries such as stores, transport depots, etc. as well as of the operating establishments. The questions were framed in such a way as to enable estimates to be made of the contribution of each enterprise and establishment (main or ancillary) to gross product.

11. Strictly speaking, of course, this can only be done for enterprises. However, many enterprises operate in several industries and in several States, and thus, for both industry and regional product calculations, it is necessary to make estimates on the basis of the establishment, which can be allocated to a single industry with more precision than the enterprise because an establishment is defined as a separate location, and is unequivocally classifiable by region. The census returns will in

general provide full details of purchases, sales, stocks, wages, rent and other operating expenses on an establishment basis, and the enterprise return for multi-establishment enterprises asks for the balance of expenses which were considered inappropriate or undesirable for inclusion in the establishment returns. These miscellaneous expenses which are reported on the enterprise return but not on the establishment return will be allocated to the individual establishments in proportion to their value added. Compared with the possible alternative of allocating the whole income of enterprises to particular regions, this adjustment, which in any case applies only to multi-establishment enterprises, is concerned with only a relatively minor element of costs. Errors or uncertainties in their allocation cannot seriously affect the results based directly on establishment returns though, of course, there are other reasons for wanting to know their precise magnitude in total. In addition, for all enterprises, the wages and salaries of head offices and ancillary units are "charged" to the establishments they serve. By this process the product of the establishment proper is reduced and the product of the ancillary unit (defined as equal to its wages and salaries plus, in the case of head offices, net rent received)⁽²⁾ can be correctly classified by region. For establishments belonging to some "out-of-scope" enterprises (i.e. those operating predominantly in industries other than those listed above), and for which there is no enterprise return, appropriate estimated adjustments may have to be made, but in many cases (such as government factories) no adjustment will be necessary.

12. Therefore, it is expected that the results of the Integrated Economic Censuses will provide estimates of gross product by State (as well as by industry) which are on a consistent establishment basis and, although based to some extent on convention, more accurate than anything which has been available in the past. In the past the Bureau, like statistical services in many other countries, has had to work from statistics of income on an enterprise basis, and it has frequently been pointed out that the arbitrary assumptions involved may pre-judge the very matters on which the statistics are intended to throw light. Of course, the method adopted is not without its assumptions and limitations. The valuation of transfers between establishments inevitably lacks precision in some cases even if firms try to follow the preferred method, and where they do not the adjustments made in the Bureau must likewise lack precision. It is also recognised that allocating expenses on the basis of "value added" may not be the theoretically ideal solution. Conceptually, the least satisfactory link in the whole process is perhaps the implicit assumption that the product of head offices and ancillary units can be measured by the wages and salaries that they pay, plus the net rent they receive in the case of head offices. As this might be thought to underestimate the product of head offices in particular, it could have some effect in underestimating the product of any regions (whether States or metropolitan areas) where head offices tend to be

(2) Rent received is collected only for enterprises as a whole, and is allocated to head offices.

concentrated. However, any judgement about the gross product generated by head offices and other ancillary units must rely heavily on some convention.

13. Until such time as the results of the economic censuses are available and studies are made of the relationship between establishment and enterprise estimates, the precise proportion of the whole economy covered by the census industries in terms of gross product will not be known. On the basis of currently available estimates by industry, which as indicated above make use of income tax and other statistics of enterprises, it is expected that the coverage will be about 50 per cent or more. Estimates for agriculture are based on establishment returns readily classifiable by State even though this sector is not regarded as part of the integrated census programme, and the problem of the multi-State enterprise is not of great importance in agriculture. Ownership of dwellings, wages of government employees (see below), and wages and incomes of unincorporated businesses in most other industries present no serious conceptual or practical problems. What remains to be estimated are the incomes of companies and Commonwealth Government enterprises in construction, transport, communication, finance, and various service industries. These in total should not account for more than about 3 or 4 per cent of gross domestic product. It will be necessary to rely on fairly arbitrary assumptions to allocate some of these incomes by State, after solving at the national level the more general problem of how to incorporate the new census data into the present data base and estimation system. In the case of the transport and communications sector there are, of course, difficult conceptual and practical problems of determining where these industries are located and allocating their transactions accordingly between regional product and expenditure and external transactions.

Gross Expenditure by States

14. At the present time the Bureau publishes estimates of personal consumption expenditure by States. These are, perhaps, best regarded as a dissection of the national figures, and they represent expenditure taking place in the State rather than direct estimates of expenditure by residents of the State. For example, retail sales figures are used to allocate much of the expenditure, so that conceptually some adjustments are necessary to correct for shopping by persons (including mail order shopping) in States other than the State of residence, but the errors are probably not very significant. Slightly more important, perhaps, is the fact that no adjustments are made for interstate tourism. It is difficult to see what could be done about these factors at any justifiable cost and any process of estimating expenditure on the basis of income (as is done with a few items) seems likely to assume away just the problems that the estimates might be expected to throw light on.

15. Over the years there have been frequent demands for estimates of private capital expenditure by States and these estimates would certainly be desirable for the study of regional growth and development. However,

10(b) - AUSTRALIA

much of the available information is based on enterprises and it has not been possible to link it satisfactorily to the limited establishment data available. Consequently, it has not been possible to compile fully satisfactory figures for areas smaller than Australia. Here again it is hoped to use the economic censuses' collections to provide additional regional statistics but the difficulties are rather more formidable than in respect of the product side of the estimates. The problems of using census data for some industries in conjunction with tax or other enterprise data for other industries will not be easy. Without developing new collections for the industries not covered by the economic censuses, there is no adequate basis for distributing capital expenditure of multi-establishment enterprises to industries and States. In the case of expenditure on assets to be rented there are conceptual as well as practical difficulties. It is not clear whether expenditure should be recorded in the region where the assets are located or where the ownership and/or management is assumed to be located. It may be difficult to get consistent reporting on a location basis if this is the preferred alternative. As for other items in the accounts, in the case of capital expenditure there is the general problem of reconciling economic census data with data from other sources. Subject to that, for industries within the scope of the economic censuses details of capital expenditure are collected on an establishment basis which in turn are reconciled with the enterprise returns. Transfers of capital equipment between establishments are reported on establishment returns as though they were purchases and sales. Buildings for letting are reported only on the enterprise return, but other equipment for leasing is distributed as far as possible to establishments. For these sectors therefore, within the ownership concept at the enterprise level, information will be available on equipment used directly by the enterprise and classified by establishment and therefore by States. Information on equipment for rental (other than buildings) is distributed to the extent that there is an operating establishment with which it can be identified. However, for other sectors, unless special new collections are developed, very little information will be available on how enterprise data for multi-establishment enterprises should be distributed by States.

16. Rented equipment poses a special problem. Applying the System of National Accounts (SNA) principles to regional accounts, non-transportable assets (buildings, etc.) should be classified to the region in which they are located - if necessary, imputing an economic agent resident in that region. In the case of transportable equipment, it should be allocated to the controlling establishment. For the former it is often difficult to obtain consistent reporting of expenditure and rent receipts on a location basis, whereas for the latter the area of location may be a more useful basis than place of ownership or management. The differences might not be very important for State estimates but would become more important in any attempt to obtain data for regions smaller than States. Finally, there is the problem of capital equipment which is readily moveable from State to State,

and specifically transport equipment. It is difficult to ensure consistency between allocations of product and capital expenditure as an interstate transport operator (e.g. an airline) may record all its purchases of vehicles in the State of the head office yet operate them with staff based in other States. Some working rules are necessary which are consistent with the way in which their operations are conducted regionally.

17. The Bureau has recently published figures for individual States of expenditure by State and local government bodies. This component of expenditure presents little difficulty unless attempts are made to obtain figures for regions smaller than a State. Central (Commonwealth) Government expenditure, however, presents both conceptual and practical difficulties in a federal system. A strong case might be made on conceptual grounds for treating the Commonwealth Government in the way suggested in the SMA for international bodies, and the West Australian study to which reference has already been made goes some way towards treating the Commonwealth Government as non-resident. This would mean that the wages paid are treated as part of regional income but not of regional (domestic) product, and other Commonwealth purchases would be treated as exports from the several regions. This would appear to be a somewhat esoteric and not very useful result, but "useful" is the question begging term. To illustrate, when pressed by a State government for information on the State distribution of other Commonwealth expenditure, it was not clear whether the State was concerned with the idea of the Commonwealth as a customer for its local industries or with the question of projects within the State, and it was not possible to get any real clarification of whether they were interested in payments in the State, purchases in the State, or purchases for use in the State. The second interest is appropriate to the non-residential concept above, whereas the last, to which payments in the State may or may not be a fair approximation, calls for a system in which the Commonwealth Government (and its product and expenditure accounts) are broken up into several units resident in the various regions. Under such a system, which appears to the Bureau to be probably the most generally useful, expenditure on wages is quite easily and properly attributed to the State where employees are located and they form part of the product of the region. But even under this system the concept of purchases for use in the State will leave some arbitrary decisions to be made, notably in relation to home and overseas purchases of defence equipment, including ships and aircraft.

Personal Income and Expenditure

18. There have been requests from time to time from planning bodies, market research organisations, etc. for statistics of personal income and expenditure on a small area basis. It may be feasible to provide this information for personal income. The estimation of personal expenditure by States would, however, be subject to the same difficulties discussed in the preceding paragraph. So far as personal income is concerned, two possibilities suggest themselves - the inclusion of questions on income in the

10(b) - AUSTRALIA

ion census schedule and a more detailed area dissection of taxation
ics. (Conceivably a sample survey might also be used but the size of
ple needed could make this impracticable if the number of regions was
numerous.) There has, however, been a firm decision not to include
ome questions on the 1971 population census schedule. The use of
n statistics presents formidable problems of coding residential
es, although this may become more feasible with the further develop-
computer operations over the next few years. The other important
lty in the use of taxation statistics is that it will be necessary to
timates independently for items which are excluded entirely from tax
ics or which tend to be seriously under-reported.

Financial Accounts

At this stage it is not intended to construct a regional system
nts beyond that of the product and expenditure accounts and personal
and outlay. It is considered that the compilation of a complete set
cial accounts would be impracticable at this time, especially in
problems associated with transactions of the Commonwealth Government
y important companies which cross regional boundaries. This is not
that selected transactions cannot be classified quite accurately and
y on a regional basis, but there are enough important items (such as
part of receipts of company tax and loan receipts from institutions)
annot be so classified to make the exercise of doubtful validity.

APPROACHES TO REGIONAL ECONOMIC ACCOUNTING IN CANADAIntroduction

1. The definition of regional economic accounts as a "set" of statistics implies that it should be capable of some analysis of its component parts and amenable to an examination of some of its details, but does not necessarily denote a set of balancing and/or double-entry type books. The reference to a national accounting is not meant to prejudge the technical issue as to whether regional accounts should conceptually be on a "domestic" or "national" basis.
2. Since Canada is a federation of provinces which have significant economic responsibilities these areas have been selected as the basis for the discussion and the main emphasis will be on provincial accounts. The term regional accounts will henceforth refer to accounts for sub-provincial areas.

Demands for Provincial Accounts and DBS Response

3. Although the need for indicators of provincial production -- and its disposition -- has been expressed in many quarters, the most articulate requests came from provincial governments and found normal expression at a series of federal-provincial conferences on economic statistics. An outline of the reasons why the D.B.S. has so far found it impractical to actually construct provincial accounts may be of some value.
4. The absence of a conceptual framework within which one can construct some of the components such as corporation profits, and the lack of data in such vital areas as provincial exports and imports are examples of past and to some extent continuing obstacles to the creation of sound provincial accounts.
5. For multi-regional companies, it is difficult to determine the economic region of origin for profits and other company derived statistics, such as interest and depreciation, other than by some arbitrary method such as of the location of the head office, the distribution of salaries and wages, sales, employment or some judgementally or empirically derived combination of these factors. The development of establishment (i.e. sub-company) type information in the past has been such that serious consideration of either allocating company type financial statistics on the basis of -- or deriving financial statistics directly from -- establishment based surveys was very difficult. Lack of data, rather than conceptual doubt impeded work on the expenditure side. No information on the provincial breakdown of investment in inventories, federal government expenditure and consumer expenditure exists.
6. There also exist conceptual and data difficulties with respect to the provincial allocation of a national at market prices concept of production. Similar difficulties with respect to the industrial allocation of these concepts have been encountered in Canada in the construction of country-wide input-output tables and in the real industry of origin type approach to national accounts. Since the problem in the latter two systems has been solved by opting for a domestic at factor-cost concept of production, the adoption of a similar concept i.e. domestic production at factor cost for provincial accounts will probably prove viable.

7. As a matter of fact, in the late fifties the Bureau undertook a provincial estimation of the income side of the income and expenditure accounts for the year 1957 with the then available data and on a domestic and factor cost basis. The estimates were never released as they were not judged to be of sufficient reliability, however, a paper outlining the methodology was distributed to the provinces.
8. Attempts to structure provincial accounts along the traditional income expenditure accounts have, however, been questioned in a much more fundamental manner. Problems of monetary and fiscal management, the external relation of the currency and the avoidance of the ups and downs of the business cycle have been and are in the main the responsibility of the national government. One of the chief - if not the primary - objective of the income expenditure accounts has been to furnish a meaningful framework within which decisions on these issues could be made. At the sub-national level, on the other hand, "it is clear that the major regional problems are different from those posed at the national level, and it does not necessarily follow that they require a complete set of national income and product accounts as their basic framework of economic information."⁽¹⁾
9. Even if there had been no basic conceptual problems and data for provincial accounts had been more readily available, the immensity of a properly executed job appeared in the past simply overwhelming. Such an undertaking would involve compiling every year ten provincial aggregates with all their appropriate details, making them consistent with all other provincial data, integrating them with national totals, and analyzing them in terms of national price and volume movement. The possibility of being faced with provincial demands for the latter dimension as well as eventually quarterly data, must also be taken into account before one commits resources to such a project.
10. It can indeed still be argued that the resources of a central statistical agency would be more efficiently employed in developing a regional potential or data base than to undertake such estimates itself. Regional analysts could then build up their own macro-economic models from this base. This would give a much greater degree of flexibility to the construction of provincial accounts and they could be much more readily related to regional analytical and policy needs. Further, the builders and users of accounts would thus be the same people -- or at least work in close proximity to each other -- and the assumptions which must inevitably be made at the detailed working level could be more clearly recognized. Analysis could thus be based on an intimate knowledge of the limitation of the data and would presumably therefore be more reliable. Furthermore, as this approach would spread statistical expertise more widely over the country and result in more experimental solutions to given problems, it can be argued to have more long run potential benefits than a centralized construction possesses.
11. As a matter of fact, though it is recognized that a central statistical agency should be ready to set standards and provide technical and coordinating assistance and make available data, it is partly because of the factors outlined above that there has been no really serious demand on D.B.S. to undertake regional input-output studies for specific regions. There is, however, another aspect which differentiates provincial input-output models from other macro-economic aggregates. In the main such provincial input-output

(1) R. and N.D. Ruggles "Regional Breakdown of National Economic Accounts", Design of Regional Accounts, W. Hochwald, ed; John Hopkins Press, Baltimore, 1961 p. 132.

models as exist in Canada have been built for specific analytical purposes and no serious proposals have been made to use them for inter-regional comparisons or to construct them in such a way that a national table can be derived.

12. Although none of the considerations described so far have lost their validity, development at all levels of government and in the political, economic and statistical areas have illuminated the problems from different angles and involved DBS in an ongoing process of reassessment.

Recent Developments

13. The needs and demands for provincial accounts vary from the relatively unsophisticated desire for one figure -- be it production, demand or income -- against which the overall economic performance of the province could be compared to detailed information to be incorporated into advanced econometric models built by provincial experts. In terms of people and resources involved, development in the provinces is uneven.

14. It is obvious that the different needs of all the regionally-oriented analysts, planners and policymakers as well as variation in available economic expertise prevent the emergence of a consensus on even the broad specification or purposes of provincial economic accounts. However, it is possible to describe in general terms the types of analyses for which they might be used.

15. Problems of analysis in growth and development are probably of most concern to the provinces. The analysis of such problems fractures itself on closer examination in many directions. Under the heading of ongoing growth and development, provincial authorities want to discover not only what has been the past performance of the region as a whole but also its causes. Depending on the type of analysis pursued and the prevailing economic and legal constraints, policies designed to influence such growth can be directed towards either the production or demand side.

16. Another problem area is that of interregional income disparities. The Economic Council of Canada has been very much concerned here and has suggested "two main, interrelated considerations in moving towards a better regional balance. The first is the importance of reducing the relative disparities in average levels of income as they presently exist among the regions The second consideration is the need to assure that each region contributes to total national output on the basis of the fullest and most efficient use of the human and natural resources available to the region."(2) While output data are obviously needed to quantify the issues raised in the last half of the quotation, personal provincial income has been found adequate for a quantitative look at the first problem.(3)

17. While there exists a long history of federal government activities designed to correct regional economic imbalance, these have in the past mainly consisted of ad hoc measures with little attempt to design such policies in an integrated manner. The federal government established the Department of Regional Economic Expansion which may now give a federal dimension to the demand for some form of provincial economic accounts.

(2) Economic Council of Canada, Second Annual Review, Ottawa, 1965 pp. 98, 99.

(3) See S.E. Chernick, Interregional Disparities in Income, Staff Study No. 14, Economic Council of Canada, Ottawa 1966.

18. There exists some possibility that the federal department will require not only value data but regional volume and price data as well. The absence until now of equally strong pressures by the provinces for provincial price analysis is probably caused not only by their realization that such price data might be more difficult to obtain than value data, but also their basic inability to influence the course of prices. Furthermore, while the provinces' interest has in the main been restricted to annual data, it is not too unlikely that demand for timely sub-annual regional data may also emanate from the current federal approach towards regional problems.

19. There have been some efforts by the provincial authorities to construct their own provincial accounts which clearly demonstrate that the desire for such provincial accounts is strong enough to have prompted the provinces to invest their own resources into such projects. By now there exists, therefore, in some provinces, an analytical and policy formulation capacity which with the passage of time can only grow and improve. At the same time the federal demand for regional statistics and their increasingly sophisticated use is growing. In this changing climate it has been felt at DBS that while it is still too early to decide to undertake the actual construction of one or more sets of provincial accounts, it has become incumbent on the Bureau to speed up its development of the regional data base from which reliable and good quality accounts could be constructed either by federal or provincial agencies, and tentative priorities have been set for the course of this development.

20. It may also be noted that from a strictly statistics management point of view, the development of data within a given framework -- even without any definite commitments as to the actual implementation of the complete framework permits a smoother, more integrated and more coordinated program than ad hoc and piecemeal extensions.

21. The fact that DBS establishment statistics are now fairly well developed makes it not unlikely that provincial establishment-based data for macro-economic aggregates will be constructed first. Combined with this is the suitability of establishment-based data for the type of analysis which many provinces want to carry on. Furthermore, location specificity of such data give a capability of sub-regional aggregation. The already existing program for the further extension of establishment-type data by DBS briefly dealt with below, reinforces these arguments. The further development of regional establishment data would render feasible the construction of reliable value added as well as input-output type of accounts. It might eventually make it possible to give such accounts a price and volume facet as well.

22. It cannot be gainsaid that the economic analysis of a region which does not consider the demand side is fully adequate. The fact remains, however, that sub-national governments are in many cases not able to have too significant an influence on the demand side, either in terms of their own goods demanded by outsiders or the goods demanded by their own residents from outside. This is even more evident in the small area type of analysis. Nevertheless, a reasonably strong desire for data on the demand side is also present in Canada. One of the main reasons for this stems from the fact that taxes at the sub-national level are quite often aimed at final demand and provincial taxing authorities are therefore understandably interested in viewing their tax base and potential within an integrated macro-economic framework. Provincial planners, furthermore, are interested in the export-import aspect of the demand side. The difficulties of constructing a framework and filling its cells for a sub-national balance of payments are too well-known to require elaboration.

23. Since there exist institutional and constitutional constraints which seem to circumscribe the potential economic impact of regional demand oriented policies, and since it is statistically more difficult to construct a base for a provincial gross expenditure framework, its development has under present circumstances been given a lower priority than that of the establishment type data base.

24. These conclusions are reinforced by the existing plans for the development of statistics which now exist in the DBS and which have been drawn up after a very careful process of weighting and considering all the demands made on the Bureau. The realization of these plans will constitute a significant strengthening of the Bureau's data base. If the regional dimension is given proper attention -- and although this will involve some new data collection it appears to be to a substantial extent a question of appropriate coding and tabulation of available data -- the present data short-falls will, one hopes, largely disappear in the future. In areas where the major difficulty is the substantial data deficiency (for example service industries, construction, etc.) programs are being formulated to collect new data.(4)

25. Like many other statistical agencies in the world, DBS is of course making increasing use of more and more complex electronic data processing installations and as such installations make possible the progressive automation of the entire data collection activity, one can expect that in the future more easily retrievable regional characteristics can be incorporated into the data base.

26. It is evident, in summary that while the DBS is planning to broaden and deepen its regional data base, it is not in a position to undertake the construction of provincial accounts. DBS has also not reached any firm conclusions as to the type of provincial accounts which would be most useful analytically, although there exist preferences, prompted by feasibility, towards an establishment based data system.

27. A summary outline of the conceptual problems to be resolved might be:

- (i) the allocation of head office expenses;
- (ii) the allocation of overall company expenses in the case of multi-establishment companies;
- (iii) the determination of income originating from such special industries as transportation,
- (iv) there may also be problems of determining income originating in (a) government, (b) communication, (c) and finance. It should be pointed out that the establishment is thought of as the most feasible general unit from which to obtain regional and industrial information. In specific industries or regions or for certain specific information items, the establishment may prove not to be the most suitable unit of collection. In such instances one should approximate the establishment concept as closely as possible.

28. Appearances perhaps to the contrary such a course of action which emphasizes the broadening of the data base without assuming responsibility for construction of specific aggregate accounts, has at this time several very significant advantages:

(4) Dominion Bureau of Statistics, Brief to the Special Committee on Science Policy, Senate of Canada; Ottawa, 1968 pp. 80, 81.

- (a) It permits further exploration by the provinces into the kind of system which they would eventually like to see and from which they will in the long run derive the most analytical and policy formulation benefits.
- (b) It provides time both to the provinces and DBS to build a competent statistical and analytical staff in this area and there are also some indications that the application particularly of econometric methodology may have some significant pay-offs.
- (c) The development of a broad regional data base without commitment to any specific provincial accounting system prevents such development from being locked into a precise format at too early a stage. On the other hand, the expansion of regional statistics will benefit from the coherence and integrative influence of the system of national accounts.
- (d) In due time it may also be possible to remove the obstacles to the development of a provincial system of factor incomes. As more data become known and articulated, establishment and legal entity type may become reconcilable one with the other. At that point, a rational allocation of company type forms of income to establishments -- and hence regions -- could be attempted.

Conclusion

29. It appears that at this time no insuperable conceptual problems exist for some types of provincial accounts or at least none which are not amenable to any worse conventional circumventions than are practised in other areas of the national accounts. In view of the improvements which are planned for the national data system as a whole, the regional information for Canada is bound to improve over the next few years and emphasis is being placed in the DBS on the development of a regional data base rather than on the advocacy of any particular provincial accounts system.

SYSTEMS OF REGIONAL ACCOUNTS

1. In a large country like India with a federal set-up of polity, it is of interest to compare the progress in the economic and social fields made by one constituent region with that of another. The Republic of India consists of 17 States and 11 Union Territories. The States/Union Territories are further divided for administrative convenience into districts, tehsils, towns, villages etc. In India the 'region' for the estimation of income and other aggregates is taken as co-terminous with various administrative divisions of the country. At present, however, the need for regional income data has been felt for the States/Union Territories and districts and not for administrative sub-divisions lower than the districts. Again, as the inter-district economic and social transactions are more difficult to measure than inter-State transactions, it is recognised that the problem of regional accounting would need to be first resolved at the State level before considering the district level estimates. In view of this, the estimation of income has been largely restricted to the State/Union Territory level although certain attempts have been made by some official and non-official agencies to work out the estimates at the district level. Some studies have also been carried out in the field of consumption expenditure and capital formation at the State level. It may be mentioned that there is no conceptual accounting framework at the regional level in India. Efforts would have to be made to evolve a suitable framework considering the availability of data at the regional level.

STATE DOMESTIC PRODUCT (SDP)

2. The estimates of State Income could be built up either by the 'accruing' approach or the 'originating' approach. The estimates of State income accruing and state income originating are both essential for a comprehensive insight into the economy of a State. The State income originating may be considered as the income originating from the productive enterprises situated within the boundaries of a State and it corresponds to the domestic product concept at the national level. The state income accruing may be defined as the income accruing to the factors of production supplied by the normal residents of a State which corresponds to the concept of national product at the country level. The difference between the two arises on account of the flow of factor incomes across the boundaries of the States. Such flows of factor incomes are much more significant at the State level than at the national level. It is also much more difficult to measure them at the State level as the economy of the States is essentially open and there are no administrative arrangements such as the exchange control or the customs

department which enable the measurement of such flows at the national level. In view of these difficulties the emphasis in India has been on compilation of estimates of State Domestic Product (SDP) by the income originating concept.

Work done so far

3. The earliest estimate of state income by Slater appeared in Madras Year Book 1923. However, the details of the method of estimation are not known. The earliest effort at estimating regional income was that of Natarajan⁽¹⁾ who allocated the national income for the years 1938-39 and 1949-50 by three broad sectors amongst the then British Provinces on the basis of the sectorwise indices specially constructed for the purpose. The estimates of national income used by him were based on Rao's⁽²⁾ estimates for the year 1931-32 which were carried forward to 1938-39 and 1949-50 with the help of indicators of physical growth in the three broad sectors. Price effect was superimposed by the use of index numbers of wholesale prices published by the Office of the Economic Adviser to the Government of India. The next was the study by Tiwari⁽³⁾ presenting the estimates for the years 1921-22, 1931-32 and 1938-39 for the state of United Provinces both at current and constant prices. He collected and utilised a large amount of local data on output, income per worker etc. Production approach was used in the sectors of agriculture, livestock, fishing and hunting, forestry and mines while for the remaining sectors the income approach was adopted. The two alternative procedures - one of allocation of national totals among States and the other of direct estimation - followed by these studies were typical of the methods used in

-
- (1) Natarajan, B. An Essay on National Income and Expenditure in India, 1949.
 - (2) Rao, V. K. R. V., National Income of British India, 1931-32 (1940)
 - (3) Tiwari, S. G. - Economic Prosperity of the United Provinces, 1951.

later⁽⁴⁾ attempts in the field by official and non-official agencies and research workers. The estimates for the States of erstwhile Bombay, Bihar and Uttar Pradesh were released by the respective State Statistical Bureaus (SSBs) around the year 1950. In the meantime the Working Group on State Income was formed by the Central Statistical Organisation which stimulated greater interest in the subject. By 1958, as many as 8 SSBs viz. Assam, Bihar, erstwhile Bombay, Himachal Pradesh, Madhya Pradesh, Punjab, Uttar Pradesh and West Bengal had undertaken the work of the estimation of State income. Presently all the States (excepting Nagaland and some Union Territories) are regularly publishing the estimates. Annexure shows the position regarding the availability of State official series and also indicates the base year in case of each State for the compilation of estimates at constant prices. Some of the SSBs have recently revised their earlier series by broadly following the Central Statistical Organisation (CSO) as per the methodology given in its brochure on revised series of national product⁽⁵⁾. The data-base and the methodology adopted for deriving the official estimates vary from State to State. The base year for the estimates at constant prices are also different. As a result, the estimates currently published by the SSBs are not strictly comparable. The need for having a series of comparable estimates of State Domestic Product has gained considerable recognition in recent years because a part of the allocation of resources from the Central Government to the States is being based on the relative per capita incomes of the States.

- (4) (i) B. G. Kunte:- Economic Prosperity of Bombay Province and Sind, 1919-1939 (ii) Mohinder D. Choudhry & Bert F. Hoselitt: State Income of Delhi State 1951-52 and 1955-56 in Economic Development and Cultural Change, April, 1963 (iii) Mohinder D. Choudhry: Regional Income Accounting in an Underdeveloped Economy, 1966 (iv) K. N. Raj: Some Features of Economic Growth of the last Decade 1951-61 in India. Economic Weekly, February, 1961 (v) Ravi Verma: Estimation of State Income by allocation method. Third Indian Conference on Research in National Income, 1961 (vi) National Council of Applied Economic Research: Estimates of State Income for 1950-51, 1955-56 and 1960-61 and Distribution of Income by States 1960-61 (vii) Institute of Public Opinion: The Character of Regional Changes in Income: 1955-56 to 1960-61. Quarterly Economic Report, 25th Issue (viii) R. Mehrotra: Social Accounting and Regional Incomes in India, 1966.
- (5) Brochure on Revised Series of National Product for 1960-61 to 1964-65, CSO, 1967.

To meet the requirements of the policy makers, the CSO prepared in the year 1968 comparable estimates of SDP of various States and made the same available to the Planning Commission, Finance Commission, etc., for their use. The comparable estimates prepared by the CSO are discussed later in paragraphs 14 to 16.

DISTRICT INCOME

4. In recent times the demand for the district income estimates as a measure of the level of development of such a region has gained considerable importance. The difficulties of preparing income estimates by the accruing method at the district level are all the more greater. Apart from the non-availability of basic data at the district level, the 'originating' concept is also relatively much more difficult in the case of the district than for the State in respect of supra-regional sectors, covering the boundaries of more than one region, like Railways, Post and Telegraphs etc. Despite these difficulties, some SSBs viz. West Bengal, Gujarat, Madhya Pradesh, Rajasthan, Punjab, Andhra Pradesh and Assam have attempted district income estimates by the income 'originating' concept although all the reports are not yet published. The National Council of Applied Economic Research⁽⁶⁾ have also published district income estimates for 14 States of the Indian Union. All these estimates are rather imprecise. Their usefulness lies in the fact that they help in focussing attention on the gaps in statistics at the district level. If district estimates have to be meaningful, special and consistent efforts would have to be made to collect the required information at the district level.

CONSUMPTION EXPENDITURE

5. The National Sample Survey (NSS) has been collecting data on household consumption expenditure from its very first round⁽⁷⁾ Estimates are available by rural and urban areas for all the States from all rounds except the first two in which data on only rural households were collected. The estimates of household consumption can thus be built up at the State level by the use of NSS data although such estimates may be subject to larger error as compared to the estimates at the

(6) National Council of Applied Economic Research: Inter-district and Inter-State income differentials- 1955-56 (1963)

(7) The first round of the NSS covered the period October, 1950 to March, 1951. From the fourteenth round (July 1958-June 1959) each round of NSS covers one full year.

all-India level due to the small sample size at the State level. As regards government consumption expenditure, the CSO, as a part of their work on the preparation of 'Accounts of the Public Sector', compile information on this item. A number of SSBs have also brought out 'Economic Classification of the State Government Budget'. These documents provide data on the consumption expenditure by the State Governments. As in the case of all-India estimates compiled by the CSO, the SSBs estimate only (i) compensation of employees, and (ii) net purchase of commodities and services - the break up of the latter item into a detailed commodity classification has not so far been attempted.

CAPITAL FORMATION

6. The first notable estimates of capital formation by States were prepared by Mehrotra⁽⁸⁾ for the year 1958-59 by a combination of both direct and allocation procedures. The estimates of capital formation based on local data have been compiled by the SSBs of Assam and Maharashtra.

7. The Maharashtra SSB⁽⁹⁾ prepared the estimates of capital formation for the years 1961-62 and 1962-63 by important industry groups as well as by the public and private sectors. In the case of public sector, use was made of the budget documents of the State Government and local bodies while data on railways and posts and telegraphs were obtained from the concerned agencies. In the case of private sector, the estimates were built up on the basis of Reserve Bank of India's 'All India Rural Debt and Investment Survey 1961-62' (AIRDIS), NSS estimation for the urban households for the year 1961-62, annual survey of cottage and small industries conducted by the SSBs and Annual Survey of Industries (ASI) etc.

8. The estimates of the Assam SSB⁽¹⁰⁾ were also compiled separately for the public and the private sectors and cover the period 1950-51 to 1964-65. In the case of private sector, the estimates for rural households were built up from the AIRDIS and the estimates for other years obtained

(8) Mehrotra R.: Social Accounting and Regional Incomes in India, 1966.

(9) "Capital formation in Maharashtra", SSB Maharashtra 1966 (unpublished).

(10) "Estimates of Capital Formation in Assam, 1950-51 to 1964-65", issued by the Department of Economics and Statistics, Assam, December, 1968.

on the basis of suitable indicators. The estimates of capital formation in construction in urban areas were based on the 'Survey of Urban House Building Activity 1956-57' conducted by the SSB and use of index numbers of value of import of cement into Assam for the use of the private sector. For manufacturing industries, the estimates were built up on the basis of the ASI. For want of data, the estimates in respect of non-household small scale establishments were not compiled. The estimates for the State Government (administration) were based on the 'Economic Classification of the State Government Budget', while data on construction expenditure of the Central Government (administration) were obtained from the Accountant General, Assam and Nagaland. In the case of railways and posts and telegraphs, the requisite data were obtained directly from the concerned agencies while the estimates of departmental commercial enterprises of the State Government were culled out from their annual accounts. The estimates of change in stocks in the private sector were restricted to the factory establishments, mining and commerce. The SSB presented the estimates both for gross and net capital formation at current prices only.

Coordination of State income work

9. The CSO in cooperation with the SSBs has played an important role in the attempts made towards evolution of State income estimates on a comparable basis. Apart from being the central agency responsible for preparing and releasing the estimates of income and other related aggregates at the all India level, the activities of the CSO in the field of regional accounting also cover (i) provision of technical and secretarial assistance to the Working Group on State Income (WGSII), (ii) conducting on-the-spot studies at the headquarters of the SSBs by sending central teams for evaluation of the available data and the methodology and suggesting measures for improvement, and (iii) provision for training facilities in the field of national/regional accounting for the technical personnel of the SSBs.

Working Group on State Income and on-the-spot studies in the SSBs

10. The WGSII was constituted in 1957 consisting of active workers in the field of some in the SSBs, the CSO and other interested agencies with the following functions:

- (i) examination of methods of estimation and concepts
- (ii) investigation of ways and means to fill up the gaps regarding some of the sectors
- (iii) improvement in coverage and quality of information
- (iv) improvement in comparability of estimates of different States and

- (v) technical scrutiny of the estimates to the extent possible.

The Working Group examined the various methods in vogue in the SSBs for estimation of State income and suggested various measures for improvement. It also discussed the position regarding the availability of agricultural price statistics in various States and the appropriate price data for valuation of agricultural output. It recommended that in the case of sectors where the income approach was followed in the estimation of state domestic product, periodical assessment of data on employment, per head earnings etc., by the SSBs would be desirable. For this purpose, special studies suitably phased to cover various sectors over a period of years in cooperation with the CSO have been recommended. The CSO has prepared five brochures⁽¹¹⁾ describing the methods adopted by the States. Subsequently teams were sent out to the headquarters of the various SSBs for on-the-spot examination of all aspects of State Income estimation. The five brochures, one each on (i) agriculture, animal husbandry, forestry and fishery, (ii) mining and manufacturing, (iii) electricity, gas, water supply and sanitary services, (iv) other services and (v) trade, storage, hotels and restaurants, explain in detail the basis of preparation of estimates of SDP by various SSBs in these sectors. The brochures are intended to serve three basic purposes. Firstly, they would help the Planning Commission and other policy making bodies in using the estimates prepared by the SSBs with the full knowledge of details and limitations of the estimates. Secondly, they would enable each SSB to appreciate in detail the methodology and data-base used by other SSBs. Thirdly, they would help the CSO and the WGSJ in formulating and laying down standard methodologies in various sectors.

-
- (11) These are
- (i) Methodological Questionnaires and Replies on State Income Estimation: Agricultural Sectors, CSO, April 1964.
 - (ii) Methodological Questionnaires and Replies on State Income Estimation: Mining and Manufacturing Sectors, CSO, Sept., 1964.
 - (iii) Methodological questionnaires and Replies on State Income Estimation: Services Sector, CSO, March 1967.
 - (iv) Methodological Questionnaire and Replies on State Income Estimation: Electricity, Gas, Water Supply and Sanitary Services, CSO, Feb., 1968.
 - (v) Methodological Questionnaire and Replies on State Income Estimation: Trade, Storage, Hotels and Restaurants, CSO, Sept., 1969.

11. Evolution of Standard Methodology: After examining all aspects including the methodology and data used by different SSBs in compiling their estimates of SDP, the WGSII has prescribed standard methodology for various sectors. These sectors correspond in scope and coverage to the revised series of national product brought out by the CSO with suitable further breakdown wherever necessary. The sectors considered are: (i) agriculture and animal husbandry, (ii) forestry and logging, (iii) fishing, (iv) mining and quarrying, (v) large scale manufacturing, (vi) railways, (vii) communication, (viii) banking and insurance, (ix) public administration (Central Government), (x) public administration (State Government and local authorities), (xi) educational services, (xii) medical and health services, (xiii) sanitary services, (xiv) electricity, gas and water supply, (xv) small scale manufacturing, (xvi) construction, (xvii) transport by means other than railways, (xviii) trade, storage, hotels and restaurants, (xix) real estate and ownership of dwellings and (xx) other services.

12. The first fourteen sectors may be classified into three categories:

- (i) Category -I of sectors for which the production-approach was recommended viz. agriculture and animal husbandry, forestry and logging, fishing, mining and quarrying, large scale manufacturing and electricity.
- (ii) Category -II of sectors for which the income approach has been recommended viz. public administration (State Government and local authorities), educational services, medical and health services, sanitary services, gas, and water supply.
- (iii) Category -III consisting of four sectors, viz. railways, communication, banking and insurance and Central Government administration.

In respect of the last category, the estimates at the all-India level were recommended for allocation among States on the basis of standard indicators prescribed for the purpose. The CSO allocates the national totals amongst the States in respect of these sectors. As regards the last six sectors, listed in paragraph 10 above draft standard methodology has been prepared by CSO but final recommendations have yet to be made.

13. In prescribing the standard methodology, the availability of Statewise data and usability of such data on production, prices, inputs, working force, earnings, indicators of temporal and spatial pattern relating to various sectors have been the primary guiding factors. The

10(b) - INDIA

base year recommended for estimation at constant prices was 1960-61. Standard methodology prescribes the choice of the relevant data, lays down the procedures for their use and suggests ways and means to procure the minimum essential data.

14. Training in Regional Accounting: In the Statistical Officers' training courses run by the CSO which are in the nature of in-service training of technical workers of the Central and State Governments, there is a provision for special training in national and regional accounts for the trainees opting for this subject. A need has, however, been felt for instituting an intensive in-service training course in regional accounts specially for statisticians working in the field of State Income in the SSBs. A scheme for this purpose has been processed and a six-month course is expected to start sometime during the fiscal year 1970-71. The emphasis in the course would be on concepts and definitions, techniques of examination of source material, standard methodology etc. It is proposed to devote about 130 hours to theoretical aspects and more than 650 hours for practical/study work. It is hoped that the course would go a long way in meeting the needs of the SSBs for technically trained personnel in the field of State Income accounting.

Comparable estimates prepared by the CSO

15. As already stated the estimates of SDP prepared so far by various SSBs are not comparable because of the differences in methodology and data-base used in the preparation of such estimates. There has, however, been a consistent demand for a comparable set of estimates of per capita SDP with all their limitations which could be used for allocation of certain percentage of resources from the Central Government to the State Governments. The CSO prepared the comparable estimates for all States in association with the State Statistical Bureaus for the period 1962-63 to 1964-65. The estimates in respect of net output from the commodity producing sectors i.e. agriculture, animal husbandry, forestry, fishing, mining and quarrying and large scale manufacturing were supplied by the State Statistical Bureaus whereas estimates for non-commodity producing sectors were worked out by the CSO mostly by the allocation of national total to States by use of appropriate indicators depending upon the availability of the same. The details of the indicators used for allocation for each sector are as follows:

1. Small scale manufacturing: Statewise NSS data for 1960-61
2. Construction : Statewise working force data from 1961 census and consumption of cement
3. Electricity, gas and water supply : Analysis of accounts of

3. Electricity, gas and water supply (contd.) electricity undertakings, employment in gas industry, working force and wages and salaries in water supply.
4. Public Administration : Analysis of budget accounts
5. Transport by other means : Number of motor vehicles, consumption of high speed diesel oil analysis of accounts of port trusts 1961 working force
6. Trade, storage, hotels and restaurants : Income tax statistics 1961 and working force
7. Real estate and ownership of dwellings : Income from ownership of urban dwellings
8. Other services : Working force as per 1961 census

16. The estimates prepared by the CSO are not strictly comparable in the sense that they do not fully take account of the wage differentials. Furthermore, compiled at current prices they also do not take account of the price differentials from State to State. However, taking into account the availability of statistics, this procedure was considered to be the only alternative. In the long run it would be necessary to build up the estimates for all the sectors according to the standard methodology laid down.

AnnexurePosition of State Income Work at the SSBs.

(Estimates of State Domestic Product at Current and Constant Prices and the Base Year Used for Constant Price Estimates)

S. No.	State/Union Territory	Estimates at		Base year for col.(4)
		current prices	constant prices	
(1)	(2)	(3)	(4)	(5)
1.	Andhra Pradesh	P	P	1960-61
2.	Assam	P	P	1948-49
3.	Bihar	P	P	1948-49
4.	Gujarat	P	P	1960-61
5.	Haryana	P	P	1960-61
6.	Jammu & Kashmir	Not preparing	P	1955-56
7.	Kerala	P	P	1960-61
8.	Madhya Pradesh	P	P	1952-53
9.	Maharashtra	P	P	1960-61
10.	Mysore	P	P	1956-57
11.	Nagaland	Not preparing		
12.	Orissa	P	P	1960-61
13.	Punjab	P	P	1952-53
14.	Rajasthan	P	P	1954-55
15.	Tamil Nadu	P	P	1960-61
16.	Uttar Pradesh	P	P	1960-61
17.	West Bengal	P	P	1951-52
18.	Andaman and Nicobar Islands	Not preparing		
19.	Dadra and Nagar Haveli	Not preparing		
20.	Delhi	Not preparing		
21.	Goa, Daman & Diu	Not preparing		
22.	Himachal Pradesh	P	P	1950-51
23.	Laccadive, Minicoy and Aminidive Islands	Not preparing		
24.	Manipur	P	P	1960-61
25.	NEFA	Not preparing		
26.	Pondicherry	Not preparing		
27.	Tripura	P	Not preparing	

Note - P denotes that SSB prepares estimates.

Systems of Regional Accounts

1. In countries like Tanzania, which have embarked upon comprehensive planning for economic development it is essential to build up regional accounts within the framework of national accounts. However, the dearth of reliable and up-to-date statistics relating even to the important fields of primary production acts as the most serious bottleneck in building up regional estimates. Further where the national estimates are based on the results of sample surveys regional estimates tend to be either non-existent or in any case less dependable.
2. For Tanzania, certain ad-hoc breakdown by regions of GDP estimates prior to the recent revisions alone are available. No regional breakdown of the revised series of estimates have been compiled for any of the years. The Bureau of Statistics has plans to build up such estimates at regional level at least for the primary producing industries. Building up capital formation estimates at regional level is almost impossible under the existing availability of data. However, in course of time, as the country gets more and more committed to regional planning the necessity to have regional accounts would be felt more keenly.

SYSTEMS OF REGIONAL ACCOUNTS

1. There has for some time been an interest in separate accounts for the constituent countries of the United Kingdom, and in particular for Scotland and for Wales. More recently, the establishment in 1965 of eight Economic Regions in England, each with its own Regional Economic Planning Board, has created a demand for separate figures for these regions also. The requirements which have been put forward, both for the separate countries and for the Planning Regions, have been varied. At one extreme there have been suggestions for full sets of accounts comparable with those produced for the United Kingdom as a whole, including input/output tables, the analysis of gross domestic product by industry of origin as well as by expenditure components, income and outlay accounts for institutional sectors, etc. At the other extreme there have been more modest demands for data on the contribution made by each country or region to the gross domestic product for the nation. Between these two extremes there have also been suggestions for time series for gross domestic product or certain components such as consumers' expenditure or capital formation, designed to assess the rates of development in particular areas.

2. There has been a considerable increase in the last few years in the number of statistical series in which a regional analysis has been incorporated, but the data available still fall far short of what would be required for the compilation of comprehensive and consistent systems of regional accounts on the lines of those available for the UK as a whole. Nor do the data yet adequately meet the more modest demands mentioned above. Whichever approach to regional estimates of gross domestic product and its components is adopted - via expenditure data, income data or production data - a point is reached where it becomes necessary to allocate some part of the national totals on a more or less arbitrary basis. This is no doubt the case in other countries also but for a small and densely populated country with an intricate communications network such as the United Kingdom the problems tend to be more serious than in a larger country where regions are more self-contained. Because of this need for arbitrary estimation the regional estimates are inevitably subject to appreciable margins of error. It has therefore not been considered worth while to devote much in the way of official resources to this work. Official estimates are, however, published by the Government of Northern Ireland, which is largely autonomous and for which most statistics are collected separately, and a "Scottish Budget" which was in effect a central government account for that country, was published earlier this year. Apart from this, a certain amount of work has been done by research workers in universities and similar institutions. They have produced estimates in some detail both for Scotland and Wales and for the Economic Planning Regions in England. The work has been officially supported to the extent of provision of data, some of it unpublished, and in some cases financial assistance. The results of this unofficial work are useful in giving the approximate magnitude of each country or region's contribution to g.d.p. and the way in which it is made up, but it is unlikely that a series of such estimates purporting to show changes over time for a particular region would be very reliable, even if the work were taken over officially and additional resources put into it.

3. The remaining sections of this paper set out in more detail the data at present available and the problems in using it to produce regional accounts.

Expenditure data

4. A regional distribution is available of the data collected in the continuing Family Expenditure Survey. This can be used to produce a first attempt at an income and outlay account for the personal sector. The sample of households is, however, very small in some regions and certain sections of the population - those living in institutions - are excluded from the sampling frame. Information on expenditure on goods and services by local authorities can, of course, be readily assembled within regions and presents no problem. For central government expenditure, however, it is necessary to adopt some conventional rules. One possibility is to allocate the whole of the expenditure to the South East Region of England, which includes London. Another would be to allocate the total expenditure between regions on a per capita basis. A third would be to allocate expenditure as far as possible to the regions in which it is actually made. For fixed capital information, data are already published showing the regional allocation of public sector expenditure on dwellings and other buildings and works but data are not at present available for other types of assets. There are particular problems in such cases as expenditure on railway rolling stock or vehicles operating in more than one region. For capital formation in the private sector, some data are available on dwellings and on industrial and office building, but these are in physical terms and have to be converted to values on arbitrary assumptions. There are also some data derived from the system of investment grants, though these apply only to certain types of capital formation. Although some attempt at each of these components of gross domestic product can be made, there is a major difficulty over exports and imports. Not only is there no way of allocating overseas trade by regions but also no data are available on the flows of goods and services crossing regional boundaries within the country. These flows are of particular importance in the United Kingdom. Transport statistics yield some estimates of freight tonnages but it is very difficult, if not impossible, to use these to obtain value estimates.

Income data

5. For personal incomes, both from employment and from entrepreneurship, information is available from a regional analysis of income tax data. The basis of allocation, however, is the region in which the income is assessed, so that the problem of inter-regional movements again arises since the region for assessment is normally based on the place of work and may not coincide with the place of residence. Although the boundaries of the Economic Regions are drawn in such a way as to avoid this as far as possible, there are inevitably cases where a city near the boundary of a region draws some of its workers from a neighbouring region. The tax data are not comprehensive in that they exclude incomes below the tax limit and adjustments must therefore be made to allow for this. Nor do the tax figures coincide with the income definitions needed for national accounting purposes, so that the adjustments made in the national figures, for which there is some firm basis, have to be allocated much more hazardingly among regions. The most difficult problem on income data, however, arises over company profits. Most of the large companies in the United Kingdom have establishments in more than one region, but the profits, whether extracted from company accounts or from tax assessments, relate to the company as a whole. For tax purposes the assessment is made in the region in which the head office is allocated - frequently in London, even although the bulk of the activity may be elsewhere. The allocation of these profits between regions can

only be made on some arbitrary basis such as data on the numbers employed in the various regions, but this inevitably adds very much to the margin of error in the regional estimates. Personal incomes from interest, dividends etc. can, like income from employment, be allocated according to the place at which the recipient is assessed to tax but no distinction can be made between incomes received from sources within the same region and those from sources in other regions. Here again, the figures are incomplete since small investment incomes may be omitted because they are below the tax limit or are not fully reported. Nevertheless, the use of this tax data supplemented by data from the Family Expenditure Survey can provide regional estimates of personal income and expenditure, which are probably more reliable than estimates for other parts of the accounts.

Production data

6. In the quinquennial Censuses of Production data are available on the net output, as defined for the purpose of the Census, for each region. This is available because the Census returns are as far as possible on an establishment basis. Net output derived in this way, however, is before deducting certain expenses, particularly those incurred by the head office of large enterprises with a number of separate establishments. Such expenses, to the extent that they are known, can only be allocated between the regions in which the establishments are situated on an arbitrary basis. The Census of Production, however, covers only about half the economic activity of the country. For other industries much less information is available, and for some none at all. The production approach is therefore not a very useful one in this context.

THE GENERAL QUESTION OF CONFIDENTIALITY, ESPECIALLY AS
INVOLVED IN THE USE OF COMPUTERS

Confidentiality together with objectivity and accuracy are the corner-stones of the structure of public confidence that an official statistician must build up in order to do his job properly. Provisions for the protection of confidential information from disclosure are also complementary to the necessary provisions giving an official statistician power to conduct compulsory censuses and surveys. At least in Australia compulsory collection of data would not be possible without the accompanying protection against disclosure provided by the confidentiality provisions of the Census and Statistics Act 1905-1966.

2. The relevant provisions of the Act are wide and simply stated. However, their interpretation involves complex legal and statistical considerations. The Act provides that:

"20.-(1) The Statistician shall compile and tabulate the statistics collected pursuant to this Act and shall publish such statistics or abstracts thereof, as the Minister directs, with observations thereon.

24. The Statistician, an officer, or the occupier of a dwelling, shall not, except as allowed by this Act, divulge the contents of any form filled up, or any information furnished -

- (a) in pursuance of this Act; or
- (b) at the request of the Statistician, for statistical purposes."

The precise effect of the Act then depends on the meaning given to the word 'divulge' and the word 'contents'. In practice these are interpreted to mean that no statistics should be published or otherwise revealed to any person outside the Bureau which would even indirectly represent the operations of any single firm or the personal particulars of any individual person.

3. The present Act makes no distinction between information collected by the Bureau of Census and Statistics and information supplied to it as a by-product of the administrative processes of other agencies. It also applies with equal force to all statistics compiled by the Bureau. By contrast, in some other countries certain statistics (e.g. overseas trade statistics) are exempt from the general confidentiality rules. There is no provision for the Statistician to obtain permission from informants to publish information which would otherwise be confidential. Nor is there any scope for the Statistician to supply such information to other government authorities where these authorities may have a legitimate use for it, even if its use could be of no disadvantage to the supplier of the information. There is even some uncertainty as to the freedom of the Statistician to give back to an administrative department individual particulars supplied to him by that department for statistical purposes.

4. The stringency of these aspects of the Australian legislation creates a number of difficult practical problems, particularly in respect of large finely classified monthly statistical collections like overseas trade statistics. It also places restraints on the use that may be made of registers of establishments and enterprises maintained by the Bureau for statistical purposes. These registers are costly to establish and maintain and could have a variety of legitimate administrative as well as purely statistical uses.

5. In Australia, the collection, compilation and publication of official statistics is highly centralised, most of this work being the responsibility of the Commonwealth Statistician. This has the advantage of enabling consistency to be maintained in policies and practices relating to confidentiality.

6. But legislative and procedural rules may not be enough. It is the feeling of confidence among informants that their business and personal privacy will be protected that is essential if they are to willingly provide the Statistician with the detailed information he needs. And, whatever the legislative provisions for protecting privacy, confidence will depend on the Statistician's judgment in preserving the spirit of the law as well as the letter of it. The letter of the law would permit publication of an industry total if there were three producers, but this would break the spirit of the law if one of the producers were small. The exercise of judgment in this way enables confidentiality policies to be applied flexibly in varying circumstances and to be modified to suit changing community conditions and attitudes. It should be remembered, of course, that any change must be based on a sensitive appraisal of community attitudes and be implemented with great care to avoid any possibility of damage to the public credit of the official statistician.

7. Over a considerable period of time a working policy on confidentiality has been established in Australia within the context of the legal and practical situation described above. The effect of the introduction of computers to the work of the Bureau has been to change the basis on which the existing policies had been established. The new possibilities for statistical service provided by computers and the effect of computers on the community generally and on users' requirements for statistics, including the form in which statistics are provided to them, have created a need for a review of confidentiality policy. In this review it is necessary to return to consideration of first principles to ensure that confidentiality legislation and practices are made consistent with the new technology of data processing and the new service philosophy represented by the data bank approach to statistical services. The following paragraphs consider the objectives of confidentiality policy, the implications of computer processing, the data bank approach to supply of information and its effect on confidentiality policy and procedures, and possible future developments.

Objectives of a Confidentiality Policy

8. The formulation of a policy on confidentiality requires the needs of the community as a whole for statistics to be balanced against the right to privacy of individual persons or firms. It falls therefore into the general class of questions involving the reconciliation of the interests of the community and the individual. The details of the confidentiality policy of any country will therefore be affected by the political, economic and social structure of that country and there can be no standard policy for all countries. There will be common elements in the policies of different countries, however, because the basic purpose of confidentiality policy is the same in all cases, i.e. to obtain access to detailed information that informants would not willingly make generally available. The suggested objectives of confidentiality policy set out below may therefore have fairly general application. The conflict between the two interests referred to above can be observed in this set of objectives, and the achievement of a socially acceptable balance requires the greatest skill and the most sensitive judgments on the part of the statistician.

9. The objectives are arranged in three groups which initially have a degree of overlap:

- (a) those concerning informants and their protection against intrusion into their privacy or disclosure of their business information;
- (b) those concerning statistical offices and their responsibility to operate effectively and efficiently; and
- (c) those concerning users and their need for information.

(a) Policy Objectives Primarily Concerning Informants

- (1) to protect persons fully against the disclosure of their personal particulars to any person or government agency outside of the official statistical office;
- (2) to protect organisations fully against the disclosure of information relating to their operations. The determination of what constitutes full protection is, of course, a problem in its own right, especially in the absence of complete knowledge of ownership and control relationships in private and government enterprises;
- (3) to respect community standards relating to personal and business privacy, e.g. to take account of the fact that even to make observations or to collect or query information is in some degree an invasion of privacy, regardless of protection against disclosure, and that some information is highly sensitive. Thus a particular community at a particular time might consider that some questions are of such a sensitive nature (e.g. religion, personal income) that they should only be asked in a particular way or answered on a voluntary basis, or

(e.g. company profits) that privacy should be respected by restricting the number of officers handling identifiable returns;

- (4) to be responsive to requests by individuals and firms for consideration of particular disclosure problems (e.g. to have their returns identified only by code number to prevent disclosure in transmission or to maintain privacy on receipt); and
- (5) to be equitable in making such arrangements so that no individual or firm is refused special protection which is available to others.

(b) Policy Objectives Primarily Concerning the Statistical Office

- (1) to preserve public confidence in the security of confidential information supplied for statistical purposes so that they will supply genuine data willingly;
- (2) to be adaptable to meet changing circumstances and evolving community standards so as to provide the best service possible within the existing constraints;
- (3) to arrange that as much as possible of the total stock of information obtained by the government for all purposes is fully used for statistical purposes;
- (4) to avoid unnecessary duplication of functions by statistical offices and other administrative agencies, e.g. in the maintenance of business registers;
- (5) to establish procedures simple enough to be capable of implementation without excessive clerical or data processing costs and to be sufficiently specific to obviate the need for continuous attention by senior officers to the interpretation of the rules for individual cases.

(c) Policy Objectives Primarily Concerning Users

- (1) to be flexible in operation so as to minimise the negative effects of confidentiality procedures on published statistics and services - e.g. to avoid the danger of confidentiality policy becoming a convenient rationalisation for failure to meet users' requirements which could be met with sufficient effort;
- (2) to provide for the special needs of government programmes by specific data sharing arrangements. This may conflict with other policy objectives and is not provided for in Australian legislation or practice. It may be argued by government users that there are circumstances where certain individual record data about businesses could be disclosed to other government agencies in the interest of efficient government and with the general consent of informants. It is understood that some statistical legislation permits this, e.g. by provisions for disclosure by consent, by excluding some non-sensitive data from the disclosure restrictions or by permitting transfer of information within the government under certain definite limitations.

The Implications of Computer Processing

10. In Australia, computer processing of official statistics is performed by the Computer Service Centre, which is a division of the Bureau of Census and Statistics. The Computer Service Centre has installations in Canberra and in each of the State capital cities and virtually all computer processing of official statistics is performed by these installations. The Computer Service Centre also provides computer processing services to several other Commonwealth government departments, including the Department of the Treasury and the Commonwealth Taxation Office. It therefore combines in a single organisation and at each computer installation both statistical and administrative data processing. This centralisation of computer processing of official statistics within the Bureau may have assisted in recognition of the implications of computer processing for confidentiality policy and it is hoped will also assist in the development of a new policy and procedures. It may also have created special problems arising from the possible unfounded public belief that organisations sharing a single computer service may also be prone to share confidential information between them.

11. On a broader plane, the effect of computers on the question of confidentiality has been to draw the attention of a wider audience to a problem that was formerly confined largely to discussion in professional statistical circles. It has also revived interest in the discussion within these circles and many national statistical offices are feeling a need to review principles and practices in relation to statistical confidentiality.

12. The use of computers affects confidentiality policy and practice in the following ways:

- (i) Greater attention to confidentiality policies and procedures for implementing them is needed because -
 - (a) Information can now be processed in greater detail and a greater volume and wider range of tabulations can be produced. The possibility of storage and linkage of data on a large scale is also introduced. These developments intensify the need for disclosure analysis and for better procedures for the prevention of accidental disclosure or deliberate attempts to obtain illegal access to information.
 - (b) To some extent already, and no doubt to a much greater extent in the future, statistical tabulations will be replaced by information retrieval systems. Such a system may be conceived of as a data base of detailed information including confidential details and a set of programs for retrieving, aggregating, analysing and tabulating that information. It can be expected to provide greater flexibility and more speedy response than existing methods for the production of printed tabulations. However, new procedures would have to be provided to protect the confidential information in the data base from disclosure and to carry out disclosure analysis on the reports prepared. This disclosure analysis would have to be carried out as each report is compiled if the full potential of the information retrieval system were to be realised. In doing so, procedures would need to be adapted to avoid residual disclosure by the supply of separate reports for intersecting data sets. Although each of these reports would not contain confidential data, when combined they may enable confidential details to be imputed by subtraction.
 - (c) Widespread ignorance and misunderstanding of computers and their capabilities, reaction to change in the form of new computer systems, and resistance by groups which believe they may be disadvantaged by the extension of computer processing may lead to a public reaction against computer based information systems. This reaction, which was seen in the United States in the form of the "data bank" controversy, creates special problems for statistical offices. It may result in restraints on their activities in collecting data or in maintaining integrated data files or registers. It may also inhibit them from exploiting to the fullest possible advantage the potential of computers for a better statistical service and reduced operating costs.

- (ii) Computers provide the means for a more refined and systematic approach to disclosure analysis. The rules adopted to prevent disclosure need no longer be based on rules of thumb or intuitive judgment. They can be based on a set of detailed decision rules laid down in advance for implementation by computer program as tabulations and other reports are prepared.
- (iii) To the extent that human judgment is required to determine disclosure problems that cannot be settled by the routine application of such decision rules, computer processing can provide special reports to highlight sensitive cells in tabulations and to illustrate the results of alternative combinations of cells or classification changes designed to avoid disclosure. An example of this approach is the disclosure analysis procedures built into the computer tabulation system for the 1968-69 Integrated Economic Censuses conducted by the Bureau.
- (iv) The widespread use of computers for commercial data processing, particularly where common-carrier data transmission services or computer service bureaux are used, has raised the problem of the security of confidential business records, to say nothing of military and diplomatic information. As a result, a substantial degree of attention is being given by manufacturers and users of computers, particularly by communications utilities and service bureaux, to the means of securing information held on computer readable media, being processed by a computer or transmitted over a transmission network. There is potential for a substantial by-product of techniques that could be beneficially applied to the problem of confidentiality.

The Data Bank Approach to Supply of Information - Concepts and Terminology

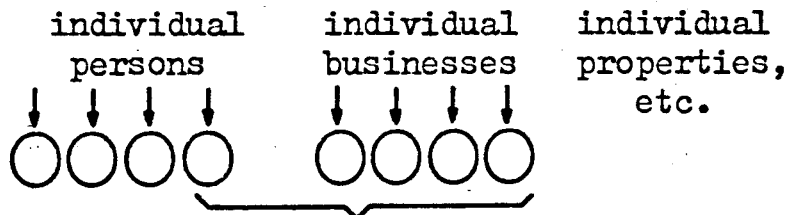
13. The issue of confidentiality arising from data storage and linkage deserves special attention. The following discussion draws on proceedings of two sessions of the International Statistical Institute. Before describing the technical possibilities of the "data bank" or integrated A.D.P. file approach for improving statistical services in Australia, and examining problems of confidentiality that these raise and their possible solutions, it is necessary to reach a common understanding about terminology. Clear and objective discussion can be difficult when some of the key terms such as "data bank" may have very different and sometimes highly emotive connotations for different persons.

14. The following concepts and terms (following particularly Vander Noot and Young¹ and Bachi and Baron²) seem to be the most useful:

- (i) Data file: a set of consistently defined observations (relating to a single subject or with at least one characteristic in common) recorded on a single medium (which can be in machine readable form, e.g. punched cards or magnetic tape or in another form such as a printed list or set of handwritten cards). The set of observations in the data file may relate to a number of different units at a single point of time or to a set of observations over time.
- (ii) Data base: a set of related data files.
- (iii) Data bank: a data storage and retrieval system operating on a data base (i.e. a number of related data files) and constructed with a sufficient degree of standardisation to enable linkage of data between files to be made readily and inexpensively with a minimum of programmer intervention. It allows single items or groups of items of information to be retrieved singly or in combination with any other item or group of items in the bank. Ideally the retrievals and data manipulations are achieved by specifying standard plain language control commands rather than by writing ad hoc programs to deal with specific requests.

15. The following indicates the relationship between the above concepts:

DATA about individual cases exist on more or less unrelated SETS OF FILES

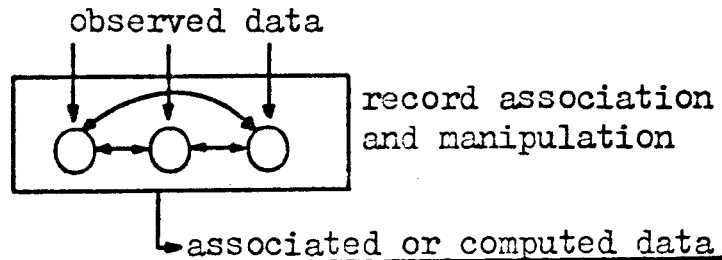


which, when organised into a system to serve a common purpose, may be called the DATA BASE for that purpose

(e.g. data base for administering controls over businesses)

(e.g. data base for statistics of household sector of national accounts)

If any part of the data base is held in a form which permits the ready association of the individual records on different files it may be called a DATA BANK (integrated file system, uniform file system, etc.)

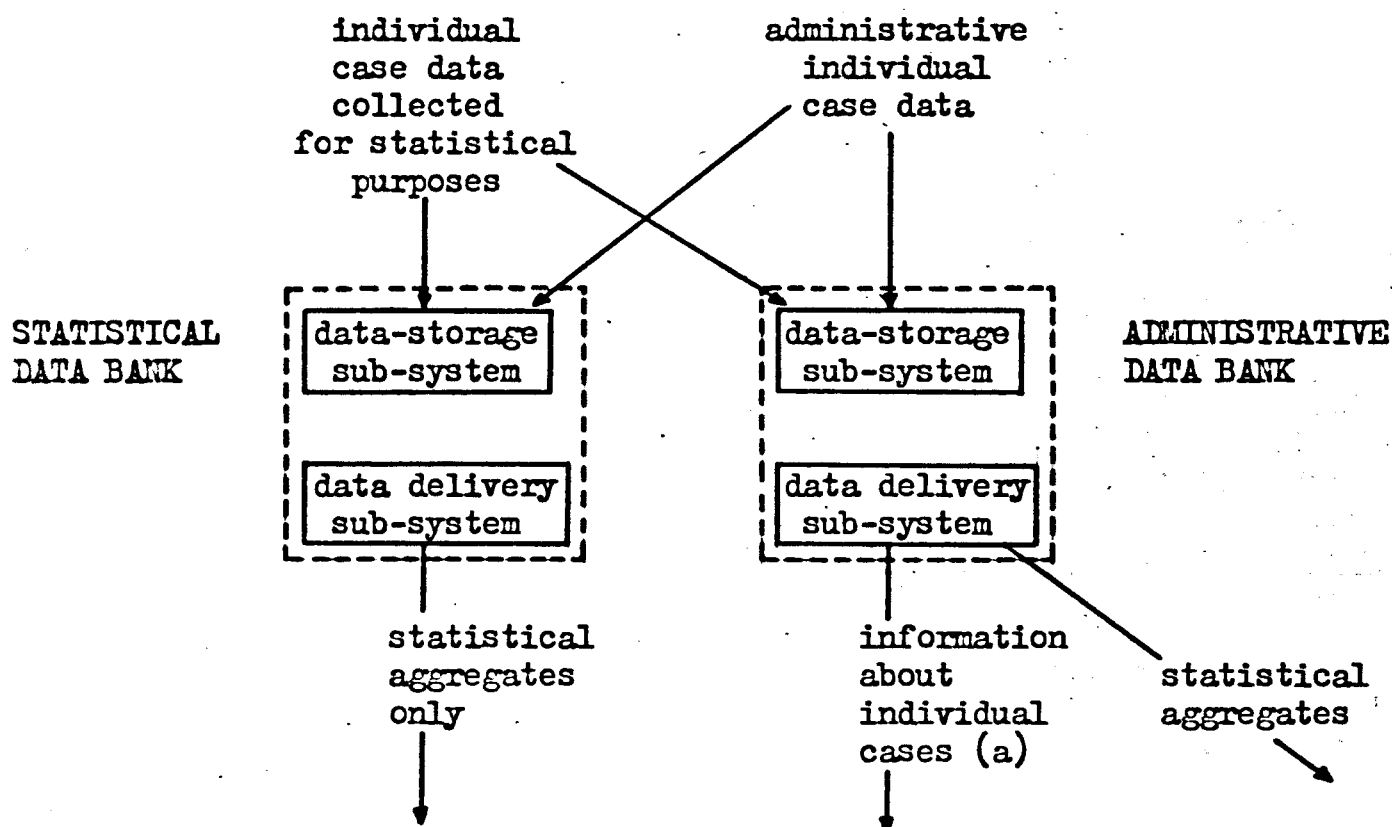


(1) Proceedings of the American Statistical Society 1966
 (2) International Statistical Institute 1969

Distinctions between different classes of data bank are suggested below as being useful in any discussion of the privacy problem.

Statistical Data Banks/Administrative Data Banks

16. Whereas a statistical data bank is a data bank which is maintained only for statistical purposes (i.e. it makes available to users only statistical aggregates rather than individual case data) an administrative data bank makes available individual case data to its users. An administrative data bank may also make available an output of statistics based on the administrative data and/or on data collected directly for statistical purposes. An administrative data bank which has an output of both administrative data and data collected directly for statistical purposes could be described as a hybrid (administrative/statistical) data bank. If the operating rules for a hybrid data bank were changed so that only statistical aggregates could be delivered from it then it would become a statistical data bank. Thus the essential difference between statistical data banks and other forms of data bank is in the bank's data-delivery arrangements. This is illustrated as follows:



In the above illustration the administrative data bank is a hybrid type because of the output (a) of information on individual cases. There could be a flow of statistics or of individual case data from the administrative data bank to the statistical data bank without altering the status of the statistical data bank. Any output of individual-case data from the statistical data bank to the user would change it to a hybrid (administrative/statistical) data bank.

17. Although the distinction between statistical and administrative data banks can be fairly readily drawn in discussion among professional

statisticians, it does not necessarily follow that this distinction will be readily accepted by the public at large, especially as it is based on the uses to which the data bank is put, rather than on the information stored or the structure of the data bank. A gradual programme of public education may therefore be necessary in order to establish a climate of public opinion in which the distinction between statistical data banks and administrative data banks is widely accepted.

18. It can thus be seen that it is participation in a hybrid type data bank that is most likely to compromise a statistical agency's reputation for respecting personal privacy. The pure statistical data bank is unexceptionable in this respect so long as it stays pure and the public have confidence that it is a pure statistical data bank.

19. Integrated file systems which are often called data banks but which contain only aggregated data (albeit perhaps at a very fine level of aggregation) present no problems with regard to disclosure. Only statistical aggregates come out of it since only statistical aggregates go into it (this assumes that the aggregates pass the disclosure tests of having sufficient cases in them). The Time Series Information System being developed in the Bureau (which is similar to CANSIM) and CANSIM itself are examples of such a statistical data bank.

20. There are various kinds of administrative data banks. It is worth noting that it is the intelligence data bank that raises most of the fears against data banks in general. In this kind of data bank the aim is to build up a general dossier of information about the individual as an individual, e.g. a data bank set up to provide information about individuals for use in crime detection or for credit rating. Statistical offices have a stake in encouraging legislative protection and ethical codes relating to all administrative data banks, including an interest in keeping open any one-way flow of information from such banks into their own system. Their most direct concern, however, is with statistical data banks.

Integration of Statistical Files in Australia

21. As the Australian statistical system is highly centralised, the Bureau does not face all the administrative problems involved in the integration of statistical files in countries such as the United States which have decentralised statistical systems.

22. The possibilities of computer processing techniques for the integration of statistical files which have been so well expounded in the Kaysen report⁽¹⁾ have been recognised and, as mentioned earlier, there are no significant organisational barriers to the substantial realisation of these technical possibilities. However, experience in the Bureau has been that the conceptual problems and practical statistical aspects of integrating statistical files are at least equal in importance to the technical data processing aspects.

(1) Executive Office of the President, U.S. Bureau of the Budget, Report of the Task Force on the Storage of and Access to Government Statistics, Washington, D.C., October 1966.

23. A great effort over the past five years has enabled the Bureau to take substantial steps towards the integration of its statistical collections. With this work as a base, the development of an integrated statistical information service is now being approached in which the storage, linkage and retrieval of statistics will be organised so as to take advantage of the degree of integration of collection and compilation already achieved and to make use of the potentialities of computer processing to provide a better statistical service to users. This development is raising some new legal problems in relation to confidentiality and has heightened the importance of some long standing ones. These problems may be considered in three groups:

- (a) those arising in the case of statistics compiled from the administrative records of another department where there are legal provisions governing disclosure of those records by that department;
- (b) those arising from the confidentiality provisions of statistics legislation (in Australia, the Census and Statistics Act 1905-1966); and
- (c) those relating to the disclosure of the contents of registers.

24. The first kind of problem can be illustrated by reference to the taxation statistics on which Australian national accounts are largely based. These are the aggregate tabulations of income tax returns which include those for the same businesses from which the Bureau of Census and Statistics collects enterprise and establishment returns in its integrated censuses of mining, manufacturing, electricity and gas production and distribution, wholesaling, retailing and associated services. Development of the national accounting system using the latter data (as well as taxation data for industries not covered) could proceed with much greater confidence if it were possible to align the aggregate industry data for enterprises compiled from the two sources. However it will be difficult to do this positively because the taxation returns are coded to industry by Taxation Office staff from only the simple description supplied by the business on its taxation return. A solution might be to provide the taxation office with a listing of the enterprises showing the more objectively determined industry codes assigned to them by the Bureau of Census and Statistics, but the Census and Statistics Act precludes this. The alternative solution of the Taxation Office supplying the Bureau individual records on magnetic tape plus identifying lists to permit the Bureau to link the records with its census files is precluded by the legislation governing the Taxation Office. This also prevents possible arrangements for avoiding duplication (e.g. the substitution of data from taxation records for that at present collected by the Bureau from small single-establishment enterprises).

25. Thus the taxation legislation, in making no allowance for the fact that the statistics legislation would prevent the individual details being used for any purpose other than the compilation of aggregate statistics, may actually operate against the interests of the business community, the statistical office and the user in preventing the data available to the government from being fully used for statistical purposes.

26. Legal questions arising from provisions of the statistical legislation can be illustrated by another example. Statistics of imports are compiled from copies of import entries lodged by importers with the Department of Customs and Excise. Under the provisions of the Census and Statistics Act it is not specifically provided that the Bureau may legally give back to the supplier of the data (the Department of Customs and Excise) the particulars included on any individual import entry. A large and complex statistical editing and tabulation system has been developed to compile statistics of imports and of shipping cargo movements. This data processing system also provides the most economical means by which statistics of imports by individual importers or cargo shipped by individual shipping operators could be compiled for administrative purposes. Compilation of this information in this way has the advantage of economy and of consistency with the more aggregated statistics used for other purposes. There would be no breach of statistical confidentiality in principle by the supply of this information to the Department of Customs and Excise because it was originally lodged with that department as part of administrative procedures associated with imports. However, the present provisions of the Census and Statistics Act prevent this being done.

27. An essential part of any system for linking and retrieving statistical data is a register, identifying the individual units whose particulars are recorded on the statistical files and providing classificatory information. Such a register is expensive to set up and maintain and has many other possible uses apart from the linkage and retrieval of data. Statistical uses include the preparation of despatch lists, collection control and sample selection. Non-statistical uses include the preparation of directories of various kinds, for example, a directory of manufacturers containing a list of the products marketed by each manufacturer. Such a directory would be useful for trade promotion purposes and in fact one is maintained by the Australian Department of Trade and Industry. At least insofar as they relate to businesses and contain only information such as the name, address, industrial classification broad size grouping, and broad product range, there appear to be no overriding statistical policy reasons why the contents of such directories should not be made generally available or at least available to government agencies. Unless provisions can be made which enable the contents of such registers to be shared between several users, the cost of establishing and maintaining them may be such as to prohibit their construction for statistical purposes. The Australian Census and Statistics Act at the moment would appear to prevent the contents of such statistical registers being made known to any person or organisation outside the Bureau.

28. These legal aspects of the integration of statistical files and the sharing of statistical processing systems with other government departments and the sharing of register facilities are under consideration at present. It is hoped that arrangements can be made to take advantage of some of these possibilities without weakening the protection afforded by confidentiality legislation to statistical data generally and without damaging public confidence in the Bureau.

Possible Future Developments

29. In the discussion above a distinction has been drawn between statistical data banks and administrative data banks and there is an associated distinction between statistical data processing systems and administrative data processing systems. With the increasing use of computers for administrative purposes, the monopoly which statistical offices once held of large scale automatic data processing systems is rapidly disappearing. Since an automatic data processing system can be adapted to produce statistical aggregates as easily as it can produce other forms of reports the traditional role of the Bureau in the compilation of aggregate statistics in such fields as overseas trade and taxation may in the long run undergo a significant change. The Bureau may in future be concerned to a greater extent with the question of overall statistical needs and collection relationships, statistical and information processing standards, conceptual framework for the definition of statistical concepts, and facilities for the storage, linkage and retrieval of information and to a lesser extent with the actual manipulation of source data or initial compilation of statistics. It would follow from such a development that the Bureau as a statistical office would necessarily take a more direct interest in the policies and procedures governing confidentiality and the protection of privacy in information processing in government agencies generally because the formulation and maintenance of standards in these matters may be as significant from a statistical point of view as the standards applied by the statistical office itself to the statistics compiled from its own collections.

30. The importance of hybrid (statistical/administrative) data banks may also increase. These data banks raise the most difficult policy questions with regard to confidentiality. As has been shown earlier, there are strong reasons on grounds of efficiency for the establishment of these data banks. However, the problems of ensuring that appropriate confidentiality policies and procedures are followed, are very complex and they may also involve the most sensitive public relations.

31. The Bureau is continuously examining specific aspects of the whole question. The underlying principle applied in any future development will be for the Bureau to work within a legal and ethical framework which balances the right of individuals to privacy against the society's view of its right to be informed, and efficiently governed.

32. An atmosphere of change often gives rise to uncertainty and to fears for the security of individuals. The Bureau is aware of the way in which basic conditions governing confidentiality policy are changing and can see the need for measures to be employed to maintain public confidence in its operations. Some clear logical thinking will need to be done in professional circles to establish a greater measure of consensus on principles, practices and terminology as a basis for more rational discussion as and when the problem becomes a matter of public debate.

THE GENERAL QUESTION OF CONFIDENTIALITY
ESPECIALLY AS INVOLVED IN THE USE OF COMPUTERS

I. Introduction

1. The subject of confidentiality invites consideration of many related topics. Their discussion would carry us beyond our central theme of confidentiality as it relates to the operations and practices of statistical agencies.
2. Confidentiality problems were recognized prior to the advent of computers. Safeguards were written into the Statistics Acts of several countries (typically decades before the first computers) as a legal expression of the concern of producers of statistics for the protection of the individual. Outside of statistical agencies information systems have also existed for a long time without the computer being involved. For instance, a telephone directory fits virtually any meaningful definition of an information system.
3. Computers, however, are such a powerful tool in the hands of both users and producers of data that the dimensions (though not the substance) of the confidentiality problem have been transformed. Computers, by enabling users to apply analytical and decision-assisting techniques to a variety of statistical information, typically in highly disaggregated form have stimulated an increasing demand for detailed information. Similarly, the increasing role of governments of all levels increases the burden on statistical agencies for data to support their administrative, regulatory, policy-making, judicial, and planning activities. Computers have also provided statistical agencies with a tool for processing, storing and retrieving information from a variety of separate or linked files collected over long periods of time. Thus computers have altered the statistical "market" on both sides: on the side of the "purchaser" as well as on the side of the "producer."
4. The statistical "market" situation, however, has a third dimension as well: the needs and concerns of the general public. The public benefits from the legitimate uses of statistics by governments, businesses, non-profit organizations, academic users, etc.; yet, the public is also concerned about the increasing burden that the collection of the statistics puts on them and by the real or imagined possibility of the misuse of "George Orwell" 1984-type dossiers about individuals. This brings us to the consideration of a seldom made distinction between confidentiality and privacy of data.

II. The nature of data confidentiality, privacy and security

5. Statistics Acts typically represent a type of "social contract" between a statistical agency and the data-providing public. It authorizes a statistical agency to collect data from individuals in order to satisfy the legitimate information requirements of the country and, on the other hand, it provides protection to respondents by prohibiting the publication or dissemination of individually identifiable data. Hence the intent of the Act is to guarantee that no harm will come to an individual as a result of his compliance with the Act. (Actually, even the publication of aggregate data may result in generalized "harm" to individuals, for example changes in legislation).

6. The protection of individual returns embodied in Statistics Acts typically includes at least the following three provisions: (1) the individual returns furnished by persons, businesses, etc. will be used only for statistical purposes and not made available for taxation, regulation or other administrative action, (2) the returns will be handled only by sworn staff of the statistical office and (3) the data will be disseminated only in a form which will not permit without authorization the identification of data relating to any individual, firm or other respondent.

7. The essence of the first provision is that information provided by a statistical office should be designed for use for statistical purposes since statistical analysis does not require information about identifiable respondents, it therefore follows that the statistical office need not and must not provide information pertaining to a particular individual or other respondent, nor should it publish or otherwise make available the records it collects in a way that permits the identification of these records as being specific to particular respondents. Thus the third provision is essentially implicit in the first. The second provision simply relates to the responsibility of the officers of the statistical agency with respect to data security, i.e. it establishes the accountability by the personnel of statistical offices.

8. Confidentiality, as a legal concept, is the legislator's imperfect translation of the following: "Don't publish or make available information collected under the authority of the Statistics Act which may be harmful to the individual respondents who provided them." This more diffuse concept of harmfulness is clearly the real concern of the public. Thus, there is a real distinction between privacy, the public concern and confidentiality, the legal concept.

9. An analysis of any country's publication programme would probably show that, in a strictly legal sense, the confidentiality provisions of the Statistics Acts are not absolutely observed. This is recognized sometimes through specific exception in the Statistics Act (the new proposed Act in Canada will permit DBS to publish a register of establishments with name, address, industry code and employment size), but sometimes even without such provisions. When for example, cross-tabulations of industry by occupation even at a moderate level of geographic breakdown are published, the tables often have several cells showing one respondent. Such entries are potentially recognizable as being the response of a specific individual. Industry and occupation, however, represent reasonably common knowledge and no one is apparently particularly concerned even though the legal requirement of confidentiality has been violated. A table concerning income distributions and showing some cells with only one respondent would, however, be intolerable since this would result in the violation of the privacy of the individual income data.

10. The distinction between confidentiality and privacy is a fine one and the Act does not distinguish between them. Yet in its daily practice, the statistical agency is attempting to draw a line between the two concepts. And therein lies another fundamental problem: computers are not capable of exercising subtle value judgments; they require precise quantitative definitions capable of being translated into numerical procedures and algorithms.

III. Types of violation of confidentiality

11. Violations of confidentiality may arise in a number of different ways and it may be worth discussing the different types.

Breach of security

12. This type of disclosure, is essentially a breach of the secrecy oath each employee of the statistical office must take. It may consist of gossip among friends, or it may consist of the sale of confidential information for profit. In either case it is a breach of data security. The fight against such disclosure consists of strict emphasis on security procedures, thorough investigation of "leaks", and enforcement of physical security (locking returns after office hours, supervisory surveillance, etc.).

13. Computer technology adds a new aspect to the old problem of making an unauthorized copy of confidential data. It is relatively difficult to determine when an unauthorized person unlawfully makes a copy of a computer tape containing statistical information. During the first and second generation of computers this was easier to avoid since usually only one job was being processed by the computer at a time. In the current generation of computers several jobs are frequently being processed at the same time. This problem is even more difficult to avoid when the statistical production is carried out on a computer which is not in the possession of the statistical agency, as in the use of commercial service bureaus, other government departments, etc. In this last case, the tapes are "under the custody" of employees of the statistical office, however they have no real control over them during their processing by the computer. The control during the time rests with the "software."

The "Old School Tie" problem, or the friendly persuasion

14. This may be a dangerous tupe of disclosure because it appears so harmless. It consists of passing information to a person known to be "thoroughly trustworthy". Typically he wants the information for himself and would not pass it along to others. He may belong to another government department, may even be the minister responsible for the statistical agency, or he could be a friendly newspaper man. The defence against this possibility lies in the thorough indoctrination of employees: information is confidential, not only when it relates to individuals but even in aggregate form before it is published or made generally available.

Processing accident

15. As long as people are involved in the processing of data there may be accidents. Copies of confidential data mailed out, computer tapes may be mislabelled and printed out for inspection by unauthorized persons, or confidential reports could be misfiled in a library open to the public.

Inadvertent direct disclosure

16. While the previously mentioned types of disclosure can be blamed on a lack of security of one kind or another, direct disclosure may result from the statistical office's publication practices. It would not involve the provision of confidential information about a person, establishment or other respondent with his full identification, but rather his inadvertent identification through the publication of too much detail. As was hinted above, precise rules are

generally missing to ensure against this kind of disclosure.

17. In the area of economic statistics some guidelines exist. For example, it is generally accepted practice to blank out information which is based on fewer than three respondents or when one or two respondents account for more than a specified proportion of the aggregate. Even in the field of economic statistics, however, the issues are not always clear: an aggregate derived from an establishment survey may satisfy all the requirements of the conditions mentioned above, yet it may be a disclosure if all the establishments involved belong to the same enterprise. The Census publications of many countries indicate that not even such general guidelines exist in the field of population or housing. Once the precedent is established to publish cells containing zeros and ones, the statistical office leaves behind at least one kind of potentially firm quantitative interpretation of the confidentiality rules and begins to rely on judgment. The line is a thin one: if occupation by industry can be published, why not occupation by industry by age?

Residual disclosure

18. This type of disclosure occurs inadvertently when two or more sets of data taken together enable the user to identify information pertaining to an individual respondent even though there is no direct disclosure. The trivial example occurs when an entry in a table is blanked out but which can be deduced from the marginal totals and the other entries in the table. A less trivial example is the publication of certain industry totals by province or other local area when one cell is dominated by a single respondent. Usually this province is grouped together with another province whose corresponding total, by itself, is publishable. If in an other table the "publishable" province is identified separately, the corresponding total for the first province can be obtained by subtraction - a residual disclosure: In Canada this is a particularly vexing problem because the Census of Manufactures reports are shared with provincial statistical agencies (with the consent of respondents). It is difficult, however, to convince a province, which is entitled to publish its results, to forego publication so as to safeguard the confidentiality of respondents in a smaller province. In order to overcome this problem a jointly agreed upon publication format is developed each year.

19. Residual disclosure is difficult enough to guard against when the method of dissemination is restricted to a fixed set of publications. As statistical offices are developing a capability for retrieving summary data from individual responses, the problems of residual disclosure will multiply exponentially.

IV. The definitional problem of confidentiality

20. It has been said that the heart of the difficulty in dealing with issues of confidentiality is that confidentiality has not been defined in quantitative terms. There is a genuine conflict between production methods which allow the inexpensive compilation and publication of more data in greater detail and rules of confidentiality which are essentially based on the paper and pencil age of processing. Computers could handle the problem of testing for confidentiality if confidentiality could be defined non-subjectively. Without such definitions our best hope for using computers to tackle this problem lies in trying to identify at least those data which are clearly and obviously not violating confidentiality (leaving only the remainder for human judgment).

In the absence of hard and fast definitions, the next best thing would be some quantitative conventions.

21. Conventions are in the process of being established through precedents. For example, our U.S. counterparts "published" after the 1960 Census individual census data without name or geographic identification (broad, regional identification were retained). These tapes were based on a sample drawn from the Census at the rate of 1 in 1,000 (another sample of individual records at the rate of 1 in 10,000 was available on punched cards). The precedent, once established, became the precedent for the similar dissemination of data collected in the monthly Current Population Survey (at about the same sampling ratio). The British are similarly disseminating data collected by the Social Surveys. Yet these conventions are hard to defend. Why stop at a sample of 1 in 1,000? Why not 1 in 100? Why not 1 in 10?

22. One such possible convention concerning direct disclosure seems to offer itself through some methodological considerations (see Section 7). Residual disclosure could then be defined as the publication of aggregates from which, through mathematical manipulations, data could be derived which would represent direct disclosure. Such a definition would at least conceptually permit a precise determination in any concrete situation as to whether or not a particular tabulation, if released, would result in disclosure. This would be the first step toward "mass production" of tests of disclosure.

V. The issue of confidentiality and the problem of duplication of collection

23. The public also needs to be protected against the burdens of providing the mass of information required by the modern state. Information for statistical purposes is only one part of the burden. Statisticians have a clear responsibility to reduce the burden if the needed data can be extracted from data already provided by the public for administrative purposes. This is an increasingly rich source of statistical information since the administrative activities of modern governments encompass ever wider spheres.

24. All concerned would benefit if statistical agencies obtain access to administrative data, yet the public may regard this with some misgivings. It is felt that the accumulation of information about individuals within an agency will allow the linkage of information from various sources and thus create the feared "George Orwell" 1984-type dossier. In this context, the provision of the Statistics Acts that identifies the purpose of data collection (whether directly from the individual or indirectly from an administrative agency), as being statistical has a bearing. A statistician must strive to clarify the fundamental distinction between information to be used for statistical or other purposes. Statistical agencies normally already have authority to collect much of the information held by administrative agencies. It benefits no one to duplicate the collection of information from the individual when it can be compiled from administrative sources. Yet such access puts a responsibility on the statistical officers to protect the genuine privacy of individuals. Information flow between the statistical office and administrative agencies must be in one direction only - from the administrative agency to the statistical office.

25. Access to administrative records for statistical purposes can be a mixed blessing under some circumstances. Needs may arise for access by the administrative agencies to individual information which is already collected by the statistical agency. If the statistical agency does not provide the

required individual data, the administrative agency may argue that the only way to avoid duplication would be for the statistical agency to abandon the collection in favour of the administrative agency. It may promise, in return, access to the statistical agency. This may or may not be the best solution, depending on the circumstances.

26. The abandonment of primary data collection by the statistical agency in favour of administrative agencies can be dangerous. The statistical agency cannot easily enforce appropriate statistical standards on the administrative agency unless some incentives are possible or the collection of adequate data is almost self-enforcing. An example of such incentives is when the statistical agency attaches standard statistical codes to the forms collected by an administrative agency or when it provides the administrative agency with access to carefully maintained, high-calibre registers (this is often explicitly permitted by Statistics Acts). Such a relationship exists in Canada between the Dominion Bureau of Statistics and the Unemployment Insurance Commission: the latter provides much of the updating information for the register of establishments maintained by DBS, in return it receives from DBS its own forms appropriately coded by industry. It is also likely to be both a contributor and beneficiary of a central register of business units that is under development by DBS. Examples of data collection by administrative agencies whose standards are almost self-enforcing are Vital Statistics, and External Trade.

27. More generally, abandonment by a central statistical agency of the collection of statistics to administrative agencies may result in the erosion of the function and authority of the statistical agency. As other government departments develop easy access to computers, it may be tempting for them to compile statistical tabulations from their administrative data instead of, or in addition to, passing the original data to the statistical agency for this purpose. Such an erosion could also have as a result a direct deterioration in the level of reliability of the particular series concerned. This could occur since the administrative agency compiling statistical tabulations for general use is less likely to adhere to the methodological requirements of estimation, editing, imputation and other uniquely statistical techniques of data processing that are part and parcel of the production of statistics.

28. In summary, safeguarding confidentiality can lead to duplication of effort by the statistical agency and the administrative agencies, or in fact even between statistical agencies. There is no single best solution to this problem. Leaving the collection to administrative agencies can be a useful, in fact, sometimes unavoidable method of avoiding duplication and can result in high-quality statistics under the right circumstances. Joint collection, with the primary duty of collection vested in the statistical agency, may be another possible solution. In the latter case, care must be exercised that the authority of the Statistics Act to collect data is not used as an umbrella for the provision of individual returns to administrative agencies, unless this is done with the explicit permission of respondents. Statistical agencies should have unimpeded access to administrative records for statistical data (but this should not become an excuse for abandoning data collection to administrative agencies). At the same time the statistical agency must strive to maintain in the public mind the fundamental distinction between data for statistical and administrative purposes.

VI. Technical issues of confidentiality

29. It has been mentioned that the computer as such did not create the confidentiality problem. It is a fact, however, that computers have extended enormously the users' ability to utilize data (and generated therefore an increased demand). At the same time they also extended the ability of the statistical offices to cope with this demand (even though never to the full satisfaction of users or, in fact, to the statistical offices themselves). These changes, though quantitative in nature, are of such dimensions that one can ask at what point the accumulation of quantitative changes becomes a qualitative one. In this sense, therefore, though they did not create the problem, computers have increased it.

30. From the point of view of a professional data processor, confidentiality is the problem of the data collector. From his point of view retrieval from, for instance, census files is no different from selecting a book in a library. Viewed in this fashion, statistical files are not very different from a business inventory, credit bureau or library record system: they all consist of a data base consisting of a record of each separate unit. Information is added to or retrieved from the base on demand. The methodology of these types of systems is identical. The data processor needs to know the answer to such questions as which data items are to be secure, what confidentiality rules should apply to what kinds of retrieved tabulations and what is the procedure in precise terms to determine the confidentiality of a particular tabulation. And so we are back to the definitional problem.

31. The issues can be illustrated by two Canadian systems. One is a data bank of time series called the Canadian Socio-Economic Information Management System (CANSIM). The overwhelming majority of the time series stored in CANSIM are either published or publishable aggregates, hence non-confidential. However, CANSIM is equipped to store time series which are sufficiently disaggregated that they are confidential (accessible only by officers of DBS). Technologically CANSIM is capable of dealing with this type of confidentiality requirement. Time series that are confidential can only be accessed by persons who know a specified password and different series (in fact, different data in a single series) may be accessible through different passwords. The key to the success of computer methods in this case is that the status of data with respect to confidentiality is essentially determined outside the computer system, prior to input. There are many articles in the literature on data processing dealing with ever more sophisticated procedures designed to prevent unauthorized access to data that is known to be confidential. Actually, these are not methods to deal with the confidentiality problem: rather, they are methods to deal with the security of data, given that their confidentiality status is already determined.

32. The problems of checking the confidentiality of the retrieved tabulations are entirely different in the case of another system currently being developed in DBS. It is called the Geographically Referenced Data Storage and Retrieval System (GRDSR) and is designed to store individual data from surveys and retrieve arbitrarily specified tabulations requested by users, referring to an arbitrarily specified geographic area. The first concrete implementation of GRDSR is planned in connection with 1971 Census data. The design of this system is quite advanced and there is little doubt at this time that the objectives of the system mentioned above will be achieved. From the point of view of confidentiality, however, GRDSR presents altogether different problems from CANSIM. Whether or not the tabulations retrieved from this base are confidential is, in contrast with CANSIM, not known until

after the retrieval. Procedures have to be designed, therefore, to screen the retrieved information prior to passing it on to users. In the absence of precise quantitative definitions as to what constitutes disclosure, the system can provide the reviewing statistician with a useful guidance: it should be possible to build into it conservative criteria to indicate when a particular tabulation is not confidential. The remaining cases would therefore appear as potential violations of confidentiality and will probably have to be reviewed by humans. With respect to residual disclosure, our hopes are essentially similar: to design a system which will assist a human reviewer to carry out his task more efficiently. Actually, under the definition of direct disclosure implied by section 7 (or something similar to it) a mathematical solution has been developed to determine whether or not a particular tabulation may represent residual disclosure.

33. To close this section on a more positive note, an example of a situation where the computer was successfully used to assist in safeguarding confidentiality can be given. The Census of Manufactures is based on the collection of data from individual establishments and most of the tabulations are also carried out on establishment data. Prior to the computer era the tabulations were, of course, checked for confidentiality, but this was restricted to checking for the disclosure of establishment data. Now, computers enable us to carry out a much more advanced check: the central list of business units identifies establishments by companies and, using this facility, the Census of Manufacturers tabulations are checked not only for disclosure at the establishment level, but also at the company level (when data are collected from establishments it may happen that a particular tabulation appears to be non-confidential even though the establishments involved are mostly belonging to a single company).

VII. Methodological issues of confidentiality

34. Developments in statistical methodology, similar to developments in computer technology are, of course, not the cause of problems of confidentiality. But just as in the case of computer technology, they may well sharpen some of the issues involved. Developments in the field of record linkage are closely related to the concern of the public for privacy. Privacy has been defined by Professor Weston of Columbia University as the right "to determine what information about ourselves we will share with others". The concern about privacy, in fact, centers around the question of making such information available to others who are not known to the respondent. Professor Weston argues that an attack on a man's ability to control what is known about him represents an assault on his humanity. Yet, in the modern state, the life of individuals is widely described and documented by written records (which typically are converted into machine-readable records). When two records dealing with different aspects of an individual's life or activities are brought together and merged into a single record by the process of record linkage, the extent of the recorded information about him is increased geometrically.

35. The fine, yet vital, distinction as to whether record linkage is performed within a statistical agency, which will carefully guard against the disclosure of individual information, or by an administrative or control agency which may use the information in an altogether different fashion, is seldom made with sufficient clarity.

36. Record linkage, of course, is accomplished most easily when each individual has a unique number which is reported by him on the different records. Yet it is easy to see intuitively that it is not necessary to have a unique number to identify two records as referring to the same person or event. If enough information is recorded about the person in both records (such as his age, sex, marital status, address, ethnic origin, etc.) then, even in the absence of a unique identifier, each additional piece of information narrows down the number of possible persons to whom the record could refer, until the choice is narrowed to a single individual. Methodological developments have been reported in the literature during the last two years which may well lead to an economically feasible process of record linkage even in the absence of a unique identifier. The methodology involved is probabilistic. One can never be 100% certain that the linked records are in fact referring to the same individual, but one can specify a level of certainty, no matter how high, as long as it is less than 100%. Even in criminal courts, juries are only asked to determine guilt beyond reasonable doubt (not with certainty).

37. When the methodology of record linkage is fully developed (and there have been major advances in the last few years), it almost certainly will be used. Yet, in light of the justifiable public concern, statistical agencies must use record linkage with extreme caution and protect with double care the linked records from disclosure. In most instances the Statistics Act would entitle them to collect directly the additional information that record linkage might put at their disposal. It is not economical to collect directly information which can be created through record linkage within the statistical office: individual respondents would have to bear the burden of response as well as, through their tax money, the cost of it. Hence, the efficiency of record linkage must be weighed against the public's concern.

38. Record linkage is related to the problem of confidentiality in yet another way. Disclosure involves making available information concerning a unique and identifiable individual. The statistical office, of course, should never disclose information about an individual who is identified by his name. But, the concept of disclosure also implies restrictions on disclosure of information on an individual who can be identified through his characteristics. In this latter case, therefore, disclosure occurs when a user can identify a respondent through an implicit or explicit process of record linkage. From this standpoint violation of confidentiality might be defined as the disclosure of information that goes beyond that required for identification alone. Hence one may argue that a table in which some of the cells contain entries of one but in which none of the marginal totals are ones does not represent a violation of confidentiality. In fact, in this case the entry of one can only be identified as referring to a particular person if the reader knows a priori that the particular person has all the characteristics indicated by the table. However, should another dimension of breakdown be superimposed on the table then disclosure would clearly occur. At that point, information is disclosed about a person which goes beyond that required for his unique identification; in a sense at that point the reader may learn something new about a particular person. To illustrate by an example, suppose that a particular census table published for a given municipality relates to the cross-classification of persons by industry and occupation. Suppose that one of the entries in the table is one, but none of the marginal totals are ones. The reader at that point may recognize the person to whom the one refers: he may say, "Joe Smith is a statistician working in a synthetic textile mill; the table shows that there is one such person; that entry must therefore refer to Joe Smith". For the reader to recognize this entry as referring to Joe Smith, he must know in advance that Joe Smith is a statistician and he works in a synthetic

textile mill. Since none of the marginal totals are one neither the industry nor the occupation identifies Joe Smith by itself: both are needed simply for identification (actually, if the entry is based on a sample, the reader must also know for full identification that no other person has the same occupation and industry). If this table, however, is extended to a cross-classification of industry by occupation by age, at that point the reader may learn Joe Smith's age.

39. Should this analysis be correct and acceptable, it would go a long way toward conceptualizing what is involved in disclosure and hence it might carry us closer to arriving at an acceptable, yet precise definition of disclosure. Direct disclosure might then be defined as an entry based on a single observation (or other specified number of observations) provided that at least one of the corresponding marginal totals is also based on the same single observation (or other specified number of observations). Given a precise definition of direct disclosure, a corresponding definition of residual disclosure can be derived.

40. Under the heading of methodological issues related to confidentiality, it is sometimes argued that response errors, automatic error corrections, etc., in a sense moderate the problem of confidentiality since all of the information that might be disclosed in a tabulation is not "real". This seems to be a relatively minor issue. First of all, if enough information on an individual is contained in a record he can be identified using the statistical theory of record linkage briefly referred to above, even in the presence of errors in some of the recorded information. Second, if appropriate statistical standards are maintained, the extent of such imputations should be at a low enough level to ensure that its effect is small.

VIII. Confidentiality and the organization of the statistical system

41. It was mentioned above that a centralized statistical system makes it easier in practice both to protect confidentiality, and to do so in a publicly visible fashion. It is important to here again distinguish carefully between the use of information for statistical and for administrative purposes. This distinction is much easier to make if the agency involved in the collection and publication of statistics has only statistical functions. When the statistical function is an adjunct to an administrative function, it may not only be more difficult to appear to maintain confidentiality, but it may in fact be more difficult to do so.

42. On the positive side, when a statistical activity is an adjunct to an administrative activity, it may be easier to gain access for statistical purposes to the records of the particular administrative agency involved. It may, however, be more difficult to gain access (even for statistical purposes) to administrative records of other agencies. A central statistical agency can argue with conviction that it be allowed access to administrative records for statistical purposes. Whatever the confidentiality constraints specified in an act governing an administrative agency, a Statistics Act is probably more stringent in this respect. Moreover, the central statistical agency would probably be entitled to collect the information in question directly from the respondents.

43. Any central statistical agency has a variety of information about respondents within its walls. Record linkage performed on this data does not require inter-agency agreement. For those concerned about privacy this

may be an additional source of worry. Objectively, however, the public would benefit by the elimination of redundant data collection and by the increase in the information gained from data already collected.

IX. Confidentiality aspects of different forms of data dissemination

44. In this section which can be considered to be a summary of the paper, the issues mentioned above have been regrouped under the alternative forms of data dissemination.

Publications

45. This is, of course, the traditional and still most important form of data dissemination. Disclosure is easiest to prevent in a printed publication. A publication, no matter how large, involves only a finite number of fully foreseeable cells. Once disclosure is precisely defined it will be possible to devise algorithms to check the publications for both direct and residual disclosure. Even in this case serious problems remain since mere identification of "disclosure cells" does not guarantee a method to handle them. No fully satisfactory method of suppressing disclosure cells in a table, even after they have been identified, has yet been devised. Rows or columns may be collapsed, but this "ruins" some other publishable cells in the table. It also tends to diminish the value of the whole publication. Another, even less satisfactory way of dealing with the problem is simply to blank out any cells which may result in disclosure. This, of course, would result in disclosure through the discrepancy between the totals implied by the body of the table and the actual totals presented on the margins.

46. Neither the concept of confidentiality (which is at the present largely based on judgment), nor the practice of protecting it is uniform. It appears that conceptually identical tables are treated differently depending on the information they contain. A tabulation containing several zeroes and ones may be published if it refers to a cross-tabulation of say age by marital status; but a similar table may be considered to be a disclosure if it represents a tabulation of income by occupation.

User requested retrieval

47. The problem involved in this type of data dissemination has been highlighted by contrasting retrieval from a data bank containing aggregates whose confidentiality status is determined prior to input (such as is the case with CANSIM) with a data base containing individual respondent data (such as the GRDSR base) where confidentiality must be determined after retrieval. In the latter case methods have to be devised not only for checking for direct disclosure, which is not an easy task by itself, but also for residual disclosure which involves keeping track of all requests previously processed. DBS is actively working toward finding a solution to this problem which may be a crucial roadblock in the way leading to the full use of such systems as GRDSR. It might be mentioned that, given a definition of direct disclosure a corresponding, mathematically precise definition of residual disclosure and a mathematical theory involved in testing for it is possible. The economic feasibility of the procedures involved are currently being investigated. Particularly the cost of testing for residual (as opposed to direct) disclosure will be high irrespective of the method used. This is due to the fact that the very concept of residual disclosure involves keeping track of all tabulations and retrievals from a particular file and this can be an enormous task.

???

"Summary tapes"

48. This method of "publication" was introduced as part of the data dissemination programme for both the 1961 and the 1966 Censuses. These are computer tapes containing basic tabulations from the Census, i.e. minimal or no cross-tabulations but by very small geographic areas. The problems of confidentiality are conceptually similar to those of standard publications. To the extent, however, that these summary tapes contain information referring to exceedingly small areas (areas containing 150 to 200 households) the "standard" problem of confidentiality of regular publications is compounded.

Sample of individuals on tape

49. Disseminating data for a sample of individuals is a recent innovation. To the best of our knowledge it is only applied in the area of population and income tax household surveys. A sample of records is selected, all direct identification and geographic codes removed (at least the detail of such codes). The sample records are typically released in machine-readable form. The advantage of this form of data dissemination is that it permits users to "play around" with the data, and permits certain types of analyses which cannot be carried out on aggregate data. It has also been found to be a useful aid in teaching survey courses since it is a source of "live" data.

50. The question whether this form of data dissemination does or does not represent disclosure runs head-on into the problem of definition--what do we mean by disclosure? The most liberal interpretation of the concept is where violation involves breach of security. The most restrictive interpretation is one wherein disclosure occurs whenever someone, on the basis of prior knowledge about a respondent can find out some additional information about him (or her). Should our definition of disclosure coincide with the latter, then the dissemination of individual records on tape, even after the removal of formal identification, is clearly a disclosure. After all the Census, for example, contains upwards of 70 fields of information about most individuals. Even if each field of information is only a categorical "yes-no" type of variable (and obviously most of the variables involved are coded to more than two classes) the theoretically possible number of different records is equal to 2 which is a number exceeding 10. There can be no doubt that with so much information available about each person on the tape, every person is uniquely identifiable (i.e. someone who knows enough about a given person can recognize him on the basis of the information that is recorded on the tape). Actually, it is very probable that one-tenth of the information contained on the tape about each person would uniquely identify most persons and, therefore, the remaining nine-tenths of the information would represent disclosure. In other words if, for example, an outside user, say a credit bureau or the police, tries to match through record linkage the tape of individuals released by the statistical office against its own files, then if its own records contain sufficient information about each person (information such as age, sex, marital status, household size, number of children, age of children, size of municipality, educational attainment, etc.), then almost certainly every person on the sample of records from the census file could be matched against a record on the other file (though not conversely). The fact that such a tape of individual records is only supplied on a sample basis does not alter the fact that for those persons it is disclosure. Sampling simply decreases the potential benefit that someone looking for the characteristics of a particular person would derive through linkage with the sample tape: the probability of his finding the particular person is equal to the rate at which persons were selected for inclusion on the sample tape. If this rate is 1 in a 1,000, the

odds are one against a thousand that someone searching for the census records of a particular person would find him.

51. Sometimes the argument is put forward in defence of this scheme that no one can be sure about the identification of a person on the sample tape because there may be other persons with the same characteristics who did not happen to get included in the sample. This argument is relevant in marginal cases (such as in section 7 in connection with the discussion in the context of a publication of aggregates, of the level of cross-tabulation at which disclosure may potentially occur). In the present context, however, when so much information is released separately about each person in the sample, as is the case in connection with a sample of records from the census, such an argument is spurious. The chances of two people being identical with respect to more than 70 characteristics are infinitesimally small.

52. Allusion was made at the beginning of this paper to the need for developing conventions in the absence of rigorous rules. It seems that the case of the release of individual data on a sample basis is in dire need of such conventions. Should we release such information? If so, what is the safe sampling ratio, i.e. at what level of sampling is the potential value to someone who wants to break confidentiality small enough to make the exercise too expensive in relation to potential benefits? Is the practice more defensible when fewer items of information are recorded and, if so, how many items of information can safely be so released without the danger that individuals get identified?

THE GENERAL QUESTION OF CONFIDENTIALITY, ESPECIALLY AS
INVOLVED IN THE USE OF COMPUTERS

Reappraisal of Confidentiality Issue

1. The issue of confidentiality of individual information (personal or business unit) is receiving renewed attention in New Zealand as in many countries at present because of the impact of central computerized processing with the imminent development of computerized files, permitting data linkage and vastly improved retrieval systems. The term 'renewed' is justifiable, since this issue has always represented a problem. The arrangements made to secure it and which have been effective in the past have now to be reassessed for adequacy in the light of modern developments. New Zealand is, like many other countries, actively engaged in this process. Confidentiality of statistical records must be considered under a number of situations likely to be encountered under present and future processing systems. These depend on whether or not records are subject to linkage or association and the kind of control systems employed.

New Zealand Situation

2. While computerized statistical files can be distinguished according to whether primary (individual) records or data outputs are stored and/or linked, New Zealand has up to the present had little or no actual experience in machine-aided linkage of primary records from different sources. It is however planning linkage of statistical data outputs for common and small sized areas and (see paper under item 5(a)) is planning for a central data file in computer storage. The confidentiality problems are much greater in the former setting than they are in the latter one.

Classes of Records to be Examined

3. Data records or information may be obtained for statistical purposes -
- (a) From direct statistical surveys subject to guarantee to persons or firms that individual particulars will not be disclosed either at the collection or publication stages as stipulated in enabling statistical legislation (in New Zealand the Statistics Act 1955).
 - (b) Indirectly as by-products of administrative requirements which in some instances are covered by specific guarantees of confidentiality (Inland Revenue) or by generalised provisions, or under official Secrets Acts, standard administrative procedures, etc. The confidentiality problem

still remains, although the statistical purpose component, vis-a-vis the administration element in such primary records, may vary widely in extent between different kinds of administrative records.

- (c) Through mechanized data linkage systems in which information drawn from one or more areas of statistical records and/or relevant administrative records (perhaps also in future from contributed data or records from non-official agencies) can be associated.

Types of Institutional Arrangement Controlling Storage of Statistically-relevant Records

4. There are three main variants found in New Zealand in the type of data holding agency.

- (a) In the first, which applies specifically to the Department of Statistics, statistical data or information from direct surveys from agencies is both stored and directly controlled both in data processing and in granting access.
- (b) A second variant is where a specific agency of Government is entrusted with the operation of a Computer Bureau or Centre and carries out EDP processing on behalf of the Department of Statistics or other Government departments doing statistical work.
- (c) In the third variant, some external agency, for example the United Nations, a Regional Authority, or a Research Organisation, seeks to obtain statistical records (generally in aggregative form or at least with direct identifications such as names and addresses removed from individual records) from the national statistical office for the purpose of assisting planning, research or further processing.

5. The problematical aspects of any such arrangements relate to the unknown capability of such organisations to obtain cognate or even more remote kinds of data which, when associated by linkage methods, could lead to identification of personal or business respondents, possibly by association of a number of characteristic variables in fine reference regional areas.

Present and Future Procedures for Processing and Custody of Statistical Records

6. Basic statistical records with identifications such as names and addresses omitted, which result from the processing of either direct statistical surveys or from the extraction of data from administrative documents were, until recently,

generally stored clerically and controlled by the Department of Statistics. After data had been transferred to computer input media or otherwise processed, the original records containing identification names and addresses were destroyed or returned to the originating source if they were administrative documents subject to less vigorous security classification. More restricted administrative documents were of course never removed from the premises of the originating source, appropriate secondment or other mutually acceptable arrangements having been made for extraction of statistical data. This situation then conformed to the first institutional variant mentioned earlier and was the traditional pattern in New Zealand.

7. To a large extent this situation still applies in New Zealand, where a transitional phase of transfer to the second variant is in progress. The standard situation is likely to become, however, this second variant, in which statistical records from the processing of surveys and administrative documents will be stored and controlled by the New Zealand Government Computer Centre. Comprehensive and meticulous rules and procedures are at present being developed by the Department of Statistics or agency providing the basic records, which will govern access to or processing of records and the release of information from them in any form, either aggregative or individual.

8. At the present time no linkage of primary records is being undertaken by machine methods although attention is being given to the possibility of permitting linkage of areal statistical outputs with other data series for similar areas. Here again, however, original statistical survey records will be destroyed after processing, the appropriate customary arrangements being made by the responsible authorities for administrative records.

9. It is quite possible for the third mentioned variant to apply concurrently with either of the other variants. New Zealand, as doubtless have other countries, encounters considerable pressures from outside organisations to release primary level data. If this were to be considered permissible, it would undoubtedly magnify problems of confidentiality.

General consideration relating to Privacy of Unit Records

10. Current-type arrangements for protection of privacy of detailed records have not presented undue problems in the past. However, the enhanced or potential value of detailed statistical records for research or administration which have become possible by the development of computerised linkage and retrieval systems has added a new dimension to the confidentiality problem. In this setting, the following considerations apply:

- (a) The problem in essentials reduces to the issue of privacy or protection of personal or single respondent records - whether this problem arises at the basic record custody or at the data release stage.

- (b) Confidentiality of aggregated data from which information cannot be identified back to any person or firm is not an issue, for example the publication of larger personal incomes statistics on an all-New Zealand basis.
- (c) Arrangements need to be made to ensure that parts of aggregated released data cannot be so identified, for example publication of statistics is restricted in New Zealand to three or more independent units in the Census of Distribution.
- (d) Single cell data need not infringe any confidential provisions so long as identification cannot be made, for example, Census of Population income statistics for larger regional areas.
- (e) With records bearing fine location references the possibility of identification to one person becomes increasingly possible, the smaller the area the greater the chance of apparent identification, for example analyses of persons by marital status by age or higher income levels by age for areas less than city subdivisions or in very small towns present problems.
- (f) Similarly greater attention to ensure privacy becomes more pressing where the extent of sub-categorisation is enlarged with the more sensitive characteristics being shown in finer detail, or where a combination of a large number of classes is accompanied by greater detail within classes, for example, family heads by sex by age group by income of head and number of children as analysed from a Census of Population.
- (g) While nearly all of the preceding considerations apply to even unlinked records or data output, the impact of record linkage extends the horizontal range of possibilities where confidentiality could be breached, and in the vertical sense magnifies the risks by attachment of associated characteristics relating to any one person.
- (h) Increasing demand for more and more data for small and still smaller geographical areas, a pronounced feature of regional planning requirements, imposes a challenge to central statistical offices to simultaneously assist in giving much needed data which it is within their capacity to supply, yet to maintain adequate standards of confidentiality preservation in a situation which is more conducive to a breakdown of such safeguards.

Prerequisites for Introduction of Linkage Systems

11. Assuming that a satisfactory means of effecting linkage can be implemented without recourse to unduly inconveniencing or antagonizing survey respondents by requiring them to know or use some individual identification device such as a unique reference number, there would be no strong case for an embargo on linking records under fully guaranteed conditions of non-disclosure to unauthorised enquirers. While some objections will undoubtedly be raised to the principle of linkage, adverse reaction will be less influential or prevalent if the fear of misuse or abuse of linked information can be allayed or dispelled. It is of course asking too much of human nature to expect complete elimination of initial vocal opposition, particularly in a democratic political environment; only time reduces the influence of conservative attitudes.

12. One of the primary requisites for the successful introduction and gradual extension of linked systems must therefore be the mounting of a continuous and well presented public relations programme to engender understanding and confidence of the proposed measures reinforced by legal and practical measures to guarantee and insure maintenance of privacy. This should demonstrate both the useful contribution which linked information systems make to the range and efficiency of the information basis for community decision-making and planning, and the facilitation of research as well as emphasizing the effectiveness of safeguards against abuse.

13. It should not be overlooked, however, that the effectiveness of safeguards rests ultimately upon the political will of the nation. Linked storage and retrieval systems will function under the aegis of Governments which, it is assumed, will ensure by legislation and by the normal delegation of administrative authority that effective control of computerized statistical storage systems are vested in competent staff imbued with a strong sense of professional ethics in the discharge of their duties. Although there will be some lack of acceptance of the fact, in a public relations programme it is desirable to educate the community that abuses of confidentiality provisions are theoretically possible under current systems of uncoordinated, fragmentary or unlinked records. While computer methods give an impetus to the use of integrated classifications and therefore make possible more effective data linkage the difference is in degree not in substance. The public reassurance of confidentiality preservation will, as it always has, remain founded on the twin bases of Government legislative prescription and professional standards of statistical staff handling data.

Unique Numbering Systems of Reference

14. At this point it is useful to briefly touch on the ancillary question of unique numbering systems of reference. The concept of assigning to each possible survey respondent a unique reference number which serves as a unifying or linkage medium appears on the surface to be quite unexceptionally simple and statistically attractive. Translation of the concept into practical reality however would appear to present formidable problems in a country like New Zealand.

15. The initial problems of introduction are somewhat similar to those of linkage systems in general, e.g. public acceptance of number-oriented references rather than the traditional name and address reference. Indications are that administrative problems of introduction are very substantial in coping with requirements for initial introduction and continual updating. Any first application to some specific field should however be carried out in a manner which will be capable of later extensions to the scheme.

Arrangements for Preservation of Privacy of Primary Records in New Zealand

16. Safeguards are of a three-fold nature: (a) legislative; (b) professional; and (c) administrative. For unlinked records the arrangements made under these headings have been generally quite adequate. Legislation such as is contained in the Statistics Act together with the Inland Revenue and some other Acts contain explicit provisions relating to confidentiality of returns from respondents, and prescribe offences and penalties for non-observance by officials. However one of the principal safeguards rests in the competence and sense of responsibility with which official statisticians at all levels approach their work. Moreover, the measure of statistical independence possessed by the central statistical office and, more explicitly, its permanent head in the discharge of statutory statistical duties is a strong contributing factor in securing the preservation of adequate standards of confidentiality.

17. Administrative procedures and in-house rules also assist in maintaining secrecy, by spelling out in detail precautions to be taken to avoid disclosure, whether at the survey data collection or data release stages, or in arrangements for proper custody of survey documents.

18. Such arrangements will still be necessary under any statistical system, but there must also be additional measures if the advantages to be procured from computerised data and record linkage are to be effectively realised. These additional requirements will be somewhat more extensive when the storage and retrieval centre is in some degree removed from the immediate control and supervision of the statistical office supplying basic records as inputs to the storage system.

19. In particular the staff operating the centre need to be technically competent and to comprehend the potential dangers which can arise from neglect of or inadequacy of detailed rules designed to ensure confidentiality. The senior directing staff must prepare in conjunction with the statistical or other authority concerned, and ensure compliance with, policy directives and approved procedural arrangements. A number of programmed check features must be devised to prevent unauthorised operators or agencies or individuals from interrogating the computerized file. The Department of Statistics is currently at the stage of commencing the development of procedures for these purposes. The use of lock-up storage for magnetic tapes and desks with access available only to nominated EDP staff, the use of codes to ensure the authorization of a data inquiry and other devices are being examined, to ensure that confidentiality in the Government Computer Centre is maintained and is seen to be maintained.

Summary

20. Confidentiality has not been a statistical problem area of great moment in New Zealand up until recently because traditional methods of preserving secrecy have been adequate and not subject to public misunderstanding. However with the storage of statistical data records in a detached Government Computer Centre, urgent attention is being given to the implications this holds for confidentiality; such additional arrangements as may be required are currently being devised.

21. Computerized statistical file systems have not yet been implemented in New Zealand except for one example of aggregated time series data but planning has commenced and solutions to the technical problems and public relations angle will have to be found.

22. An Extract is given below from the report of the Technical Committee on Statistics to the National Development Council on an improved statistical system. Under the discussion of the recommended establishment of a central computerized file of economic and social statistical data it was stated:

" A statistical information centre as discussed here, if not properly organised and conducted could, quite clearly, pose problems because of illegitimate access to and use of statistics stored in the centre. In arriving at its recommendation the Committee has carefully considered this point and it would not propose that any action be taken to establish a New Zealand statistical information centre unless it could be assured that improper use of confidential information would not result.

It needs to be emphasised that what is in mind is a purely statistical information centre. There is no question involved of the preparation of a computerised file of "dossiers" on individuals or firms containing information collected from administrative sources of Government. The statistical information centre should store only statistical detail obtained from statistical

surveys and exercises and should provide no service beyond what is being done at present by cruder methods, nor should there be any relaxation of access to stored statistics for non-statistical purposes.

This problem has also been given consideration in other western countries, of course, before firm action to establish a statistical information centre has been taken. As an example, in the United States the question of a national statistical information centre has been recently investigated by a task force on the storage of and access to Government statistics. The task force's report, known as the Kaysen report (see 'American Statistician', June 1969) proposed that an inventory of statistical data in the Federal system be established along the lines proposed by the Technical Committee, with the same limitations of function and the enforcement of uniform disclosure standards.

Statistical information centres exist in a number of western countries and have proved their value in practice. The Committee considers that such a centre should be established in New Zealand as a second stage extension by the Department of Statistics. Only the Department, with its statutory function limited to statistical work and with its statutory duty to preserve confidentiality of collected statistical detail and to use it for only statistical purposes, can be considered as the operator of the centre. No data should be stored in the centre other than data collected for statistical purposes and the present stringent provisions of the Statistics Act 1955 relating to the maintenance of confidentiality of statistical detail and the limitations on its use to statistical uses must apply. The Department of Statistics is transferring all its computer processing to the Government Computer Centre. Only computer operators on the staff of the Department can handle magnetic tapes and disks used for the purposes of the statistical information centre. Precautions will have to be built into every EDF system used by the statistical information centre to prevent unauthorised attempts to gain access to the statistical data held by the centre."

833

11 - TANZANIA

The general question of confidentiality especially as involved
in the use of computers

1. At the state owned computer confidential files are not specially protected, neither for private persons, trading companies or government institutions. However, no files are left out to persons not authorized for the project in question, without a written consent from the owner of the file. A locked safe is available in the magnetic tape library room, but so far nobody has used it to keep confidential tapes in.

CONFIDENTIALITY OF BUSINESS STATISTICS

1. The practices described in this paper relate generally to information collected under the 1947 Statistics of Trade Act. The Business Statistics Office safeguards information supplied voluntarily just as securely as information supplied under the Act and hence the practices described can be taken to relate to all information the Business Statistics Office collects.

The Statistics of Trade Act, 1947

2. The prohibition of disclosure of information collected from individual businesses is contained in Section 9(1) of the Statistics Trade Act, 1947, as follows:

"No individual estimates or returns, and no information relating to an individual undertaking, obtained under the foregoing provision of this Act shall without the previous consent in writing of the person carrying on the undertaking which is the subject of the estimates, returns or information, be disclosed ..."

Section 9(5)(b) further provides that:

"in compiling any such report, summary or communication the competent authority shall so arrange it as to prevent any particulars published therein from being identified as being particulars relating to any individual person or undertaking ..."

This provision is reinforced by Section 9(5)(a):

"No such report, summary or communication shall disclose the number of returns received with respect to the production of any article if that number is less than five."

3. An exception to these general prohibitions is that the total quantity or value of any articles produced, sold or delivered may be reported after taking account of any representations made by any person directly affected by the disclosure Section 9(5)(b).

Practice in the Business Statistics Office

4. The Business Statistics Office is concerned with the compilation of statistics in large scale enquiries like the census of production and distribution; and with monthly and quarterly inquiries about the production of different commodities and cognate matters. The statistics obtained are published as reports and in articles in e.g. the Board of Trade Journal. In the major Business Statistics Office inquiries, three kinds of possible disclosure are distinguished:

a. Primary disclosure is in respect of key items for which the number of contributors is shown such as total goods produced

or sold, work done, net output, employment, etc. (i.e. most of the aggregates appearing in Table 1 of the 1963 Census of Production reports, and corresponding items in Census of Distribution Reports), together with specific items of goods produced or sold. Disclosure occurs if there are entries against these items for less than three enterprises (the establishment details being accumulated on an enterprise basis for assessment). It may occur if there are three or more enterprises concerned, and one or two of them held a dominant position in respect of the item under consideration.

b. Secondary disclosure concerns non-key items for which the number of entries is not shown and cannot be accurately deduced, such as employers' contributions to private pension funds, new building work etc., specific items of goods produced or sold where there are separate figures shown with and without quantities, and non-specific items such as "other materials for processing" or "other sales". In these cases disclosure occurs if there are entries for one enterprise only or if one enterprise has a dominant position.

c. "Disclosure by difference" arises where a figure has been published, and a subsequent analysis produces a figure differing from it only to the extent of an amount attributable to one or two enterprises. The clearest cases occur where the boundary of a geographical area is altered between the time of one tabulation and another; or where a tabulation for one area (e.g. a standard region) overlaps that for another (e.g. a television area).

Effect on Publication

5. The Statistics of Trade Act forbids the disclosure of "information" relating to a particular undertaking (see para 2). Hence (quite apart from the constraint on publishing figures relating to a firm's activity even though the firm itself is not identified) there is an embargo on the dissemination of lists of firms or establishments, showing the heading of the Standard Industrial Classification to which they have been classified on the basis of information in statutory returns. Information of this kind is much in demand. Because of this, the permission of respondents to the 1968 Census of Production is being sought for the inclusion of their names and addresses in published lists of firms contributing to the results for different industries. No other information about them will be published. In some of the Business Monitor Production Series of publications, lists of participants in the inquiries have been published but covering only those firms who had not objected to their names being included.

6. The prevention of disclosure accounts for a substantial part of the procedures involved in handling a statistical inquiry. So far as the Census of Production 1963 was concerned for instance, disclosure exercises accounted for perhaps one-tenth of the total costs of the census. The proportion of the costs of clearing special tabulations is usually a very much higher proportion of the total costs of such tabulations. As has been indicated, it is not sufficient to examine tabulations of the returns as they are received. Disclosure must also be considered in terms of enterprises,

and returns must be grouped accordingly, and as activity becomes concentrated in larger enterprises, the amount of information which must be suppressed increases. (Incidentally, the inquiry register must in consequence contain adequate information about the inter-relations of establishments and their parent companies.)

7. The suppression of figures involving disclosure is easy to handle where broad aggregates are concerned, like work-in progress, or sales of plant which can be included in figures of total stocks or deducted from purchases of new plant to produce a figure of net investment. It is much more difficult to suppress items relating to individual commodities. Only in some cases can the sensitive item be included in a total of, for instance, "all other sales" or a more precisely described residual heading. Often it is better to combine the item with another of a similar nature but of a different size, or material, or use. Much depends upon the relative size of the figures, since the combined figure must of course still meet the disclosure criteria. The choice between different alternatives often presents considerable difficulty, and judgment must be exercised in the light of technical knowledge and the advice available to the office. Whatever the conclusion, the effect is of course that a good deal of useful information which is clear of disclosure has to be hidden together with the sensitive items with which it is combined. Consideration is being given to the possibility of adapting the practice followed in publishing figures for different kinds of business in the field of retailing, where disclosure items are not combined with others, but additional items are suppressed so that the disclosure figures cannot be deduced. A similar procedure in handling output data is followed by the United States Bureau of the Census.

8. Another factor which must be borne in mind in the suppression of the primary data is the likelihood of embarrassment at a later stage in producing other analyses because of the risk of disclosure by difference. Combining figures for the purpose of an industry analysis may make it less or more likely that a subsequent geographical analysis will produce a disclosure situation. The Business Statistics Office does not normally consider itself obliged to ascertain what alternative sources of information may be available to readers of reports. For example, market research undertaken by a firm may enable it to deduce particulars from published aggregates which have otherwise been carefully hidden. Only in an exceptional case would the Business Statistics Office take into consideration, in settling its own procedure, special intelligence of this kind.

9. The risk of disclosure by difference, combined with the fact that the statutory prohibitions are not subject to any limitations as to time, make it increasingly difficult to produce new analyses of past statistics. Occasionally, indeed an amendment has to be suppressed because the change from the previously published figures represents the addition of information for a single organisation, but this kind of case is rare. Far more serious is the problem presented by a revision of industrial classification or some of the regional or local authority boundaries, which may involve such fine adjustments that comparative figures for past periods cannot be produced to set alongside new data collected and analysed according to the new definitions.

10. In checking for disclosure the circumstances which are taken into account are normally those which were relevant in the period to which the

data relate. Between the end of the period for which returns are received from firms and the date of publication of the statistics obtained from them, changes inevitably take place, and amalgamations may result in new disclosure situations. Changes of this kind are ignored in preparing tables.

11. Ratios derived from the basic data collected are often regarded as not involving any risk of disclosure since "individual undertakings" cannot be identified. Thus a single figure of net output per head, obtained from census data about sales, purchases, stock movements, employment, etc. could not be traced to any particular business. However, if a set of ratios of this kind (others being, for example, wages/sales, stock/sales, capital expenditure/employment and so on) were known to relate to a single organisation, there might be some possibility of that organisation being identified. Hence, in an article published in the Board of Trade Journal of 16 February 1968 a series of median ratios were shown, each of which related to a different organisation, although a series of ratios for a single median organisation (according to some criterion) would have been more illuminating.

Application of Disclosure Procedures

12. The suppression of data which involves the risk of disclosure is carried out as standard procedure in all cases except where the figures relate to the sales or purchases of specific commodities. In all these cases the firms concerned are approached for their agreement to publication; and this is readily forthcoming in the majority of cases. However, the items for which agreement to publish is withheld are numerous and often of considerable importance. Different departments of an organisation may have different views about publication, so that an objection to publication may result from a careful balance of opinion.

13. An example of the practice described in the previous paragraph is that a special approach was made by the Board of Trade to the chairmen of about 60 large companies in anticipation of the publication of industry reports on the 1963 Census of Production. The items for which clearance was sought had been subject to refusal of agreement to publication in earlier census reports, and collectively accounted for over £400 million of output. As a result it has been possible to publish for the first time particulars of more than £350 million of this output, as well as of a large output which would have had to be hidden as a consequence of suppression of the sensitive data.

14. The reasons given by firms who make representations against the publication of information relating to their business almost invariably concern the competitive situation in which the firm concerned is placed. Most often the fear is of the advantage which would accrue to a direct competitor. In many cases there is fear of potential rather than actual competitors, particularly overseas. Here price differentials may be of critical importance, so far as they can be deduced. Occasionally other aspects are stressed, such as the relative labour content of output in this country as compared with others. Frequently it is argued that information about United Kingdom output should be suppressed because corresponding information for other countries is suppressed abroad.

Requests for Information

15. Against this background and with these statutory obligations, the

Business Statistics Office must refuse to meet requests for information which would disclose information supplied by individual firms. The requests come from bodies like the National Board for Prices and Incomes, (now the Commission for Industry and Manpower) Industrial Training Boards, local authorities and planning organisations, trade associations, business undertakings (including market researchers) and from university and other research workers. The most frequent type of request is for a list of firms classified to an industry or which make a given product. Since the aim is often to conduct a survey, this sort of enquiry often is not permissible for the Business Statistics Office to reveal information of this kind. Another frequent type of request falls in the field of distribution activities. Market researchers wish to know the number of outlets for given commodities in particular areas; and planning authorities often wish to have a description of the structure of particular shopping areas. Detailed analyses of this nature often show that there is only one outlet or type of organisation in a small area, and this information is not revealed.

16. Among the requests for information from the public which the Business Statistics Office has had to refuse on confidentiality grounds in recent months are:

frequent requests for breakdowns of published commodity information which would involve disclosure;

a publisher of trade and other directories wanted a list of larger establishments in selected industries;

a consultancy firm wanted a list of manufacturers by area and size;

a manufacturing company wanted the names of the six largest manufacturers of a product in which its own output was an important element;

a market research firm wanted more detailed information about size and form of retail organisation than is published so that they could design their own sample survey;

a local authority requested information about kinds of businesses, with their turnover and number of persons employed, in various shopping areas within its boundaries;

a trade association asked for a list of traders operating in the field covered by the association.

17. While the Business Statistics Office in no circumstances provides the public with a list of firms, it occasionally helps official agencies to conduct their own inquiries by sending to the firms it has classified to the field in which an enquirer is interested a questionnaire (or other document) covered by a letter of explanation from the Office.

Confidentiality and the Computer

18. The question of confidentiality as involved in the use of computers is only beginning to be explored in this country as far as the Government Statistical Service is concerned. Hitherto computers have been used to process and store information which has only been made directly available to statistical sections of other Government departments. They have also been employed in identifying primary and secondary disclosure as described

in paragraph 4 prior to the publication of data, but the task of dealing with disclosure situations remains largely one of human judgment. With the introduction of data banks and other sophisticated, and direct methods of disseminating information a system will have to be devised which incorporated efficiently and mechanically the procedures laid down in the statutes. This consideration will have to be borne in mind when the question of what should be included in data banks is discussed, and this is still very much an open question as far as the United Kingdom Government Statistical Service is concerned.

19. The whole question of the privacy and confidentiality of information about individual persons and businesses is being examined in a wider context by the Younger Committee which has been directed by the Home Secretary to review the situation in the private sector. Coupled with this, the Prime Minister has undertaken to examine a suggestion that "a code of conduct" should be established in the use of computers within central government for dealing with personal records. It is currently intended that this exercise should also take account of information about individual businesses.

20. Ideas are being exchanged and examined on an international front also and, in fact, the Younger Committee might be seen to have resulted from a report in January 1970 by Justice (British Section of the International Commission of Jurists) on Privacy and the Law, since its proposals were introduced into the parliamentary debate. The Council of Europe has also discussed the problems of confidentiality, and the Organisation for Economic Co-operation and Development is currently carrying out a "Computer Utilisation Study" within member countries which embraces among other things the legal and technical aspects of security with regard to data banks.

21. The position in the United Kingdom - and in many other countries - is therefore that the question of confidentiality, including its significance in relation to computers, is receiving considerable attention, but the situation remains, as yet, in the exploratory stage.

INTERNATIONAL TRADE STATISTICS AND THE SITC

Introduction

This paper on international trade statistics and the Standard International Trade Classification (SITC) discusses some topics of interest in relation to the current revision of the SITC under the following four headings:

- (a) Need for revision of the SITC - reviewing the case for a policy of regular revision of the SITC;
- (b) Relationship with the Brussels Tariff Nomenclature (BTN) - reviewing the present relationship and discussing possible means of overcoming some of its less attractive features;
- (c) Differing classification requirements for imports and exports - outlining the difficulty involved in applying the same degree of detail in export statistics as in import statistics and suggesting a possible solution for incorporation in the SITC; and
- (d) Structure of the SITC - describing a way along which it may be profitable to proceed in the revision of the SITC structure, and reviewing some of the possible objections to revision of the structure.

(a) Need for Revision of the SITC

2. Following the first revision of the SITC in 1960, when the one-to-one relationship with the BTN was introduced, the first specific move (known to Australia) for further revision of the SITC came in 1966 when the Organisation for Economic Co-operation and Development (OECD) suggested a revision of Division 64 - Paper, paperboard and manufactures thereof. The United Nations Statistical Commission, at its Fifteenth Session in 1968, considered that the proposal to revise the classification of paper and paperboard should be deferred pending review of the entire structure of the SITC. Since that time a number of other areas of the SITC have been suggested as possible areas in which review and revision might profitably be carried out. (1) These areas include the following:

Section 0, Food and live animals
 Division 67, Iron and steel
 Division 71, Machinery, other than electric
 Division 72, Electrical machinery, apparatus
 and appliances; and
 Group 735, Ships and boats

3. In addition, there have been suggestions for revising the SITC in other ways (1), including revisions of the broad structure and of the degree of detail at the lower levels. With regard to the revision of the broad

(1) The suggestions were made at the meeting of the group of experts held in November 1969 to consider general guidelines for the revision of the SITC.

structure of the SITC there has been a wide range of suggestions. One suggestion is that the broad structure of the proposed Classification by Broad Economic Categories (BEC) should be adopted as the broad structure of the SITC. Other suggestions propose the transfer of individual categories from one part of the SITC to another. These suggestions, if implemented, would have the general effect of sharpening distinctions in the SITC structure between food and non-food categories, between categories of materials and finished goods, and between categories of primary and processed materials (and thereby, incidently, increasing the correspondence between the BEC and SITC structures). Suggestions for review of the degree of detail at the lower levels of the SITC have arisen as a result of difficulties encountered in a number of countries in implementing the full range of SITC 5-digit items for export statistics and of the recognition that the strict one-to-one relationship with the BTN has led to excessive fragmentation of detailed items in the SITC (see paragraphs 7 and 8).

4. The need for regular periodic review of the SITC was expressed at the Fifteenth Session of the United Nations Statistical Commission. There is probably general agreement that the SITC, along with other international economic statistical classifications, is subject to obsolescence, arising from changes in patterns of economic development and changes in the commodity content of production and trade due to technological development. It is presumably also agreed that an element of stability is desirable in a classification, in order to give it value for presenting a series of statistics over time. Experience with international classifications seems to suggest that a period of about ten years between revisions provides an acceptable balance between obsolescence and stability, and such a period was suggested at the Fifteenth Session. The point was also made at the Fifteenth Session that it is desirable to plan revisions of the SITC well in advance. This is particularly important because of the widespread use of the SITC and the relationship between it and the BTN (which may necessitate corresponding changes in the BTN). A desirable means of planning revisions may be to adopt a policy of regular decennial reviews of the SITC.

(b) Relationship of the SITC with the BTN

5. As is generally recognised, the one-to-one relationship between the SITC and the BTN is of great benefit because it, firstly, makes the precision of definition achieved by the BTN applicable to the SITC and, secondly, provides an internationally agreed system whereby statistics compiled according to the BTN are convertible to the SITC, and vice versa, if required. This one-to-one relationship is between the 1,312 basic items of the SITC and the 936 (of the 1,124) headings of the BTN, plus the 376 statistical sub-headings of the remaining 188 headings of the BTN. In other words, each basic SITC item is identical with either a heading of the BTN (when the heading is not divided into statistical sub-headings) or a statistical sub-heading of a heading of the BTN (when the heading is divided into two or more statistical sub-headings).

6. Although there is great advantage in having the one-to-one relationship between the SITC and the BTN, this has necessitated the establishment of many detailed items in the SITC, some of which may not be of great significance in international (and, particularly, in national) trade,

and may not even be of appreciable or general economic interest. In the case of some countries, this would have caused problems in collecting and, consequently, in presenting export statistics classified according to SITC (e.g. where trade in these items has been negligible). These kinds of problems were encountered in Australia.

7. At the meeting of the Group of Experts, held in November 1969 by the United Nations Statistical Office to consider general guidelines for revising the SITC, it was recognised that departure from the strict one-to-one relationship between the SITC and the BTN was possible while still allowing the precision of definition of the BTN to be applied to the SITC. At the meeting it was agreed that the Statistical Office would examine all SITC items which account for a very small proportion of total world imports and exports with a view to amalgamating two or more such items or combining such items with other more important items. This would be done in such a way that a resulting broader item corresponds to either two or more headings of the BTN or two or more statistical sub-headings of the one heading of the BTN (thereby obviating the necessity for the two or more separate statistical sub-headings).

8. The possibility of defining an item of the SITC to correspond with either (i) one or more headings of the BTN together with one or more statistical sub-headings of other headings of the BTN or (ii) two or more statistical sub-headings drawn from different headings of the BTN, however, was not given general endorsement by the Group of Experts. Nevertheless, in Australia it is considered that the process of defining an item of the SITC in this way differs only in degree but not in principle from defining an item to correspond with two or more headings of the BTN, and would allow much greater flexibility in establishing items in the SITC while still maintaining the present important link with the BTN for definitional purposes. This aspect merits further consideration during the current review of the SITC.

(c) Differing Classification Requirements for Imports and Exports

9. Although the basic items of the SITC bear a one-to-one relationship with the headings and statistical sub-headings of the BTN, which is designed as a classification for customs tariffs purposes, the SITC is intended for use in export as well as in import statistics. It is therefore useful to consider the differing classification requirements for import and export statistics.

10. The compilation of import statistics according to the SITC is made easier by the correspondence between detailed categories of the SITC and of the BTN, at least for the high proportion of countries whose tariff systems are based on the BTN. In the case of Australia, the customs tariff includes not only the 1,124 headings of the BTN, but also the statistical sub-headings which, when applied to the BTN, bring it into complete correspondence with the 1,312 items at the 5-digit level of the SITC. The Australian Import Commodity Classification (AICC) is based on the SITC, and the 5-digit level of the AICC is identical with the 5-digit level of the SITC. Thus there is no serious practical problem at present in compiling import statistics for Australia on a precise SITC basis.

11. The application of the SITC in the compilation of export statistics does, however, pose problems of some concern. Australia is one of the countries which, in United Nations reports relating to the International Trade Statistics Centre, is described as reporting "essentially at the 5-digit level" of SITC. This statement is true in a sense. In devising the Australian Export Commodity Classification (AECC), Australia went to great lengths to incorporate in the classification every 5-digit SITC item for which it appeared that there might be a significant volume of exports and for which it would be practicable to obtain adequate descriptions of goods on customs export entries for the purpose of identifying SITC items. It was found that in the AECC Australia could adopt 847 of the 1,312 SITC 5-digit items. However, it was found that for the remaining 465 SITC items either the volume of trade did not justify adoption of these SITC items separately or that it was impracticable to implement the SITC distinctions in the AECC because of difficulties in collecting exports data in the required detail. Accordingly, these 465 items were combined to form 116 more broadly defined items at the 5-digit level in the AECC. A number of these 116 5-digit items have, of course, been further subdivided on the basis of characteristics considered more appropriate for national purposes. The impression in Australia is that, where other countries have attempted to base their export classifications on the SITC, problems of a similar character have been encountered. For example, it is understood that in the U.S. Export Statistics Classification SITC groups are identified fully at the 3-digit level, but that a proportion of the 4-digit and 5-digit items has been re-defined - apparently in much the same way as it was necessary to re-define 5-digit SITC items for the AECC. It would therefore seem that from the point of view of achieving full acceptance as an international standard classification for use in compiling export statistics the present SITC is too detailed, in some areas at least.

12. In the current review of the SITC, more emphasis should, consequently, be given to the problem of devising a classification which will be suitable for export as well as imports statistics. The problem involved in devising such a classification is (as implied above) that a classification suitable for export statistics may not be sufficiently detailed and, therefore, suitable for the compilation and presentation of import statistics, and for the maintenance of the link between the SITC and the BTN.

13. This problem might be resolved by establishing a "basic" level in the revised SITC according to which all countries could report both import and export statistics in the full detail of this level; below this "basic" level a more detailed level could then be established for countries to report import statistics only. In designing the "basic" level, specific attention would need to be given to the practical problems encountered by countries in reporting export statistics according to an international standard classification. Naturally, some difficulties would have to be overcome to obtain a consensus on what would be practicable for most countries, but a method such as this could do much to enhance the acceptability of the SITC for use in export statistics. Once this basic level had been devised, it would be extremely valuable if all countries were to adopt this level fully, not just "essentially", for exports and imports in their reporting to the International Trade Statistics Centre and, wherever practicable, for national purposes. In what follows, this suggested level is referred to as level A,

the most detailed level of the suggested "basic" revised SITC. If a greater degree of detail in import statistics should be required internationally, the "basic" revised SITC could be subdivided to provide a more detailed level (approximating perhaps the detail at present 5-digit SITC level) for use by countries in reporting import statistics to the International Trade Statistics Centre on a standard basis, and perhaps also for use in national import classifications. This more detailed level is referred to below as Level B. Thus Level A would represent the most detailed level of the "basic" revised SITC to provide an international standard for both exports and imports which could be fully adhered to by all countries, while Level B would provide more detail for imports only - full adherence to Level B would possibly be not as essential as to Level A. Both levels should be defined in terms of BTN items. A possible method of achieving this would be to have a precise correspondence between the detailed headings of the BTN and the items at Level B. The items at Level A would then be defined in terms of specific combinations of the BTN detailed headings. This may not be the only solution which would meet the competing objectives, but it presents a possible line of approach, in conformity with the views expressed by the United Nations Statistical Commission. In the Report of the Fifteenth Session, it is stated: "Substantial problems (in revision of the SITC) were envisaged, particularly with regard to the correspondence which at present exists between the SITC and the Brussels Tariff Nomenclature (BTN). The commission considered that this correspondence remained extremely desirable and that every effort should be made to maintain it ... The commission also noted the fact that, in the event that a change of the BTN could not be expected, it would be necessary to re-examine the value of maintaining the strong one-to-one correspondence between the two classifications, because it was felt that the alternative of an excessive fragmentation of the SITC would greatly hinder its operation as an international classification." (paragraph 127, page 39)

(d) Structure of the SITC

14. Although the structure of the SITC provides commodity groupings of general economic interest (i.e. the sections, divisions and groups) it has been found desirable by a number of countries to distinguish broad economic categories of commodities of imports and exports in addition to the sections of the SITC. In 1964 the Statistical Commission recommended that data on broad economic classes of commodities of imports and exports, distinguishing food, industrial supplies, capital equipment, consumer non-durables and consumer durables, should be compiled to supplement the summary data already compiled on the basis of the sections of the SITC.

15. Development of these broad economic classes, now known as the Classification by Broad Economic Categories (BEC), has proceeded since that time. In the course of this work, it has become clear that it would be desirable to include in the BEC additional categories to those recommended in 1964. The development of the BEC has proceeded with the objective of evolving categories to provide meaningful aggregates for economic analyses of the usual uses to which goods are put, in such a way that the categories, as far as practicable, can be aligned with the basic classes of the SNA, namely, capital goods, intermediate goods and consumer goods. The BEC is to provide

for conversion of data originally compiled according to the SITC which, as it stands, is not suitable for this type of analysis.

16. The degree of interest expressed in the BEC in international discussions suggests that the broad aggregates compiled according to the BEC could be of greater use than the totals compiled according to the sections, etc., of the present SITC. If this is so, there would seem to be a case for changing the present SITC structure so that it will correspond more closely with the broad categories of the BEC. There would be a number of different ways of achieving a closer correspondence; three possible ways are described below.

17. Firstly, there is a possibility of adopting the BEC structure as the broad structure of the SITC. If this were done, the two classifications would, in effect, become one, and the composition of the BEC categories would be automatically identifiable in terms of the more detailed categories in the SITC. It could be questioned, however, whether all of the categories of the BEC could be implemented in the SITC without detracting from the usefulness of some of the lower level SITC groupings. For example, it might be thought that the implementation of the distinction "for industry"/"for household consumption" would unduly fragment existing SITC groups and could introduce an element of conventionality into some parts of the SITC beyond that which might be considered acceptable. Similar considerations might apply to the distinction between "durable"/"non-durable" in the BEC consumer goods category. Furthermore, there would appear to be advantages in maintaining the present SITC practice of arraying "parts" with (or near to) the relevant equipment categories - this consideration would argue against introducing the "Parts and accessories" categories, as recognised in the BEC, at a broad level in the SITC. If the BEC structure were to be adopted as the broad structure of the SITC, the kind of SITC structure that would result is illustrated in Table A below.

TABLE A

Section O. Food and Beverages					
Division	Major Group	Group	Sub-Group		
01	011	0111		Food and Beverages - Primary	
				Food and Beverages - Primary - For industry	
				Live animals	
			0111.1	Bovine cattle (including buffaloes) (001.1)	
			0111.2	Sheep, lambs and goats (001.2)	
			0111.3	Swine (001.3)	
			0111.4	Poultry, live (001.4)	
			0111.9	Live animals (chiefly for food) n.e.s. (001.9)	
			0112	0112.0	Wheat (including spelt) and meslin, unmilled (041)
			0113	0113.0	Rice in the husk or husked, but not further prepared (042.1)
			0114	0114.0	Barley, unmilled (043)
			0115		Cereals, unmilled, other than wheat, rice, barley, maize and oats
				0115.1	Rye, unmilled (045.1)
				0115.9	Cereals, unmilled, n.e.s. (045.9)
				.	
	.				
	.				
	.				
	.				
02	021		Food and Beverages - Processed		
			Food and Beverages - Processed - For industry		
	0211	0211.0	Meal and flour of wheat or of meslin (046)		
	.				
	.				
	.				

Some of the possible reservations, expressed earlier in this paragraph could probably be overcome by amalgamating or otherwise rationalising some of the fragmented old SITC groups into new SITC groups. For example, in the table, Group 0113 could be deleted and shown as a sub-group in Group 0115.

18. Secondly, there is the possibility of adopting only the broadest level categories of the BEC as the broad structure of the SITC (that is, using these categories - with some minor modifications to preserve as far as possible existing whole SITC divisions and groups - as the sections of the SITC), so that the composition of the broadest categories of the BEC is explained through the hierarchic structure of the SITC. Thus BEC category "Food and Beverages" would become a new section in the SITC which would be based mainly on existing Sections 0, 1 and 4. The groups within this new section would not be re-ordered within lower level BEC categories but would be left essentially in the existing order with, where possible, the existing

division, group, etc., numbers. The lower level categories of the BEC (e.g. those categories which distinguish "primary" and "processed" commodities separately) could then be defined in terms of lower level categories of the SITC by designating each lower level category of the SITC as being appropriate to a particular lower level category of the BEC. For instance, the distinction between "primary" and "processed" might be introduced in a section at, say, the group level by designating each group wholly as being either "primary" or "processed". However, such an approach could mean accepting a greater degree of conventionality in what is being designated, for example, as primary or processed than has been accepted in the BEC if, say, an SITC group is designated primary but contains an item which is included in a processed category in the BEC. Alternatively, the composition of SITC groups could be revised so that the degree of conventionality in the SITC is the same as in the BEC. A separate BEC would not be needed if this approach were followed but annotations indicating the designation of groups as primary or processed, etc., would be required. The result on the structure of the SITC of following this approach has been illustrated in Table B.

TABLE B

Section O. Food and Beverages			
Division	Group	Sub-Group	
00	001		Live Animals
			Live Animals - (Primary)
		001.1	Bovine cattle including buffaloes
		001.2	Sheep, lambs and goats
		001.3	Wine
		001.4	Poultry, live
		001.5	Horses, asses, mules and hinnies
		001.9	Live animals (chiefly for food) n.e.s.
01	011		Meat and Meat Preparations
			Meat, fresh, chilled or frozen - (Processed)
		011.1	Meat of bovine animals, fresh chilled or frozen
		.	
		.	
		.	
	012		Meat, dried, salted or smoked, whether or not in airtight containers - (Processed)
	.		
	.		
	.		

This approach would have the advantage of retaining much of the existing lower level SITC structure and of having similar commodities presented contiguously in arrays of data for the section, while providing the facility to array data for the section in such a way as to indicate the composition of the BEC categories such as "primary" and "processed". For example, commodity

categories could be arrayed in tabulations in SITC order but with data for "primary" and "processed", etc., groups presented in separate columns. Other BEC distinctions, e.g., "for industry" and "for household consumption" could be similarly provided for at, say, the sub-group level.

19. Thirdly, there is the possibility of retaining the two separate classifications, the SITC and the BEC, but of amending the SITC with the aim of obtaining easy conversion between the two (this approach differs from the other two in that in the previous approaches the structures of the two classifications were to be integrated into one structure). This approach envisages the minimum possible disturbance to the existing SITC structure and numbering system consistent with making as many SITC groups, divisions, etc., correspond to, or fit into, BEC categories (i.e. the aim would be to key as many as possible whole SITC sections, divisions and groups to relevant categories of the BEC). Amending the SITC with this aim in mind would involve transferring certain SITC items, sub-groups, etc., from one section or division to another (and possibly the creation of some new SITC sections) and the preparation of a key which would link the two classifications. In this approach, as in the other two, it is assumed that changes to the SITC items (that will be made during the review of the SITC) will include changes to those SITC items for which trade was to be apportioned among several BEC categories, i.e. these items would be split so as to fit them more readily into BEC categories without the need to apportion trade in them.

20. Of these three possibilities of changing the SITC structure, the third would entail least disturbance to the present structure of the SITC. For this reason it could be thought to be the most desirable choice. It could also be thought the least satisfactory of the three possibilities discussed in the sense that the simple direct link between the detailed SITC items and the BEC categories through the hierarchic structure of the SITC, as envisaged in the first two instances, would not be available; the link between SITC categories and BEC categories would be an indirect one in this case involving aggregation of SITC categories to form BEC categories by means of a key. Moreover, two separate classifications would be required in place of one. Thus, if it is considered desirable to avoid the proliferation of different classification structures either one of the first two possibilities would appear preferable to the third. It might be thought, however, that in this computer age it will be possible to re-arrange data in accordance with different classification structures without too much trouble and that considerations of structure would therefore not be important. Nevertheless, one has to settle for one classification structure as the main basis for arraying detailed statistics and hence it would not be true to assume that questions of classification structure are not of importance.

21. Revision of the structure of the SITC would involve revision of the numbering of SITC categories. It has been suggested, however, that revision of the structure of the SITC should not be undertaken, because the main requirement in the structure of the SITC is continuity with the past and that items in the SITC should, as far as possible, carry the same code numbers after the current review as they carry in the present SITC. However, leaving aside consideration of changes in the structure of the SITC, there may be considerable changes in the code numbers of SITC categories as a result of the current review of the SITC for four possible reasons, namely:

- (i) the amalgamation of items which account for a relatively very small proportion of total international trade (mentioned in paragraph 7);
- (ii) subdivision of items which account for a relatively large proportion of total international trade;
- (iii) transfer of categories from one part of the SITC to another (mentioned in paragraph 3); and
- (iv) elimination of certain inefficiencies in the present numbering system.

This last reason arises when countries which use the SITC as the basis for their own national classifications find that the numbering system provides, at the most detailed level of the SITC, insufficient scope for expansion of the detailed level for national purposes. For instance, Australia has experienced difficulty in accommodating all the desirable dissection at the 7-digit level in the Australian Import Commodity Classification (AICC) of SITC item 581.20 (polymerization and copolymerization products), and would want to see this SITC item (and others) subdivided into more than one item in the revised SITC.

22. Further, in national trade commodity classifications it is common to have a considerable degree of change in items and item numbers at the most detailed level of the classification from year to year, as trade patterns are modified or as additional detail is provided to meet expanding demand for information, and Australian experience has been that users have little difficulty in understanding these changes (it is, of course, important to provide a key between new and old items where changes have occurred). In the previous revision of the SITC, a considerable degree of change in item numbers was accommodated. Moreover, it should be fairly simple, when necessary, to convert data according to the old SITC to a new framework during an interim change-over period, utilising an appropriate conversion key. This would be facilitated to some extent, at least, by the effects of amalgamation of less important existing SITC items, and subdivision, in the main, of only those items which are important in international trade - this should make it easier to convert data from the existing SITC to the proposed revised SITC than was the case with the previous revision when large numbers of existing items were subdivided.

23. Consequently, in view of the probability, firstly, that there would be considerable changes in the code numbers of SITC categories resulting from reasons other than structural changes; secondly, that these changes can probably be fairly easily accommodated by users of the SITC; and thirdly, that it would be relatively simple to convert data from the present SITC to the proposed revised SITC by means of a key, it seems that the resulting revision of SITC code numbers should not be taken as a substantial objection to revision of the SITC structure. Therefore, if it is accepted that the aggregates compiled according to the BEC are of much greater general and economic significance than the statistics of international trade summarised for the existing sections, etc., of the SITC, it would seem to be desirable to adopt the broad categories of the BEC as the broad categories of the SITC.

INTERNATIONAL TRADE STATISTICS AND THE SITC

1. The United Nations Statistical Office is engaged in a program for revision of the SITC in association with the Brussels Tariff Co-operation Council with a view to maintaining the one-for-one concordance between the SITC and BTN if at all feasible. The present is an appropriate time, therefore, to review the existing and probable future rôle of the SITC not only as a classification for international trade statistics but in the wider context of commodity statistics generally.
2. Canadian international trade statistics are compiled and published on the basis of National trade commodity classifications derived from Canada's Standard Commodity Classification (SCC). The SCC was developed by the Dominion Bureau of Statistics as a vehicle for integrating commodity statistics derived from different sources (e.g. imports, exports, production, freight) within Canada. For purposes of reporting to the United Nations and other international agencies as well as for making direct international comparisons of trade figures, a convertibility list has been established by which Canadian trade statistics can be re-grouped on the SITC(R). This is done directly on the computer.
3. Recently, a program for more detailed comparison of the Canadian Customs Tariff with the BTN has provided material that can be used to improve the convertibility list between the SCC and the SITC. The improvement comes through cross-classification of Canadian import statistical classes by Canadian tariff items to provide detailed categories that can be fitted quite precisely into BTN classes. A by-product is more precise ad hoc conversion of import data to the SITC. Export conversions are not affected but these are being studied with a view to improvement.
4. When comparisons are made between the Canadian SCC and the SITC the different rôles and objectives of the two systems should be taken into account. The SCC is a "general purpose" commodity classification designed for national use while the SITC has been designed especially for international comparisons of international trade statistics. Each system, of course, is intended to provide data for economic analysis but the range of analyses contemplated is different in the two cases. In addition, the different fields of application of the two systems lead to different approaches to the grouping of commodities. The classification systems are therefore not competitive but complementary.
5. Considering only international trade statistics, an internationally accepted standard classification cannot be expected to serve satisfactorily as a national classification without modification, due to the marked variations in commodity trade patterns experienced by different countries. Modification of an international standard classification for national use generally involves subdivision or aggregation of its most detailed categories, the basic structure of the international classification being taken as given. Linking the international standard classification for trade with the international customs tariff categories (i.e. the SITC with the BTN) makes the situation more complicated because the definition of items is dictated by tariff considerations.
6. Past Canadian experience with a national classification system for trade statistics linked closely with the tariff classification is not reassuring. The structure, definitions and concepts of the customs tariff are dictated primarily by considerations of revenue and/or protection of industries

Considerations of economic analysis tend to be assigned a lower priority. The regrouping of tariff-dominated categories into larger groupings more suited to economic analysis does not change the nature of the basic statistical items. This is particularly relevant when the statistical classification for trade is based upon a national standard commodity classification used also for statistics of production, industrial materials and transportation (freight). The Canadian SCC employing criteria most generally applicable to economic analysis in a variety of statistical series on commodities is therefore a purely statistical classification system divorced from tariff considerations.

7. Those who designed the SCC were aware that a commodity classification system designed for general use in a variety of fields could not be specialized in terms of any particular data collection system or pattern of data use. Such a classification structure must be based upon principles that will endure in the field of commodities despite inevitable changes in commodity characteristics due to new materials and technology as well as changing patterns of commodity content. The principal criteria used for grouping in the SCC are stage of fabrication, chief component material, and purpose. The broadest groupings are five in number, being: Live Animals; Food, Feed, Beverages and Tobacco; Crude Materials; Processed Materials; End Products. With the exception of Live Animals and Food, Feed, Beverages and Tobacco, therefore, commodities in the SCC are grouped broadly on the basis of stage of fabrication. Such a pattern can be expected to endure. Crude materials and fabricated materials are further subgrouped on the basis of chief component material while within the end products section (i.e. fully manufactured articles) the principle of use or purpose is employed.

8. Versions of the SCC have been used to classify Canadian exports and imports since the early 1960's and in commodity freight statistics (water, rail and road) beginning in the mid-1960's. It is planned that by 1971 a version of the SCC will have been introduced for statistics of industrial products and materials. Because of the different requirements of the various statistical fields, versions of the SCC have been developed to suit the particular data collection and use patterns needed. Each of these working classification is made up of combinations of SCC classes and breakouts of a particular SCC class as required. Each can, therefore, produce data on the SCC pattern and in addition a list of some 2,000 commodity groups is being designed especially to obtain comparable data from the different sources. None of the categories on this list crosses SCC class boundaries. In this way the SCC provides a set of basic commodity definitions for all uses and attains its objective as a standard commodity classification system.

9. An international standard classification such as the SITC cannot be subjected to minor revision frequently (annually) as is the practice in most countries for their national classifications for international trade. In fact all "standard" classifications must maintain greater stability than do "working" classifications. This relative "rigidity" of the SITC, arising in part from its "standard" rôle and in part from its international rôle (i.e. amendment requires agreement by a number of countries) in addition to the link with BTN, imposes a special obligation on the United Nations to undertake a thorough, well considered revision program at regular intervals. The present revision is the first of this type for the SITC.

10. DBS experience indicates that the development and effective implementation of a standard commodity classification takes time and requires the cooperation of producers and users of data from a variety of sources. Important among these are the trade statisticians and the users of trade data.

However, commodity statistics of international trade have limited usefulness in isolation and their value is enhanced when they can be used freely in conjunction with other commodity data. The SITC, although designed for international trade, is already being used for international transport statistics in Europe and some of its categories are being adapted for use in classifying industrial materials and products (see E/CN3/405, 25 November 1969, a Statistical Commission paper on the 1973 World Programme of Industrial Statistics). Thus the need for an International Standard Commodity Classification is becoming apparent and the SITC is in a favourable position to assume that rôle.

11. The structure of the SITC should be examined to improve its relevance to current analyses of international trade as well as to facilitate its use for other commodity data. The most glaring deficiency of the present structure for these purposes is the mixing of fabricated materials and end products, especially in Section 6, Manufactured Goods Classified Chiefly by Material, but also in Section 5, Chemicals. Section 6 in particular contains a hodge-podge of materials and articles yielding data that defy analysis. Clear differentiation between materials to be used in further processing (processed materials) and articles ready for use (end products) would provide more stable groupings suited to analytical requirements. The added stability and precision in groupings would also make the classification more applicable for general use.

12. The necessary restructuring of the SITC could be accomplished with minimum disturbance of the existing sections and groups by the following kind of rearrangement:

Eliminate two relatively small sections, i.e. Section 1 Beverages and Tobacco; and Section 4 Animal and Vegetable Oils and Fats; combining Beverages and Tobacco with the present Section 0 Food and Live Animals, and including Animal and Vegetable Oils and Fats with those groups in present Section 6 that contain processed materials to form the basis for a new Section 6 Processed Materials n.e.s.

The groups and sub-groups in present Section 2 Crude Materials, inedible, except fuels, made up of unmanufactured goods should be converted to a Section for Crude Materials.

The balance of Section 2 should be included in the new Section 6 Processed Materials n.e.s.

Section 5 Chemicals should be confined to a Section for Chemical Materials.

Other Sections and the end products from Section 5 and 6 would then form the basis for suitable sections of End Products.

(It would be necessary also to transfer some subgroups and items from one section or group to another to accomplish the above, but examination of the SITC indicates that changes of this kind would not be extensive.)

13. Development of the SCC in Canada and its implementation for international trade statistics have improved the quality of Canadian commodity trade statistics per se in addition to enhancing their usefulness through comparability with other series. The improvement has resulted from closer contact by those compiling international trade figures with all sources of commodity information resulting in greater awareness of the technical and market factors that affect particular commodities and influence the relationships between

commodities. These are dynamic factors requiring constant attention if commodity data are to be kept up to date and relevant to analyses of current market and other economic conditions. Examination of the SITC in connection with the current revision program indicates that it has not been kept up to date and that therefore it requires substantial amendment of this type in addition to the restructuring outlined above. The sections of the SITC most in need of internal revision are those in which new materials and techniques of production have been particularly dynamic, for example: Section 5, Chemicals; Section 6, Division 62, Rubber Manufactures n.e.s. (affected by the increasing use of plastics); Division 64 Paper, Paperboard and Manufactures Thereof; Groups 651-655, Textile Yarns and Fabrics; Section 7, Division 62, Electrical Machinery, Apparatus and Appliances (particularly with respect to electronic equipment).

14. In view of the significant developments that have occurred both in the volume and commodity content of world trade since the SITC was established, a revision such as that sketched above would provide a more useful instrument for trade analysis. Even more important, however, would be its influence in the development of comparable commodity statistics from other fields at the international level.

INTERNATIONAL TRADE STATISTICS AND THE S.I.T.C.

1. The Third Session of the United Nations Statistical Commission recommended that the revision of the League's Minimum List be prepared to make it more suitable for the analysis of international trade data and to make it more appropriate for the purpose of international comparability of trade statistics. In cooperation with the Governments and with the assistance of the Statistical Experts the United Nations Secretariat drew up the 1950 edition of the S.I.T.C. (Original). By 1960, governments of countries accounting for about 80 per cent of the world trade and the major international agencies had adopted the S.I.T.C. as the basis for trade statistics. A number of countries in Latin America and territories of the British Commonwealth used the original SITC as the basis of their customs nomenclature also. In many European countries and in a number of countries outside Europe, customs tariff nomenclature based on the 1955 Brussels Tariff Nomenclature (BTN) worked out by the Scandinavian countries was adopted as the basis of both the trade and tariff classifications. The re-grouping of BTN data into the form of the original SITC however presented considerable difficulties. To improve the situation the original SITC was revised maintaining its basic structure at the 3-digit level in the revised SITC.

2. Since the SITC (Revised) consists of the BTN items re-arranged in a statistical order, it can be used by governments in a number of ways. Countries whose tariff nomenclature is based on the BTN can compile economic statistics through the re-arrangement of customs data by means of a key between the two classifications or they can avoid the necessity of re-arrangement by requiring importers and exporters to enter in their declarations the commodity number according to the SITC (Revised) which automatically identifies the item of the BTN since the two sets of numbers are in reciprocal one-to-one correspondence; the latter procedure results in loss of customs information and, incidentally, makes it entirely unnecessary to require that both a custom and statistical number be entered in declarations.

India's Trade Classification

3. The Indian Trade Classification as revised on the basis of the Standard International Trade Classification (Original) was implemented with effect from 1.1.1957. It comprised nearly 4,850 articles with 9-digit codes retaining the frame work of the SITC (Original) classification upto the 5-digit item levels. The list was common for imports, exports and re-exports. Thereafter several proposals for changes in and addition to the list were received from the trade as well as from Government Departments. Only in special cases, however, the desired

amendments in the Classification could be carried out.

4. With the development, and diversification of the country's trade and industry, the list was revised again in 1965 keeping in view the requirements of various Government departments, Export Promotion Councils, Chambers of Commerce and other important trade and industrial organisations in the country. The revised Indian Trade Classification adopted the SITC(Revised) fully upto the 5-digit codes. The Revised Indian Trade Classification was drawn up after sub-dividing them into 7-digit headings to meet Indian requirements. Many unimportant articles of the old list were replaced by new ones and the total number of commodities in the Revised Indian Trade Classification, according to which the foreign trade of India is recorded, increased to 5,000.

5. In cases where no sub-divisions after 4 or 5 digit (sub-group and item) levels of the SITC(Revised) have been made, the description of the articles has been replaced with 7-digit code (by adding 3 or 2 zeroes, as necessary). In some cases the descriptions have been modified on the basis of the corresponding BTN headings to make them more explanatory.

6. In a few cases deviations from the SITC(Revised) had to be made to satisfy national needs. In those places, however, adequate explanations have been given so that there may not be any difficulty in correlating those headings with the corresponding SITC(Revised) items. For some articles requiring a more detailed break-down, new code numbers have been allotted for the concerned group, sub-group or item of the SITC(Revised). In all such cases blank spaces of the SITC(Revised) have been utilised. There are also instances where unsplit sub-groups(4-digits) of the SITC(Revised) have been sub-divided into new items(5-digits). There have been minor deviations at the division(2-digit) level under Section 0 where 054.2 and 055.41 have been shifted to new codes 049.1 and 049.2 respectively in order to bring together all grains, pulses and flour. The coverages under groups 653 to 657 are also slightly different.

7. As against 177 groups of the SITC(Revised) sub-divided into 625 sub-groups, there are 184 groups of the RITC, sub-divided into 650 sub-groups(4-digit level). Out of the 650 sub-groups, 296 have been sub-divided into 1088 items (5-digit level), 610 of which have been further sub-divided into 4168 commodities(7-digit level). Total number of commodities separately specified in the RITC is 5,000 (including 354 undivided sub-groups and 478 undivided items).

Practical Difficulties

8. As explained above, the SITC(Revised) signifies a common set of codes upto 5 digits facilitating to that extent comparison of trade statistics of one country with another. There are, however, instances where commodities (detailed upto 7 digits) of export importance for one country do not find exclusive place in the import statistics of major importing countries. Trade statistics of these items are traceable either clubbed with other similar items or grouped under a general heading upto 5 or 6 digits, rendering it difficult to make use of the trade data. A few examples are cited below for the purpose of illustration of this point.
9. Sandalwood Oil: In the case of Sandalwood oil(RITC 551.1031) fetching foreign exchange worth about Rs.32 million a year, there is no specific code in the trade accounts of a major importing country like France, wherein it is clubbed with "other oil perfumery" under code 330.153. This makes it difficult to identify the imports of Sandalwood oil into that country from India.
10. Guar Gum: Similarly, Guar Gum(Guar Flour) - RITC No.292.9122 - is an important export item from India contributing over Rs.10 million per annum, but there is no separate code for it in the trade accounts of major importing countries of interest to India like France, Italy, Japan, Federal Republic of Germany and Switzerland. This is usually clubbed with vegetable saps and extracts, pectic substances, mucilages and thickeners derived from vegetable products and Agar Agar.
11. Frog Legs: India exports Frog Legs(Frog meat) - RITC No.001.8901 - mainly to USA and France, current exports being of the order of Rs.6 million. However, in the trade accounts of France, the item finds place only in the general category, 'other meat' under code 020-431.
12. Mangoes: Mangoes (RITC No.051-9502) worth Rs.2 million are exported from India at present to UK, France, Federal Republic of Germany, Switzerland etc. However, in the trade accounts of UK, mangoes are grouped with other tropical fruits like guavas and avocados under the code 051-93. In the trade accounts of France, the item is found under code 080-142 clubbed with other fruits under the general heading "other Fresh Fruits". Similarly, in the trade accounts of Federal Republic of Germany and Switzerland, there is no separate code for Mangoes.
13. The above are only a few selected examples. It is also true that in most of the cases cited above, comparison has been sought to be made with countries following BFN in their

trade statistics. Nonetheless, these examples suffice to focus attention on the practical difficulties experienced by a country like India in the utilisation of the trade statistics of other countries. Two major problems need to be considered in this context. The first one relates to the possibility of evolving a common system of codes to be followed by all countries. This is necessarily a very complex problem. The second question, which is not only relatively simple but also perhaps more relevant to this conference, is whether it is possible to come to a common agreement on a more detailed system of codes than SITC(Revised), so as to ensure separate codes for a larger number of commodities which are of common interest to different countries. A uniform system of 6 or 7 digit codes as against the 5 digit codes in the SITC(Revised) may be considered in this context, with individual countries being free to add two more digits to accommodate appropriate sub-classifications. It is realised that there are a large number of nations outside the Commonwealth fold which are currently using SITC(Revised), and as such this problem needs to be tackled eventually in a forum with a much wider base. But this Conference could certainly provide a useful lead in this direction for consideration of the U.N. Statistical Commission. The evolution of a more detailed international trade classification will facilitate the formulation of not only trade and tariff policies of all countries in the world but will also assist in the conclusion and implementation of bilateral trade agreements and in the task of export promotion generally.

12 - NEW ZEALAND

PROPOSED REVISION OF THE STANDARD INTERNATIONAL
TRADE CLASSIFICATION, REVISED

1. This paper sets out certain proposals regarding the pending revision of the S.I.T.C., Revised.
2. Division 01 : Meat and Meat preparations - For New Zealand conditions (and probably to meet the needs of other meat exporting countries) the breakdown of division 01, meat and meat preparations, could be improved. Because of its importance to New Zealand, subdivisions to fine levels are needed. These include class of animal, whether fresh, chilled, frozen, etc.; whether bone-in or bone-out, packaged in cartons, or not packaged in cartons; quality rating; whether ox, heifer, cow, bull; and weight ranges.
3. The present S.I.T.C. does not make full use of the third digit. Only three numerals are used in this column to subdivide meat into "fresh, chilled or frozen", "dried, salted or smoked, whether or not in airtight containers", and "in airtight containers, n.e.s. and meat preparations, whether or not in airtight containers". This means that if separate statistics of "fresh", "chilled", and "frozen" (for example) are required a further digit has to be used to accomplish this breakdown. However, better use of the digits mentioned above would obviate this need. For instance, instead of the three broad headings as currently used, it would be possible to use the following:
 - 1 Fresh
 - 2 Chilled
 - 3 Frozen
 - 4 Dried
 - 5 Salted (including in brine)
 - 6 Smoked
 - 7 In airtight containers, with or without vegetables or other ingredients
 - 8 Sausages, meat extracts, and meat juices
 - 9 Other.
4. It would be preferable to make the subdivision by type of animal in the third digit and the type of processing as the fourth digit. Set out below is a suggested three and four digit breakdown so far as bovine animals are concerned. Similar breakdowns (4 digit) where appropriate for other animals are also envisaged:

01 Meat and Meat preparations

- 011 Meat of bovine animals
- 012 Meat of sheep
- 013 Meat of goats
- 014 Meat of swine
- 015 Poultry (including poultry offals)
- 016 Meat of horses, asses, mules and hinnies

- 017 Meat of game animals (deer, wapiti, etc.)
- 018 Edible offals of the animals falling in
headings 011,012,013,014,016 and 017
- 019 Meat and edible offals of other animals.

011 Meat of bovine animals

- 0111 Fresh
- 0112 Chilled
- 0113 Frozen
- 0114 Dried
- 0115 Salted (including in brine)
- 0116 Smoked
- 0117 In airtight containers, with or without
vegetables or other ingredients
- 0118 Sausages, meat extracts, and meat juices
of bovine animals
- 0119 Other.

5. This proposal omits the original 0118 (1); "Poultry liver, fresh, chilled, frozen, salted or in brine" has been omitted and included with poultry. On the face of it, there appears to be no logical reason why poultry liver should have been separately enumerated. If any country wished to record poultry liver separately, could not this be done as a fifth or sixth digit under the poultry heading? Because of varying importance to exporting countries it is desirable to separate goat meat from sheep meats. Also, meat of deer, which is a current export from this country and on which it is desirable to keep separate records, has been added.

6. Of importance is the fact that the changes proposed above can be carried out within the existing numerical structure of the S.I.T.C. Revised, since, at present, numbers between 0139 and 0199 are not in use. The advantage is that, for countries requiring a breakdown into fresh, chilled, frozen, etc., a digit is saved and this is valuable in making one extra digit available for the many other classifications mentioned earlier.

7. Group 262 : Wool and other animal hair - Wool is another commodity important to New Zealand and various other countries. It would be very desirable to separate wool from other animal hair. In this division (26) the third digit is only used up to No. 7, so that there is room to do this. It is suggested that the groups in this division could be as under:

- 261 Silk
- 262 Wool
- 263 Other animal hair
- 264 Cotton
- 265 Jute
- 266 Vegetable fibres, except cotton and jute
- 267 Synthetic and regenerated (artificial) fibres
- 268 Waste materials from textile fabrics
(including rags).

The following sub-groups, at present numbered as under, could be transferred to the new group "other animal hair":

- 262.3 Fine animal hair, other than wool, not carded or combed
- 262.5 Horse hair and other coarse hair, not carded or combed (and the two sub-numbers)
- 262.7 The second portion - "Other animal hair, carded or combed (excluding tops)".
- 262.9 The second portion - "Of other animal hair, n.e.s."

8. One of the advantages, so far as the proposed changes affecting wool are concerned, would be that under this new arrangement, cards sorted on group 262 and tabulated would give total value for wool and total weight for wool. At present, this information has to be obtained by more indirect methods.

9. Relationship with B.T.N. - No doubt, when these proposed changes are being examined, relationships and effects on the Brussels Tariff Nomenclature will be taken into account throughout. Mention is made of what is felt to be a drawback to the Brussels Tariff Nomenclature so far as it affects the situation in New Zealand. It may have application elsewhere. In New Zealand trade statistics are prepared on an S.I.T.C. Revised basis to 7 digits while the Customs Tariff is on a Brussels Tariff Nomenclature basis to 6 digits. The particular portion of the Brussels Tariff Nomenclature concerned is 73.15 "Alloy steel and high carbon steel in the forms mentioned in headings numbered 73.06 to 73.14". This number is split into 26 subdivisions by means of letters and each of the letters is keyed to a separate 5 digit number in the S.I.T.C. Revised. In view of the limit of 6 digits in the New Zealand Customs Tariff this 26 subdivision makes further subdivision in certain cases difficult, or impossible, to attain. It would not seem difficult to break down the existing contents of 73.15 into smaller units. For example, the existing breakdown relates to "High carbon steel", and "Alloy steel", alternately throughout. It would be possible to breakdown some of the types further than they are at present. A recasting of this chapter would present no difficulties insofar as the availability of numbers is concerned since the numbers 73.41 to 73.99 have not been used. Furthermore, the economical use of digits in the main skeletons with which many countries will be concerned, can be a real advantage in using these skeletons because it can avoid the use of an extra digit in the S.I.T.C. classification. One digit or two digits may not seem of much importance, but when it is considered that importers and exporters are required to enter these numbers and that they must all be checked, the additional work and the additional possibility of error are greatly increased for each extra digit added.

10. Amendments and comparability - Although a number of changes have been proposed in this paper, it is nevertheless

pertinent to state that there are considerable drawbacks to the making of changes in a classification. The main disadvantage is that it becomes more difficult to extract time series data from existing punched cards, etc. relating to several years. If such time series are required, and if changes have been made, it then becomes necessary to seek out and repunch cards so that the material for all the years covered is on an identical basis. This is time consuming and can lead to errors. A good example here is the current proposal by the Joint Working Party on Forest and Forest Product Statistics (FAO/ECE/STATS/91 of 31/12/69) where proposed amendments to the present S.I.T.C. Division 25 would result in trade statistics for 251.3, 251.4, 251.6 and 251.9 referring to different pulp and paper products before and after the proposed changes. Only the 251.1, 251.2 and 251.6 groupings would in fact preserve absolute comparability. A rearrangement of the sequences of the proposed revision items would give desirable comparability historically. The same situation exists with Group 631, and particularly Group 641, where a one-for-one relationship at the fourth digit S.I.T.C. might be more difficult to achieve before and after the proposed amendment. However, the principle involved here needs special consideration before promulgating final changes in the S.I.T.C. or any other classification.

11. S.I.T.C. and B.T.N. - A high proportion of countries are now running their Customs Tariff on a Brussels Tariff Nomenclature basis and preparing most of their statistics on an S.I.T.C. Revised basis. New Zealand, from 1962 to 1967, operated its Customs Tariff on an S.I.T.C. Revised basis and used the same classification for its statistics. From 1 July 1967, New Zealand fell into line with the majority of other countries and adopted a B.T.N. - based Customs Tariff while retaining an S.I.T.C. Revised statistical classification. In the earlier days of promoting the use of the S.I.T.C. Revised classification, encouragement was given to the adoption of this classification both for customs tariff purposes and for statistical purposes. However, the Customs Co-operation Council had, over the years, built up a comprehensive system for the Brussels Tariff Nomenclature, complete with explanatory notes and with a service to Customs Departments throughout the world on classification decisions. This high degree of sophistication which had been given to the Brussels Tariff Nomenclature encouraged governments, no doubt pressed by their Customs Departments, to adopt a Customs Tariff on a B.T.N. basis. As a customs tariff, the S.I.T.C. Revised classification was not complete. For this purpose it would require a completely developed set of explanatory notes and a decision system tailor-made to suit the classification and this was never done. In New Zealand, it was found difficult by the Customs Department, to operate an S.I.T.C. Revised-based Customs tariff, using an adapted set of explanatory notes originally derived from the Brussels Tariff Nomenclature.

12. The development of a common classification - with supporting explanatory notes - would require considerable study

12 - NEW ZEALAND

and effort on the part of the United Nations Statistical Office and the Customs Co-operation Council. It would also need the full support from both organisations before it could be expected to succeed. It is felt that this matter merits close re-investigation. The present double classification system, where used, results in considerable extra work on the electronic data processing side and, indeed, also on the clerical side. Since frequent changes to the Customs Tariff are necessary for duty purposes, there is the constant problem of keying the two classifications together. Apart from the increased work, this creates a greater possibility of error.

13. Classification of war materials - It might be desirable to make a re-examination as to war materials. In the current issue of the S.I.T.C. Revised, group 951, "Firearms of war and ammunition therefor", is split into six sub-headings (apparently to enable keying with B.T.N. to be accomplished). One of these subdivisions is "Armoured fighting vehicles". On the other hand, "Warships of all kinds" is included in section 7 under number 735.1. So far as military aircraft are concerned, these are included under the "Aircraft, heavier than air", No. 734.1. Consideration should be given to the exclusion of all these items from the trade statistics. If, after consideration, it was decided that they should be left in the classification, then consideration should be given to the grouping together of all these items under one group number in section 9. Furthermore, it would be desirable not to split this group into sub-headings in the same manner as is presently done. The allocation of a multitude of small parts for defence equipment, even into a small number of items, is proving time consuming and difficult for the Ministry of Defence in New Zealand. It is felt that most countries would not want to disclose too much detailed information regarding its defence equipment, but, in any case, with one single number in the S.I.T.C. Revised skeleton, it could still make such subdivisions as it wished. Goods not intrinsically of a military nature but intended for military use could continue to be included in the trade record.

14. Inclusions and Exclusions - Referring to United Nations statistical paper M.34, page vii, the right hand column, it might be desirable to rewrite this definition, which is intended to cover inclusions and exclusions for import and export statistics. The question of inclusions and exclusions, insofar as they concern ships and aircraft, were the subject of U.N. Paper ST/STAT/27 of 15 April 1968. There are the following factors to be considered concerning ships and aircraft (the factors have been related to New Zealand but would apply, of course, to any country).

1. They can be new or used.
2. They can be used entirely for internal purposes in and around New Zealand, or between New Zealand and overseas countries, or even both.
3. They can be registered in New Zealand or in a foreign country.

4. They can be owned in New Zealand or in a foreign country.

It seems to be necessary to bring all these contingencies into the definitions relating to inclusions and exclusions. Possible sets of inclusions and exclusions relating to ships and aircraft, taking the above factors into account, (and again, related to New Zealand) are set out below:

Import Inclusions:

New and used ships and aircraft arriving in New Zealand, irrespective of nationality of ownership or registry, if they are to be used only for internal trade between New Zealand ports or airports.

New and used ships and aircraft arriving for the first time in New Zealand if they are to be registered in New Zealand and are to be used between New Zealand and overseas countries.

New and used ships and aircraft arriving for the first time in New Zealand which have been purchased by a New Zealand company or New Zealand citizen to be used between New Zealand and overseas countries, irrespective of nationality of registry.

Import Exclusions:

Ships and aircraft not on the New Zealand registry or owned by a New Zealand company or New Zealand citizen which are being used in transporting passengers or freight between New Zealand and foreign countries.

Ships or aircraft purchased by a New Zealand company or New Zealand citizen but full use not involving the physical presence of ship or aircraft in New Zealand territory.

Export Inclusions:

New and used ships and aircraft leaving New Zealand irrespective of nationality of ownership or registry if they have been used internally between New Zealand ports or airports (except as merely incidental to overseas operations).

New and used ships and aircraft permanently leaving New Zealand if they were registered in New Zealand and were in use between New Zealand and overseas countries.

New and used ships and aircraft permanently leaving New Zealand, sold or transferred out of New Zealand after ownership by a New Zealand company or New Zealand citizen and used between New Zealand and overseas countries.

Export Exclusions:

Ships and aircraft not on the New Zealand registry or owned by a New Zealand company or New Zealand citizen which are being used in transporting passengers and freight between New Zealand and overseas countries.

Ships or aircraft sold by a New Zealand company or a New Zealand citizen if use of such ships or aircraft had not involved their physical presence in New Zealand territory.

Since the capital cost of ships and aircraft is so substantial, it is considered that the spelling out of inclusions and exclusions for these items in some such detail as the above, is necessary.

15. It is to be hoped that discussion on the above comments could be of use in undertaking the revision of the Standard International Trade Classification.

International Trade Statistics and the S.I.T.C.

1. Statistics on trade are processed by the Customs Department of the East African Community in Mombasa, Kenya. These consist of monthly trade and revenue reports on domestic exports, re-exports, net imports, exports to Kenya and Uganda and imports from these two neighbouring countries. Quantity and value figures are given for main commodities, lumping the remaining items under "other goods". The term "domestic exports" refers to exports of domestically produced goods to countries outside East Africa. Re-exports means goods which originally came in from outside East Africa and exported again to countries outside East Africa without reprocessing; and net imports refer to goods imported from outside East Africa. Apart from the monthly trade and revenue reports there are monthly hollerith reports giving detailed S.I.T.C tabulations by items. In 1968 the Community changed from five-digit S.I.T.C codes to six digits.
2. Exports and imports to and from Kenya and Uganda refer to trade in locally produced merchandise. Hence, to arrive at a complete picture of international transactions one has to aggregate external and inter-East African trade figures.

INTERNATIONAL TRADE STATISTICS AND THE UNITED NATIONS
STANDARD INTERNATIONAL TRADE CLASSIFICATION

INTRODUCTION

The current edition of the UN publication on the SITC (Revised) (1), issued in 1961, includes a short description of the historical background to the Classification as it stands today. In 1961 governments of countries accounting for about 80 per cent of world trade were compiling data on trade by commodity according to the SITC; today such countries account for approaching 90 per cent of world trade. The United Nations Year Book of International Trade Statistics for 1967 covered 143 countries. Data on an SITC basis is shown for 127 of them. Forty of the 143 countries listed are members of the Commonwealth; data for all of these was in terms of the SITC.

2. The SITC(R) has remained in its present shape since 1959. The UN Statistical Commission in 1957 had recommended that a better relationship be sought between the SITC and the Brussels Tariff Nomenclature (the BTN) (2) which is used for trade statistics by many countries throughout the world. In 1959 a group of experts from several inter-governmental agencies and countries achieved modifications resulting in a one to one correspondence between the BTN and SITC items.

3. Further modifications to the SITC(R) have subsequently been proposed, for example to achieve better comparison of trade statistics with production statistics. One major set of proposals for changes in the headings for paper, paperboard and manufactures thereof (Division 64) was made by the OECD Special Committee for Pulp and Paper and considered by the UN Statistical Commission at its 15th session in February, 1968 (see UN paper E/CN3/384, 16 November 1967). There was widespread opposition to the proposals, for several reasons which have been set out in the UN paper itself. But perhaps the most significant in relation to the subject of this paper was the belief that changes to the SITC should not be considered piecemeal but only when a general review was warranted. The Commission concluded in 1968 that a revision of the SITC(R), in consultation with the Customs Co-operation Council (CCC), was of high priority.

(1) Statistical Papers, Series N7 No 34.
Standard International Trade Classification, Revised, United Nations, New York, 1961.

(2) Nomenclature for the Classification of Goods in Customs Tariffs, Customs Co-operation Council, Brussels, 1955; and Explanatory Notes to the Brussels Nomenclature, Brussels, 1955.

4. Work on such a general revision is now in hand. The Secretariat of the UN Statistical Office were to make a brief progress report to the 16th session of the Commission (October 1970). A group of experts from seven countries and 6 international organisations was convened under the Chairmanship of the Statistical Office in New York last November. The main efforts of the New York meeting were concerned with many problems of detail and with the identification of particular problems which would need to be studied further in depth by the Secretariat, before submitting proposals to member governments of the United Nations. Certain questions affecting the structure and principles of the SITC(R) were also aired, though inconclusively. It is the intention of the Secretariat to circulate proposals to member countries for comment probably in the spring of next year, and to re-convene its group of experts late in 1971 to draft final recommendations for the Statistical Commission to consider in 1972.

5. Between them, the Commonwealth countries share the experience of wide use of this major classification from which to draw lessons. It would be useful for our countries, therefore, to consult together in advance of the approach which will be made by the UN Secretariat in 1971 to all member countries, seeking their views on the proposals as they will then stand.

6. This paper therefore sets out to review some of the problems of the Classification and its revision.

THE PRINCIPLES UNDERLYING THE SITC

7. Before setting down our views on the problems of revising the SITC it may be useful to consider briefly what the Classification sets out to do in terms of the principles underlying its design.

8. It can be fairly said that no entirely satisfactory trade classification has been devised on the basis of a single principle, for example, the nature of the material used, the end-use of the commodities, or the industry of origin. It is arguable that a more satisfactory approach, either for the trade classification of an individual country or for an international standard, may emerge from a compromise which chooses for the different groups of commodities the basis of classification most likely to suit the interests of the majority of users of the statistics.

9. Such a compromise is the basis of the SITC. It has resulted in a hybrid scheme which cannot perhaps be precisely described (except in terms of its full coding system), and which does not always give a unique answer to a question asking where a particular commodity should go. It has to be recognised that it does not serve all purposes; that sometimes the experts who designed it had to accept a compromise on a particular heading where the argument could not be resolved to the satisfaction of all. But the Classification seems to have provided, for most users - in governments and industry - a broadly acceptable system to which countries can work and which has proved its utility.

10. Some of the elements underlying the Classification as it was first designed may be noted. Inevitably, it was affected to some extent by the classifications already in use in some of the major trading countries. Among the main features of the scheme was an attempt to reflect the normal way in which commodities were traded; some examples of this occur in Section 2, Crude Materials, which includes some manufactured materials which are handled like crude materials, e.g. synthetic rubber is in the same category

as natural crude rubber.

11. Another factor underlying parts of the main structure is the stage of production itself, separating crude from manufactured goods. But in one major sector, food, stage of production is virtually ignored - raw and processed goods occur throughout Section 0, (a feature of the system which for some advanced economies may be unsatisfactory). A similar mixture occurs in Section 4, Oils and Fats, where the products may be very simply prepared or produced in complex manufacturing plant; the separation of this Section from Section 2 itself may reflect the historical as well as economic circumstance (e.g. the great problem of shortages of oils and fats after the war).

12. In contrast, Section 3 is essentially confined to a collection of goods having common functions.

APPROACHING THE REVISION OF THE SITC

13. A major feature of the revision would consist of up-dating headings to allow for the growth of new products in trade or for changes in importance of products with some history in international trade, but other changes are not excluded.

14. The paper presented by the UN Statistical Office to the group of experts last November qualified the requirements of up-dating with a strong plea for continuity. It may be helpful to summarise the UN Statistical Office's paper here, recognising that some of its arguments will appeal particularly to those member countries which have devoted resources to introducing the use of the Classification in their compilations and which desire to maintain their series at lowest cost.

15. In addition to stressing the need for continuity, the paper made the following points:

- i. The 5-digit structure should be maintained as far as possible; the use of 6 digits should be avoided if possible.
- ii. The existing 3-digit (group) headings should be retained as far as possible, up-dating being done mainly at 4-digit (sub-group) or 5-digit (item) levels.
- iii. Economy should be exercised in increasing the number of items.
- iv. The correlation between SITC and BTN should be maintained, possibly modified so that, instead of full one for one correlation, one SITC heading might sometimes match 2 or more BTN headings to avoid retaining unnecessarily detailed SITC items.

16. Apart from the special exercise in 1959, the occasion of the present review is the first opportunity, after 15 to 20 years' experience of the SITC, to consider revisions. The Statistical Commission envisaged the possibility of future up-dating exercises every 10 years, and it is important to plan ahead as far as possible in order to make future changes less disruptive.

17. But it would be both difficult and costly to establish a radically different classification, even if a better basis could be found. It seems

necessary to recognize the risk of so altering the SITC(R) that we would end with 2 systems in use because some countries would not find the resources for a markedly changed system. The object should be therefore to improve and not undermine the Classification.

ASSOCIATED CLASSIFICATIONS

18. Before examining more closely the direction of possible and worthwhile revisions to the SITC, it would be useful to consider its adaptation for other types of classification. While the Classification does serve at present to meet several requirements in general economic analysis, some systems of classification of commodities entering international trade have been called for to meet certain more specific requirements. For example, the need has been felt to associate trade statistics with aspects of national accounts and industrial statistics. It is already possible, within certain limitations, to use the Classification to meet such needs; 2 examples are described here.

BROAD ECONOMIC CATEGORIES (BEC)

19. Earlier this year, the UN Secretariat issued a revised Draft Classification by Broad Economic Categories (1) for discussion at the October 1970 Statistical Commission. This is not directly linked with the revision of the SITC(R) but there are some aspects of this topic which may be usefully discussed here.

20. Despite the difficulties that stand in its way, there is a desire in many countries for a trade classification of commodities according to economic classes, to facilitate economic analysis of the uses to which goods are put. There is a long history of discussion among statisticians, struggling towards a solution of this problem; the Statistical Commission's session in October may have advanced matters by the time of the Conference of Commonwealth Statisticians. But against this background it seems desirable to examine whether, without unduly damaging the structure of the SITC, changes can be made in it to help towards a classification by Broad Economic Categories. One or two aspects of this are touched on later.

21. It is not possible in this paper to anticipate the outcome of discussions on the BEC at the Statistical Commission. However, after a number of attempts, the latest Draft is likely to have advanced the general design of a BEC about as far as it is possible to go. Any weaknesses follow from the limitations of the SITC which, as it stands, is not entirely suitable for analysis by end-use, and which, in our view, cannot be re-designed to make it fully suitable. It is in the nature of many commodities that at the time of import or export no definite knowledge of end-use can be obtained. This is not to say that it is impossible to amend any SITC items so that they can be allocated reasonably by end-use or main end-use. But there will always be some headings, including several major ones; which cannot be so allocated on the basis of the documentation available for compiling trade statistics; and it would be undesirable to distinguish fresh headings which depart from the standards of precision generally obtaining throughout the SITC.

(1) UN Statistical Commission. External trade statistics: draft classification by Broad Economic Categories (E/CN.3/408) 1970.

22. A second Associated classification making use of the SITC appeared in 1966, when the United Nations published a study relating the items of the SITC to the groups of the International Standard Industrial Classification (ISIC) (1); this was intended simply as a guide indicating "for the wide range of commodities entering world trade, the industries from which they normally flow". Here again, the need for this kind of correspondence needs to be borne in mind in the revision of the SITC.

DIRECTION OF POSSIBLE REVISIONS TO THE SITC

23. The rest of this paper discusses the possible directions of revisions to the SITC, and certain problems arising from them, under 3 main headings:

- a. changes of detail;
- b. transposition of headings across Divisions or Sections;
- c. overall structure of the Classification.

a. Revisions of detail

24. It is not intended here to discuss the various specific suggestions for revision of detail, but rather to review the kind of changes which it would be appropriate to introduce at this stage of the development of the SITC.

25. Changes of detail could be made for the following reasons:

- i. the separate identification at item level of important commodities in international trade, especially considering new and growing products;
- ii. the abandonment of headings which have become unimportant;
- iii. the abandonment of headings for which goods cannot satisfactorily be identified;
- iv. sub-division of headings to assist in matching SITC headings with headings in e.g. the BEC or the ISIC.

NEW HEADINGS

26. There are many examples of new and growing commodities. What is surprising, perhaps, is the absence of separate SITC items for some old-established products of major importance in world trade.

27. But there may be a problem of fitting new items into the Classification. There is first the question of the correlation with BTN. It is understood that the Customs Co-operation Council at Brussels would be willing to consider further sub-division of their 4-digit headings by the use of letters, to maintain a 1 for 1 correlation; as was done at the time the SITC(R) was introduced in 1959/60, with the proviso that any new headings can be satisfactorily defined.

(1) Statistical Papers, Series M No 43, Classification of Commodities by Industrial Origin, United Nations, New York, 1966.

28. In some areas of the SITC, however, in particular perhaps Section 6, there may be little or no room in the numbering to introduce fresh 5-digit headings. One possible way to meet this difficulty would be the introduction of optional 6-digit headings; but they would have to be optional as the sixth digit might cut across existing sub-divisions in some countries and this would weaken their value. The proposal would be a significant departure from the current shape of the SITC and on balance may introduce more problems than it would solve. Another alternative to deal with crowded parts of the Classification is considered below.

ABANDONMENT OF PARTICULAR SITC ITEMS

29. At the New York meeting of experts, the UN Statistical Office and some of those attending drew attention to the existence of some items which are not statistically significant in international trade. There was some general sympathy for the view that SITC items should exceed some pre-determined minimum percentage of world trade. No firm figure was identified, especially as it was pointed out that for some countries minor items in international trade could be very important to them individually. But several very minor items were identified and it seems sensible to simplify by omitting them where there is general agreement. If separate headings were abandoned for this or any other reason, the 1 for 1 correlation with BTN would disappear. However, the modification of this correlation suggested by the UN Statistical Office so as to allow a single SITC heading to match 1 or more BTN headings would still allow the BTN to be used for purposes of definition, and trade data based on BTN headings could still be readily combined into SITC headings.

30. A few instances may remain where goods constituting an item in the Classification cannot be satisfactorily distinguished in practice. An example of this is industrial diamonds, which cannot be satisfactorily distinguished from other diamonds.

SUB-DIVISION OF ITEMS

31. Finally, there may be a case for sub-dividing headings in order to ease the adaptation of SITC headings and data to other classifications. An example of this kind is sub-group 025.0, eggs, for which in any event the SITC already suggest an optional sub-division into eggs in the shell and not in the shell. There is no identification problem, and the sub-division would allow the 2 classes of product to be allocated separately by end-use between industrial use and household consumption for use in a classification by Broad Economic Categories (although this is not to say that the split between "shell" and "other" precisely matches the split by end use). Whether such a division should be introduced would depend in part on how important the trade was in the different sub-headings.

b. Transposition of headings

32. Many proposals for the transposition of headings between Divisions or Sections were submitted to the UN Statistical Office for consideration by last November's meeting of experts, but little progress was made on this. While it is likely that there would be in practice widespread acceptance of several of the proposals, the group of experts found themselves involved in a discussion of principles underlying the definitions of particular Sections of the SITC. It may be helpful, therefore, if this gathering of statisticians exchanged views on this topic in order to give a lead.

DISTINCTIONS BETWEEN SECTIONS OF THE SITC

33. The problem can be simply exemplified by reference to one of very few proposals for transposition considered by the New York meeting, namely, that wool tops (sub-group 262.8) should be transferred to Division 65. While there appears to have been some willingness to accept this proposition, the discussion called into question the distinction between Section 2, crude materials, and Sections 5 to 8, manufactured goods. Section 2 includes many commodities such as wool tops, in the production of which substantial and skilful processes have been employed which can only be described as manufacturing. Several other headings of synthetic goods (e.g. rubber, man-made fibres) are found in Section 2 although they are clearly manufactures

34. The group of experts put forward several views. One was that Section 2 covered or should cover simple processes, with little added value, while the goods in Section 6 were more complex, with greater added value. (As a basis of distinction for an international standard there would be difficulties flowing from the variability between countries in the value added in producing like goods.) Another view queried whether it was right to separate like things, e.g. synthetic from other fibres. It was suggested that Section 2 covered the "inputs" into Sections 5 to 8. There were arguments of principle against convenience. There were suggestions that distinctions between more and less manufactured goods should be sharpened.

35. In approaching questions of transposition and the distinctions between different Sections of the SITC, account should be taken of several factors. The SITC is a comparatively simple economic tool; its users recognise that it is impossible to design a perfectly logical system of classification that has no problems of defining the border-lines between particular classes of goods. Against this it does yield some broad classes of general economic interest.

36. Trade and production practices and relationships change, and it cannot be expected that parts of the Classification will not become out of date; nor could it be claimed that the immense task performed at the outset of drafting the Classification was entirely without inconsistencies. But one can perhaps see the intention of the drafters in, for example, the inclusion under Section 2, Crude Materials, of certain manufactured materials which are usually handled like crude materials, such as synthetic rubber along with crude natural rubber or synthetic fibres along with crude fibres. This kind of approach seems a useful one, even if there is occasion for some shifts at the margin from time to time.

37. An alternative approach to the distinction between "Crude Materials" and "Manufactured Goods" would be to widen "Crude Materials" to say, "Basic Materials", which could also include what are sometimes called "semi-manufactures" such as non-ferrous metals, steel, paper. But there would be fresh problems of identifying new border-lines. And, more important, a distinction of this kind may be difficult to accept throughout all member-countries, with their widely varying economies. What would be regarded as the manufactured end-products of some countries would be the material inputs of others.

38. Further, such a departure from the present structure of the SITC would be so at variance with its development and use that the whole continuity of this widely accepted basis for a system of international trade commodity data would be lost. As it stands the SITC is recognised and used familiarly in a considerable range of circles both internationally and in different

countries. It may be preferable to reserve groupings such as "basic materials" or "materials for industry" to a system such as the proposed BEC.

39. Clearly there will be room for considerable argument on the transposition of headings across the Classification no matter what its structure. But within its present broad form, it may be possible without damaging its simple underlying pattern or principles, to effect transpositions which if desired could take account of where items fall under, say, the BEC or the ISIC.

c. Overall structure of the Classification

40. The sectional structure of the Classification itself poses a number of problems.

41. For example, there is Section 4, Oils and Fats, which accounts for less than 1 per cent of world trade. There was, when the SITC was first drafted, a serious world shortage of oils and fats and substantial interest existed in this trade. While the products are still of major interest to many countries, it does seem reasonable to doubt whether they should still constitute a separate Section.

42. Again, over and above the question of distinguishing between "non-manufactured" and "manufactured" goods, there are difficulties of distinctions within the manufactured goods Sections themselves. The separate Sections for Chemicals and for Machinery and Transport Equipment are pretty clear and distinct. The division between Sections 6 and 8 is less clear cut. The distinction in Section 6 of goods classified according to material and in Section 8 of goods effectively according to function is perhaps too rough and ready and many users have doubts on its value. One answer to this might be to move some end-products (e.g. cutlery, carpets, pottery/china, household wares) to Section 8 (where they may be as at home as lighting fittings, furniture, clothing, etc.). Another solution, which would avoid some definitional problems, might be to run Sections 6 and 8 together.

RE-NUMBERING OF PRESENT DIVISIONS

43. There is a more practical problem facing those working on revising the SITC than any logical separation of the present Sections. Some users believe that there is a need to relieve congestion in parts of the SITC, in particular Section 6, by the creation of additional Divisions. This would serve 2 important purposes. First, it would provide room to allow the introduction of new headings at item level more conveniently. Secondly, it would enable the summarisation of significant collections of headings where there is no room at present for further Divisions, Groups or sub-groups. But the creation of new Divisions is prevented by the "decimal structure" of the Classification which restricts the full use of Division (2-digit) cells, since they are consolidated into 10 Sections (1-digit), each Section having space for only 10 Divisions. At present only 56 Division cells are in use (with no Divisions under Section 9). But the unused Divisional cells cannot be used to relieve congestion in those Sections where the variety of commodities covered could be most satisfactorily catered for by creating more Divisions.

44. An example of crowding occurs at a lower level in Division 65 (textiles), though not all possible Groups are in use. This one Division out

of 56 has nearly a hundred items (or sub-groups with no item breakdown) out of the total 1,312 in the Classification. The commodities in this one Division out of 56 coincide approximately with 15 of the 99 BTN chapters. However, it is not so much the question of apparent balance in distribution of headings which is of concern, as how to allow for any developments within the Division. There are admittedly 2 unused groups in the Division, but these might not meet the requirements arising from the different kind of suggestions being put forward in this area. One major proposal for textiles, from the United Kingdom, suggests that they be distinguished under 3 headings:

- i. pre-weaving processes
- ii. fabrics other than special fabrics
- iii. special fabrics and made-up articles.

While groups falling under such broader headings could be summarised independently of the main Division or Group structure, it would be convenient to create 3 divisions out of the one.

45. The United Kingdom have suggested consideration, therefore, of the abandonment of numbered Sections (there could be identifying letters if required); and that the present Divisions be re-numbered - still with 2 digits - and additional Divisions created where required, still leaving spare cells for future needs. The number of Sections could be reduced to, say:

- A. Food, beverages and tobacco (present Sections 0, 1)
- B. Crude materials; oils and fats (2, 4)
- C. Fuels (3)
- D. Chemicals (5)
- E. Machinery and transport equipment (7)
- F. Other manufactured goods (6, 8)
- G. Commodities and transactions not classified according to kind (9)

46. While the convenience, for mechanical summation, of limiting a Section to 10 Divisions, and having a unique first digit for all Divisions in a Section, would be lost in this proposal, the re-compilation of Divisions from past records to agree with the proposal set out above (or vice-versa) should not be costly.

47. Alternatively, extra Divisions could be created, by re-numbering and using the "vacant" 40 or so 2-digit numbers, while still retaining 10 Sections with their existing broad contents. While still losing the convenient "decimal" structure, this would allow continuity more readily at Section level.

CONCLUSION

48. It may be helpful in conclusion to bring together some of the main problems involved in the revision of the SITC. Although the points below are set out under 3 broad headings, in many respects their significance crosses between the headings.

(i) Addition or deletion of headings

49. It may be unwise to seek many net additions to the Classification. In order to make room for some new or growing commodities, should there be some deletion of trivial separate headings, by amalgamation? Should there be some minimum percentage of world trade in total below which there should be no SITC items?

50. How best could new items be fitted into any part of the structure of the Classification where there is overcrowding? This is something for detailed study by the UN's group of experts after collation and examination of all detailed proposals. But should the group be bound as far as possible (or completely) by the existing structure or would some form of relaxation be acceptable?

51. Would the combination of existing minor or difficult headings be acceptable in such a way as to modify the existing one for one correlation between SITC and BTN headings so that the correspondence was met by one SITC heading matching precisely with two or more BTN headings?

(ii) Transposition of headings

52. What guidance may be given to the group of experts in considering questions of transposition of headings from one part to another?

53. Do the Sections need more logical boundaries? Or can their long-standing looseness and hybrid nature continue to be accepted?

54. Are there any further criteria (other than the contents of Sections) which could serve to indicate what transposition of headings might be made? Should elements of the Draft Board Economic Categories be taken into account? Should industry of origin of commodities be equally or more significant than, say, commercial selling practices, or the function of a commodity?

(iii) Structure of the Classification

55. Arising from the considerations above of changes in headings and transposition of headings, suggestions for changes in the structure of the Classification could flow. In the United Kingdom view, such changes should preferably not involve major reconstructions or departures from the existing classification, but it would be valuable to exchange opinions on this. To ease some of the problems of new and transferred headings, some opening out of the Divisional structure away from the present system of a maximum of 10 Divisions in each of the Sections is desirable in our view. Our preference would be to have at the same time some adjustment of Sections to a pattern like that set out in paragraph 45; continuity with the past could be maintained in practice though there would be some changes in appearance of the Sections. Alternatively, but possibly less meaningfully and with less balance, the existing then Sections could be retained but with Divisions taking up available vacant 2-digit cells by spreading beyond the limits of the present 10 in each Section.

NEED FOR INDEPENDENT CROSS-CHECKS OF STATISTICS FROM DIFFERENT SOURCES,
THE EVALUATION OF THE ACCURACY OF SURVEY RESULTS

Whether statistics are obtained from censuses or surveys, evaluation of their degree of accuracy is important for two reasons.

2. Firstly, users should be made aware of any significant limitations of the statistics they use. As increasingly complex types of statistical inquiry are attempted, and as a greater understanding of the nature and magnitude of the limitations of traditional censuses and surveys is obtained, it is apparent that a number of unavoidable inaccuracies may have been occurring in some statistical collections. In these cases it is the responsibility of the statistical office to caution users. This may be done in general terms. General cautions, are, however, apt to be either ignored or taken too seriously, since users cannot assess their relevance for their own purposes. Hence, where there is detailed knowledge of important limitations on accuracy, such knowledge should be made available.

3. Secondly, evaluation of statistical accuracy is important because it throws light on the effects of particular sources of inaccuracy. For example, if it is known how much error is due to sampling variability and how much to response, attention can be directed to remedying that source of error most seriously affecting results.

4. The evaluation of sample errors of statistics based on samples is now virtually a matter of routine. Where full sample error calculations are not practicable, the use of replicated samples would afford some check on sample errors. Where many estimates are based on a sample, it may be neither reasonable nor necessary to give sample errors for all. Unless the errors are known to be quite small, however, they should be obtained and made available, at least in summary form.

5. Quality control and acceptance sampling are being increasingly used in processing statistical data, both to save labour in checking operations and to give better control information. One effect of this is that information on the quality of processing operations is becoming available more readily and for more operations. Similar results have been obtained from studies directed specifically to assessing the quality of certain processing operations. As a result, even where the accuracy of processing operations is not checked, it is possible to have some idea of the degree of error in them. Experience so far in the Bureau of Census and Statistics (principally in respect of Population Census processing), while not indicating errors above a level which could reasonably be tolerated, has tended to support other studies which have shown an incidence of errors in processing somewhat higher than had been previously supposed.

6. To evaluate the data obtained in statistical collections, whether for completeness of coverage or accuracy of individual responses, is in most cases an expensive undertaking. It has not in general been practicable except by using relatively small samples in evaluation; this fact has reduced the assurance attaching to conclusions from such evaluations, as have the difficulties of obtaining any completely certain check on the accuracy of

individual responses. Nonetheless it is apparent (for example, from comparison of data obtained by mail and field collection for the same units in Australian tests of an Agricultural Expenditure Survey) that the quality of response can cause results to be affected by variability or biases which may often be at least as serious as those arising from sampling. Similar situations may arise in respect of coverage. Whether remedial action should be directed to reducing sampling error or improving quality of response or coverage then depends on the relative costs of the two procedures, as well as on the extent of the resultant improvements in accuracy.

7. One further deficiency in statistics may arise, often unavoidably, from the use of concepts, definitions, classification bases, or the like which are not well adapted to the various uses to which the statistics are put. Such deficiencies are not regarded as errors since, if they are the only limitations, the statistics represent accurately what they purport to represent. From the viewpoint of the user, however, statistics of, say, imports whose documents are processed within a certain month may be just as unsatisfactory as a measure of imports made in that month as would be statistics of imports actually made during the month which were subject to various response, sampling and processing errors.

8. Measurement of deficiencies of concept is particularly difficult, since alternative uses may suggest alternative conceptual bases, and there may be no way (or no economic way) of applying precisely the concepts appropriate to particular uses. Where statistics are obtained as a by-product of administrative procedures, the statistician has often very little choice but to accept the concepts implicit in these procedures. These concepts may be appropriate for the information needs of the administrative agency, or they may merely be directed to administrative controls; in either case it is generally much cheaper to accept them substantially as they are than to undertake an independent collection.

9. It is in these last areas of response and coverage bias and conceptual deficiency that cross-checks of independently derived statistics (as for example, between wool reported produced by growers and wool reported sold by brokers) can be of particular value. These are the areas in which internally based assessments of deficiencies are impracticable and additional studies most difficult or expensive. In most cases, the cross-checks themselves will only be practicable at the level of broad aggregates. Such checks are better adapted to revealing systematic effects (of the nature of a bias, leading to consistent over- or under-statement). Response and coverage biases and conceptual deficiencies are likely to give rise to such differences; sampling errors and, to an extent, processing errors, are likely to be less evident in high-level aggregates, since their effects on components will often be in opposite directions.

10. Independent sources can have the disadvantage of creating difficulty in explaining the existence and divergence of two sets of statistics. One way of avoiding this difficulty is to arrange that only one set is published. However, if two sets of data are available without duplication of effort, and have distinct uses, it seems hardly desirable to suppress one on the grounds of divergence alone. The most common cases are those of somewhat different concepts (possibly arising from different auspices) in fact serving different needs, or of a deficient earlier result followed by a sound but later one. Whether such conflict arises within a

statistical office or between separate agencies, the most useful course where distinct uses exist will usually be to release both with a careful (and agreed) statement of the grounds for differences. (For this purpose some cross-checking or evaluation studies will probably be necessary.) Such confusion as may result will be largely among those who do not make frequent use of either set of statistics. Despite these considerations it would be preferable where economically possible for distinct needs to be met by commonly based collections, thus avoiding differences in the statistics apart from those meant to serve the distinct needs.

11. The value of cross-checks at the aggregate level in assessing accuracy will often be limited by the number of possible sources of differences between the two sets of statistics. Where differences occur, they may be hard to explain not because there is no reason for them but because there are too many reasons. Measurement of individual causes of difference may well involve the same kind of detailed investigations that would be necessary to assess accuracy if only one set of statistics existed.

12. Nevertheless, cross-checks are of considerable value, whether in indicating areas of statistics which are particularly worthy of closer investigation or in pointing to some factors as likely major contributors to observed discrepancies. Where disaggregation is enough to show, for example, that the discrepancy emanates from a particular sector of activity or type of respondent it would then be possible to go straight to the area most in need of attention.

13. In particular, the statistician should be careful of condemning other series (perhaps less technically sound) which he finds preferred to his own. Users may not always be aware of the risks attaching to deficient statistics; nevertheless those who are experienced in the use of statistics may be supposed to have a reasonably clear idea of their own needs, and if they prefer an apparently inferior product it is likely to have some virtues: timeliness may be one, and definitions and classifications more appropriate to particular uses may be others. It may be necessary, however, where it is believed that confusion may arise on the part of less experienced users, for the statistician to draw attention to the limitations of other series in the same way as he would do for his own.

14. The statistics for independent cross-checks considered so far are those largely outside the control of the statistical office, and comparable only at high levels. Where there is capacity to compare small units (e.g. households for one source with the corresponding sets of persons from another), the situation becomes similar to that of a standard post-enumeration survey. However the two sets of records arise, detailed discrepancies can be noted and a careful assessment can be made of the causes of the discrepancies. While such an operation may be cheaper than creating a second source, as in a post-enumeration survey, it is not always so, and it is certainly more expensive than checks using aggregates.

15. One further point may be noted in regard to detailed work on evaluation of survey accuracy. There are obvious advantages in knowing the actual limitations of a survey; consequently evaluation is often regarded as appropriately following the survey. From some points of view, however, it is better to regard it as a preliminary to the survey. Careful evaluation of results from pilot tests used in survey development is relatively cheap because of the small scale of the tests, and the fact that skilled staff (suitable for checking operations) are often being used for operations in the testing phase; it improves the survey results rather than finding deficiencies too late to remedy them; and it gives leads on areas of the main operation which may be worthy of further evaluation. Independent cross-checks with aggregates are, however, rarely possible at this stage; nor is any need for evaluation of actual performance met by prior evaluation studies.

16. In summary, the evaluation of components of error or deficiency in survey results is an important function of the statistical office, but is in some regards a costly one which must therefore be limited. Information suggesting particular deficiencies for investigation is thus of special advantage to indicate the best use of the resources available for evaluation. Such information can be had from the special creation of independent sources, as in post-enumeration surveys not directed specifically to obtaining particular information on discrepancies. This approach, however, is also relatively costly and may be more effective if used principally as part of pilot testing. The use of cross-checks with existing independent sources, while presenting some difficulties in interpretation, can often provide a cheap indication of existing statistics most in need of improvement and of the more serious deficiencies in those statistics.

THE EVALUATION OF THE ACCURACY OF SURVEY RESULTSI. Introduction

1. Responsible statisticians have probably been carrying out some kind of "verification" or assessment of the reliability of survey data since the earliest days of survey-taking. The first formal attempts to evaluate the accuracy of survey results most likely concerned the estimation of sampling errors in sample surveys. A more comprehensive understanding of the different sources of error that surveys are subject to was gradually developed during the last thirty years. During this period models have been set up for several of the sources of error (including some so-called "mixed error" models designed to probe not only the dynamic behaviour of individual sources of error as functions of the parameters of the survey design but also the joint effect and interrelation of several of them) and techniques have been proposed and implemented for the estimation of the magnitude of many of these error components. These attempts, in turn, led to hypotheses concerning the reduction of some of the errors under alternative methods of survey-taking. Thus the evaluation of sources of errors in surveys played a major role in the development of present day survey methodology.

2. The conduct of at least some of the types of evaluation is either very expensive, or laborious, or is predicated on the availability of some alternative sources of information for purposes of comparison. Consequently, it is not surprising that the national census programs played a particularly important role in the development and application of several pioneering types of evaluation: the major importance (and budget) of the censuses provided the impetus and often the opportunity for the necessary effort and expenditure. Another major reason is that, by their very nature, censuses are particularly vulnerable to sources of error other than the "traditional" sampling error: first of all, even if censuses are based on samples, the sample sizes are typically so large that the sampling errors of at least national or large area estimates are negligible, hence the relative importance of other sources of error ("non-sampling errors") is predominant; second, the absolute magnitude of the non-sampling errors is typically unusually large because of the enormous job that has to be carried out in a short period of time, resulting in a several-fold swelling of the staff of the statistical offices involved and, in spite of best efforts, in lower standards of survey work than those that can be achieved in the context of small or continuing surveys.

3. Approximately ten years ago, a small unit was created in the Bureau with Bureau-wide functional responsibility for survey methodology. This unit, which originally comprised only a handful of professionals, has now grown to a staff of about 40 professionals. The evaluation of the reliability of survey data and the components of error is one of the major responsibilities of this group. The first significant achievement of this staff was the design and conduct of the evaluation programme of the 1961 Census. Subsequently, their effort was concentrated on the evaluation and redesign of the Labour Force Survey. During the last few years, a major effort has been underway to convert surveys to probability sampling methods where applicable, to redesign their methodology where necessary (whether or not sampling is involved) and to design and accomplish the automation of their processing. These activities will lay the groundwork for objective evaluation during subsequent years. The present paper is uneven from the point of view of a balanced presentation of the different types of error that surveys are subject to: it concentrates on those areas where there is some relevant Canadian experience to report.

II. Objectives of evaluation programmes

4. Four major objectives of evaluation programmes can be identified.

(a) The measurement of accuracy of survey results in order to guide users. It has always been the responsibility of statisticians to try and guide the users concerning the accuracy of the statistics they are producing. This responsibility, however, has considerably increased due to the fact that ever larger numbers of users have access to statistics (and, sometimes through the power of modern technology, to statistics representing breakdowns and disaggregations for which the survey was not designed originally) and also due to the fact that statistics are clearly recognized by all levels of government as well as by business and other users as being the source of information on which decisions of often far-reaching impact are made.

(b) Analysis of sources of error with a view to subsequent improvement. The determination of optimal designs for surveys involves an attempt to find that particular design which, for a given total budget, minimizes the overall errors (in actual practice there are, of course, other considerations as well, such as timeliness and practicality, to name just two). The attempt to find the optimal design will, however, only be feasible if one is able to identify the most important sources of error (as well as cost) and also the dynamics of error (and cost). Put differently, one clearly wants to concentrate on the major sources of error, i.e. on minimizing the total error at a given level of expenditure. Although theoretical considerations and pilot experiments may help in speculating about the magnitude of errors prior to the execution of a survey, it is only after the survey has been carried out that one can really evaluate the major sources of error. Therefore, particularly in the case of continuing surveys and in the case of occasional surveys with large budgets (such as censuses), it is often worthwhile to invest in the measurement of various sources of error in order to improve the future surveys. In contrast with point (a) above, evaluation for this purpose is in the tradition of research and development: one may have hypotheses and expectations, but the end result of the evaluation programme and its impact on particular surveys is unpredictable. Experience suggests, however, that the overall result will be reduction in cost or improvement in quality or both. Some notable examples from the Canadian experience will be quoted further on.

(c) Evaluation of alternative methods of survey design. The main objective here is not the measurement of absolute levels of error, but rather the measurement in the difference of error levels between the alternatives considered. This is a somewhat special problem analogous to the standard statistical problems of hypothesis testing. Without elaborating further, it may suffice to mention that we have carried out some evaluation studies during the last four years to assess alternative methods of census-taking in order to make decisions on the methodology of the 1971 Census.

(d) Evaluation for purposes of continuing control. In the case of continuing surveys it is particularly important to incorporate into the survey a regular programme of evaluations for the following two reasons:

(i) Since the survey is a continuing one, it affords unique opportunities for effecting a design feed-back on the basis of the results of the evaluation programme. More dramatic examples of such feed-backs are complete redesigns of surveys, such as was carried out in connection with the Labour Force Survey in 1963. Less dramatic but equally important examples are programmes to check regularly on sampling errors (through the regular computation of sampling errors), on coverage errors (through comparisons

with independent estimates or through re-enumeration), on content errors (possibly through the medium of re-enumeration). These programmes may result in various actions to tighten up some aspects of the survey which appear to be getting out of control, such as sending special procedural instructions to enumerators, or to redefine some of the concepts, or to modify (e.g. re-stratify) the survey in a minor way in certain geographic areas where the sampling error appears to be increasing.

- (ii) In addition to the technical challenge and opportunity to keep the survey operating at maximum efficiency, continuing surveys often present another type of justification for evaluation programmes. To the extent that the statistics derived from continuing surveys are often used to measure changes rather than absolute levels, it is essential for statisticians to ensure that apparent changes are not spuriously introduced into the survey results by changes in the levels of errors. In this connection changes in biases are even more important than changes in the magnitude of random errors since the former can introduce spurious trends into the statistical series. For example, if in a Labour Force Survey males 15 to 24 years of age suddenly begin to be missed in larger numbers than before, then to the extent males in this age group tend to have a higher level of unemployment, the measurement of unemployment over time can be seriously affected.

III. Evaluation of sampling errors

5. With a few notable exceptions, most of the major statistical series (as well as most of the occasional statistics) published by the Bureau are based on a complete enumeration, or on probability samples, or on administrative records, or on a combination of the procedures implied by these terms. It is appropriate to emphasize in the context of the present paper that the fundamental distinction between the probability sampling method and other methods of sampling is that the former facilitates the measurement of the sampling error using only the observations obtained within the sample. It is quite possible that other methods of sampling might be more accurate, cheaper, or more timely; however, only sample designs which are based on the principles of probability permit the measurement of sampling variability, i.e. the measurement of the probability that the sample estimate differs from a corresponding hypothetical census by a stated quantity.

6. Sampling error is clearly an important component of the total error of survey estimates, particularly for estimates based on small samples (even when the overall sample size is large it may be very small for many domains of interest, e.g. small areas). In the case of surveys which are well controlled with respect to non-sampling errors, it can be the most important component of the error for most of the resulting estimates. Therefore, it is an invaluable guide to users concerning the accuracy of the statistics. In the case of the Canadian Labour Force Survey, for example (one of the largest continuing probability samples in Canada), we have recently introduced the practice of indicating the range of sampling error of each published estimate through a numerical code.

7. In the case of the Labour Force Survey the estimation of sampling errors also paid off handsomely in the context of analyzing the sources of error. This is because sampling variability is not simply a function of total sample size, but depends on the sample design as well. The analysis of the components of sampling error in the early 60's indicated some major possibilities of improving the reliability of the survey through a thorough redesign of

the sample. This was carried out in 1962 and 1963 and resulted in a major improvement of the reliability of the survey. In fact, after its redesign, the new survey yielded estimates whose reliability under the old design would have required an across-the-board increase in the sample size by a factor of 2 to 3; moreover, this increased reliability was achieved while the cost of the survey was reduced by something like 20%.

8. There is no concrete example from recent practice in Canada of a situation where alternative sample designs were carried out and evaluated in a comparative sense. The process of redesigning the Labour Force Survey involved such a comparison of alternatives implicitly, i.e. without physically carrying out alternative designs; mathematical models were established for simulating the reliability of the survey as well as its cost under alternative designs.

9. As far as the aspect of evaluation for the purpose of continuing control, it is very much a feature of all our sample surveys of a continuing nature. Control charts are maintained, for example, for the Canadian Labour Force Survey which indicate the sampling variability of the major statistics over time at both the national and the provincial levels. These are designed to provide a warning signal when external changes affect unfavourably the original reliability of the sample (for example, through changes which render the original stratification ineffective).

10. Sampling variability reflects the impact of random selection of a sample. There are other kinds of sampling error which are not the result of chance; these are called sampling biases. Sampling biases themselves may be of two kinds. The so-called technical bias of estimates is most often the result of a conscious choice of a particular estimation procedure which typically introduces negligible biases but which, if used carefully, may reduce the sampling variability by a more than offsetting amount. A more important kind of sampling bias is introduced into the estimates when operational procedures result in the inclusion in the survey of units which were not selected or vice versa. An important example of this latter type of sampling bias is likely to arise in connection with our sample for the 1971 Census. Most of the subject-matter material included in the 1971 Census will only be asked of a one-in-three sample of households. To the extent that our Census enumerators deviate from the rule of systematic selection, sampling biases will be introduced into the census estimates. The instructions given to our enumerators are designed to minimize the impact of such sampling biases; at the same time, their extent will be evaluated. Certain basic statistics will be collected both from the sample households and from the remaining (non-sample) households. For these statistics it will be possible to compare the distribution within the sample with the corresponding distribution based on 100% of the households (these statistics will include type of household, household size, age, sex, etc). This comparison will actually serve the dual purpose of evaluation and immediate feed-back in the form of "corrections" (the weighting of the sample data is so designed that the weighted up sample tabulations agree with the corresponding tabulations based on the complete counts).

IV. Evaluation of coverage errors

11. In the case of a census, a coverage error occurs when a unit (person, household, establishment, etc.) which, according to the definitions, is within the scope of the enquiry, is either not included or is included in it more than once. Similarly, in the case of sample surveys, a coverage error occurs when

a unit which is within the scope of the survey has a smaller (typically zero) or larger chance of being selected than it ought to have according to the design. In the case of a household survey, the most frequent coverage error is one which results in the exclusion of a person from the sample survey for one of three reasons: the household to which he belongs fails to get included in the list of households from which the selection is made; he himself fails to get included in the survey even though some other members of the household to which he belongs are included; or (and this may become a phenomenon with increasing frequency) he has no regular attachment to households. In the case of surveys of business units coverage errors are most frequently the results of defects in the lists (in terms of the units included or in terms of the ambiguity of definition of these units) or the methods used to update them.

12. Coverage errors affect not only estimates of totals, but also estimates of distributions and averages. This is due to the fact that coverage errors typically do not occur at random. It is well known, for example, that in the case of population surveys, young males are particularly difficult to cover. Similarly, in the case of establishment surveys, the lists that are usually available are more complete with respect to the most important establishments than with respect to the smaller ones. To the extent that those units which have a greater than expected chance of being included in the survey are different from the others, a distortion in the distributions, as well as averages, will result. Typically, distortions will also occur between geographic regions or other disaggregations to the extent that estimates for these disaggregations are differently affected by the coverage errors. In the case of continuing surveys, a change in the level of coverage errors can introduce spurious trends resulting from a change in the distribution of coverage errors.

13. Thus coverage errors result in biases. For this reason, they cannot be estimated from the survey itself: their evaluation involves a comparison with an independent, superior standard. A few concrete examples are given below to illustrate Canadian efforts to measure and control coverage errors in different surveys.

Evaluation of coverage errors in censuses

14. There are three well-known methods of estimating coverage errors in censuses.

(1) The first method, and one that is most often used, consists of selecting a sample of areas and enumerating these shortly after the census, using particular care to ensure that all households within the selected areas as well as all persons within these households are included in the evaluation survey. The results of this type of coverage evaluation survey are matched record for record with the census in the selected areas and the discrepancies are noted. These discrepancies are either directly tabulated, resulting in estimates of net under-enumeration, or a subsequent reconciliation is carried out to resolve them prior to tabulation. This method has been tried both after the 1950 and 1960 U.S. censuses, as well as after the 1961 Canadian census. Our experience, as well as that of our American colleagues, indicates that this method of estimating coverage errors may grossly understate them, particularly the error due to missing persons in partially enumerated households. This is understandable because persons who are difficult to enumerate in the census are also difficult to enumerate in the evaluation survey; moreover, it is also the experience of several countries (including Canada) which conduct a continuing household survey, that sample surveys of households are hard put to match, let

alone exceed, the coverage achieved in the census. Re-enumeration surveys to evaluate coverage errors may be invaluable tools for the assessment of the relative magnitudes of some sources of coverage errors (even though they are deficient for the estimation of the total level of coverage errors): of completely missed households, of occupied households erroneously classified as vacant, etc. Also, if the quality of the re-enumeration surveys is reasonably constant from one census to the next, it may provide useful measures of the changes in coverage errors. Finally, in the case of local censuses (or test censuses), it seems to be the only feasible method of evaluation of coverage errors.

(2) An entirely new method, and one which has been used with considerable success in Canada after both the 1961 and 1966 Censuses, is called Reverse Record Check. It consists of tracing, in the census, the members of a sample of persons from a group of lists which, between them, include everyone who lives in the country at the time of the census. The first list is of persons enumerated in the previous census (it obviously includes both people who still lived in Canada at the time of the current census and who have died or have left the country since the previous census). The second list is of persons who have come to Canada or who were born between the censuses (again, it includes persons who left the country during the same period or who died); finally, the sample of persons who were identified during the previous study of the same kind as having been missed in the previous census represents those people who were living in the country during the previous census but who were missed in that census. After a sample was selected from each of these lists, the first task in carrying out this study is to identify the current address of every selected person. This operation is called "tracing". It is the most burdensome operation of the study. It involves, first of all, mailing to the last available address of the person and if the mailing is unsuccessful (i.e. there is no response after repeated follow-ups or the response indicates that the person moved to an unknown location) then an intensive field search is originated. This field search is typically starting from the last known address and is designed to find any trace of the person in question. It includes searching various local directories, contacting the occupants of the dwelling which was the last known address of the person in question, as well as the neighbours there, the owner of the local grocery store, the local clergyman, or any other local source which appears to hold out some promise of knowledge about the person. Files of drivers' licences and other national or provincial files are also searched. The field work associated with the tracing operation is carried out by the regular staff employed in the conduct of the monthly Labour Force Survey. In spite of the obvious difficulty of this operation, both after 1961 and after 1966 about 97% of the original sample of persons was traced. The final operation, after the determination of current address, involves searching the current census reports to ascertain whether the person was, in fact, enumerated at the address where he was living. Before it was accepted that a person was actually missed by the census, these persons were recontacted to find out if they might have been enumerated at an address other than the one to which they were previously traced. The census records were then checked for these additional addresses. Once a person is found to have been missed by the census, some additional processing permits the determination of whether he was missed as part of a whole household missed or as part of a partially enumerated household. Some data may also be collected from these persons in order to analyze the characteristics of missed persons and hence the impact of coverage errors on the substantive census estimates.

The major advantage of the Reverse Record Check over the re-enumeration method of estimating coverage errors is that it provides a completely

independent approach. Thus it avoids the major problem of evaluating coverage errors; that is, that the evaluation method might itself miss the same persons who were missed in the census. This is much more likely to happen in a re-enumeration study than in the Reverse Record Check since in the latter method people who are currently in a critical age group (i.e. in an age group that is particularly vulnerable to coverage errors) are largely selected from the list provided by the previous census at which time they may not have been in this critical age group, hence they were more likely to have been enumerated. Also, here one is trying to locate particular persons rather than carry out vague instructions ("enumerate everyone"). A second advantage is the ability to distribute resources (i.e. allocate the sample), in the population groups known or suspected to have a high tendency to be missed in the census, in accordance with the usual principles of efficient sample design. The disadvantage of the method lies in the difficulty of tracing to ensure that all possible addresses at which the person may have been enumerated are accounted for and the usual matching problems (matching with the current census).

The Reverse Record Check has several interesting analytical by-products as well. It enables the tabulation of missed persons by their characteristics in the previous census (to the extent they were selected from that source). It also enables the tabulation of some information concerning intercensal migration (since it involves matching between consecutive censuses). It provides interesting possibilities for assessing the accuracy of age reporting of infants and smaller children (since this part of the study is based on locating a sample of children selected from birth registers). It provides some possibilities of checking the accuracy of reporting of country of origin (since the country of origin is known for that part of the sample which is selected from immigration records). Finally, the study appears to be the only method, for Canada, of obtaining estimates of emigration (since every person who was living in Canada five years ago is traced, including the ascertainment from friends or relatives that he or she has left Canada since that time).

(3) The third known method of estimating coverage errors is the so-called analytical method. This method uses demographic techniques of age cohort analysis to estimate the current population by age and sex. It starts from the count of persons in the previous census (by age and sex), adds to their ages the number of years between the censuses, applies the appropriate death rates, includes for the lowest age groups the known birth registrations adjusted by infant mortality and uses some advanced iterative techniques to make an allowance for the effect of coverage errors in the previous census as well as emigration. The main weakness of the method is the way it deals with external migration and persons missed in the previous census. Therefore, it is not particularly suited to "open" countries, such as Canada, where this may involve a significant number of persons. Moreover, this method is not capable of estimating coverage errors by regions (at least in countries, such as Canada, where external and internal migration is a major phenomenon). Finally, it does not easily lend itself to the classification of the number of missed persons except by age and sex (because of the difficulty of establishing a model to predict with reasonable accuracy the behaviour of other characteristics over time).

Evaluation of coverage errors in the Canadian Labour Force Survey

15. Estimates of the number of persons by age, sex and province of residence are prepared monthly during the inter-censal period. They are based on the previous census counts, adjusted for estimates of deaths, immigration,

emigration and internal migration. These population estimates have been found to be acceptably reliable at the national level and definitely useful, though less reliable, at the provincial level. They are used as the "standards" of comparison to measure the coverage errors of the monthly Labour Force Survey and to correct some of the resulting bias. The sampling terminology for the technique is "ratio estimation". It involves estimating from the Labour Force Survey itself the number of persons by age, sex and province and adjusting these numbers to agree with the inter-censal population estimates. The adjustment factors themselves indicate the magnitude of the coverage error in the Labour Force Survey by age, sex and province (at least coverage errors relative to the inter-censal population estimates which are, typically, themselves subject to similar coverage errors that the censuses are subject to). It is the Canadian experience, as well as the experience in the U.S. with similar surveys, that the coverage in the Labour Force Survey is deficient in relation to this "standard" (and hence, in relation to the coverage of the census itself). This under-enumeration in the Labour Force Survey, which we call "slippage", runs at about 3-4% but it is significantly higher in the critical age group of young males.

16. The monthly estimates of slippage are carefully watched to ensure that the coverage of the Labour Force Survey does not deteriorate. About a year and a half ago, in fact, when such a deterioration appeared to have been detected, special instructions were sent to enumerators and some of our procedures were revised. This measure brought the coverage error of the survey "under control" again.

17. In addition to measures of "slippage", the non-interview (non-response) rates are also tabulated and examined monthly. We do not consider non-interviews as coverage errors since in this case the appropriate persons were presumably found, but they were not interviewed for one of several possible reasons. However, we do recognize that non-interviews represent a source of bias and the level of non-interviews is carefully controlled.

Evaluation of coverage errors in the monthly Retail Trade Survey

18. The new design of the monthly Retail Trade Survey (to be introduced later this summer) provides an interesting application of the evaluation of coverage with "quick feed-back". This survey is based on a sample selected from our central list of establishments. The source list, however, includes in its coverage only establishments which have at least one employee covered by the Unemployment Insurance Act (this includes practically all establishments with employees). In addition to the deficiency of the list with respect to establishments with no employees (which is coverage error "by design"), the list suffers from other coverage errors as well, primarily due to late identification of "births" (new businesses). In order to supplement the list sample for its known deficiencies, a procedure has been devised which involves a monthly canvassing of all retail establishments in a sample of areas. The establishments found in the sampled areas are matched with the list maintained in the Bureau and those retail outlets not already in our central list are identified. These remaining establishments provide a sample representing all retail establishments not on the central list and are included in the Retail Trade Survey. In this instance the purpose of the area sample is not primarily the evaluation of coverage errors but rather to supplement the list sample in order to complete the survey coverage. The procedure will, however, provide at the same time a useful evaluation of the coverage errors of the list of establishments maintained in the Bureau, at least within the industry classes that

are included within the scope of the Retail Survey. In particular, it will provide estimates of the number of establishments which, by definition, are not covered by the list (those with no insurable employees) as well as of the number of those who should be on the list but are either not included or included only with a time lag.

V. Evaluation and measurement of response errors

19. Response errors may occur whenever data are requested, provided, received or recorded. Questions may be misinterpreted by the respondent; he or she may not know, may not remember, or may purposely want to distort the correct answer. Different enumerators may have different tendencies in putting and explaining questions, or interpreting responses. Response errors can be classified into two categories:

20. (i) Response variance is that component of the response errors which has a chance to cancel over a large number of responses. Some interviewers may tend to inflate a count while others tend to deflate it; some respondents tend to over-state their response to a certain question, others tend to under-state it; some questions on the questionnaire may be ambiguous (or difficult to communicate) and may invite misunderstandings in either direction. All these causes (and many others) contribute to response variance. Because this type of error may be in either direction, the net error from this source may be quite small for large areas, but it can be very large for small areas or rare characteristics. It also tends to be much larger for sensitive characteristics which are normally difficult to measure.

21. As part of the evaluation programme of the 1961 Census, a project was carried out to measure the response variance, particularly the component of the response variance that can be attributed to interviewers. Although this study was restricted to a purposively selected area of the country (an area considered to have generally similar characteristics to the remainder of the country), this project had an enormous impact on the procedures eventually adopted for the 1971 Census. A brief description of the project is provided below. It was based essentially on the method of interpenetrating samples originally proposed by Professor Mahalanobis and later developed by several other authors.

22. Suppose one wants to estimate the response variance of a census total for an area (such as the number of persons with high school education in a particular township). What one would like to have is at least two census estimates for the same area so that their difference would give an indication of the response variance. In fact, without actually repeating the census one can obtain two estimates for the same area by assigning half of the households at random to one crew of enumerators and the other half of the households to another crew of enumerators. The two crews produce two estimates for the total referring to the same area (for example, two estimates of the number of persons with high school education in a particular township). One can actually say that the two crews of enumerators produce two "census" estimates of the same total for the same area, except that each of the estimates is based on the enumeration of only half the households in the area. Correspondingly, the difference between the two estimates will reflect two sources of variance: (a) the response variance between estimates taken at the same time, under the same general conditions, but using different enumerators and (b) the sampling variance between the two samples, each comprising half the households in the area. By taking the square of the difference of the two estimates, one can estimate the sum of the response variance and the sampling variance. At the

same time, but by a different technique, the sampling variance can be estimated separately and then, by subtraction, one may obtain an estimate of the response variance. This was essentially the procedure for the project mentioned above (with some refinements described elsewhere).

23. The estimates of response variance derived from this project had a major impact on our view of census errors. It appeared that on the average the response variance due to enumerators was of the same magnitude as the sampling variance would have been if the census was carried out using a 25% sample instead of 100% enumeration. It became quite clear that for many characteristics the contribution of enumerators to the response variance was overwhelmingly the largest source of error of small area estimates. As a consequence, it was clear that our census methodology was not in optimal balance with respect to cost and error: if the response variance due to enumerators could be significantly reduced, one could introduce sampling into the census and still achieve a net reduction of errors, at the same time achieving some reductions in costs and a definite improvement in timeliness. As a result of these considerations, plus a good deal of testing, a new methodology for the 1971 Census emerged. The discussion of this methodology is not within the scope of the present paper. The point to be emphasized, however, is that the methodology of the 1971 Census was influenced in a very significant way by the response variance study.

24. No explicit measurement of response variance has been carried out as yet in relation to the other surveys of the Bureau. However, it is important to emphasize that in the case of mail surveys the use of the well-known formulae to estimate sampling errors automatically ensure that the response variance is also estimated; in other words, what are assumed to be estimates of sampling variance are really estimates of the sum of the sampling and the response variance. Hence, as far as the assessment of the overall reliability of these surveys is concerned, both for the guidance of users and for overall control purposes, the usual estimates of sampling variance are valid. In the case of surveys involving interviewers, the same fortunate situation can be secured with a bit of care: it requires that interviewers by and large should not cross the boundaries of so-called "primary sampling units". The details of this consideration, however, would lead beyond the scope of the present paper.

25. (ii) Response bias - This is, roughly speaking, that portion of the total response errors which is "left over" after all cancellation involved in the response variance has occurred. It is made up of response errors which have a tendency to occur more in one direction than in the other. It may be the result of the training and attitude of interviewers; or the reluctance of respondents to admit certain characteristics or a tendency to consistently over-rate or consistently under-rate them; or it may be the fault of the questionnaire in that it may invite errors in one direction more than in the other. Such errors do not cancel out even over large areas or a large number of respondents and may be particularly damaging for statistics at the metropolitan area, province or national level. As in the case of response variance, response bias also has a tendency to be much larger for characteristics that are difficult to measure.

26. The concept of response bias implicitly assumes the existence of "standard" or "true" estimates, response bias being essentially the net departure from such standards. Consequently, the measurement or evaluation of response bias involves a comparison of the survey in question with this "standard" may take the form of a special survey conducted with particular care, specifically designed with the objective to provide improved measures which can

serve as approximations to the "standard". Alternatively, it is sometimes reasonable to assume that certain data reported in administrative records are more reliable than the corresponding data reported in statistical surveys. In such cases, tabulations from the administrative records may serve as "standards". (An example of this latter type of situation occurs in the case of certain types of income data which are often assumed to be more reliably reported in the income tax records than in statistical surveys.) Finally, response bias may sometimes be detected (though not measured) on the basis of analyzing departures in the data from hypothetical models (internal consistency analysis). A typical example of this latter situation is the analysis of age reporting: it is observed in many surveys that the number of persons reporting ages that end with a 0 or 5 is unreasonably large in relation to the number reporting ages ending in 9's or 1's or in 4's and 6's. Response bias is notoriously difficult to measure or even approximate and sometimes none of the known methods apply, or apply by itself.

27. Several studies have been carried out with a view to estimate the response bias of some of the measurements recorded in the census. Comparisons have been made between the 1961 Census and the Labour Force Survey carried out at about the same date concerning estimates of unemployed, employed and some breakdowns of these (in this case, the Labour Force Survey was assumed to be the "standard"). Studies have also been carried out to estimate the response bias of income data. Also, an analytical study was completed relating to errors of age reporting. Partly as a result of these studies the questions in the general area of economic characteristics of persons have been considerably changed between 1961 and the 1971 Census. Similarly, the question on age has been altered and we will now be asking for date of birth instead.

28. In the monthly Labour Force Survey the technique of interpenetrating samples, briefly mentioned above, is used to ensure that the estimates of sampling variance actually include the response variance as well. In addition, a subsample of the households that were enumerated by the regular enumerators is re-enumerated each month by supervisory staff. The results of this re-enumeration are compared with the regular enumeration. To the extent that the re-enumeration survey is "better" than the original survey, the differences in the statistics obtained indicate the biases in the main survey. Although the monthly sample used in the re-enumeration survey is small, these samples can be accumulated over a quarter or a full year to provide estimates of response bias which have smaller sampling variabilities attached to them. In addition to comparing the estimates derived from the re-enumeration survey with the main survey to estimate net response bias, the re-enumeration survey data are also matched household by household with the results of the original survey. The resulting "micro-comparison" enables the tabulation of the so-called gross difference rate, i.e. the number of cases in which the enumerator and re-enumerator differed from one another expressed as a proportion of the total number of cases involved. This rate, it can be shown, is actually a minimal estimate of the total response variance (it measures, roughly speaking, that portion of the response variance which is caused by reasons other than the enumerator). The gross difference rate is easy to compute and, together with the estimate of response bias mentioned above, serves as a useful continuing control of response errors in the Labour Force Survey.

29. A particularly interesting application of the evaluation of response errors with "quick feed-back" occurs in the case of the Job Vacancy Survey. This is a survey of business units designed, as its name indicates, to estimate the number of job vacancies (unfilled jobs) in the economy. The survey is primarily a mail survey, but a subsample of the business units included in the

mail survey is re-enumerated by interviewers. The re-enumeration involves a probing approach which would not be feasible in a mail survey. Hence, the re-enumeration survey provides "improved" estimates for a subsample of the units included in the mail survey. The improved estimates are used during the processing of the survey to eliminate the response bias of the mail survey. The resulting combined estimate is subject to a smaller sampling error than that of the interview survey alone; at the same time, it is free of the response bias that the mail survey alone would be subject to. Moreover, the interview survey provides a regular continuing evaluation of what the response bias of the mail survey would be without the correction of the interview survey. This will enable us to assess the optimum sample size of the interview survey whose unit cost is considerably higher than the unit cost of the mail survey.

VI. Non-response bias

30. Non-response occurs when the required statistical information is not secured for a unit (person, establishment, etc.) which is known to exist, is known to be within the scope of the survey and is selected for inclusion in the sample. Instead of simply tabulating the results of the survey as obtained from the respondents, it is the standard practice of survey takers to assign, as best they can, implicitly or explicitly values (imputations) for non-respondents. This is done in order to ensure that the levels of the estimated statistics are not seriously under-stated and to ensure that such "imputations" are carried out in a uniform manner for the benefit of all users by the agency best qualified to do so: i.e. the statistical office which originally collected the data.

31. Clearly what is involved here is a special kind of "response error". The values assigned to non-respondents, without the benefit of actual measurement, are obviously different from those that should have been obtained from the persons, establishments, etc. concerned. Often, the imputations involve the assignment of some average values to the non-respondents (the average reported by a sub-group to which the non-respondent unit is known to belong). Alternatively, in the case of continuing surveys involving the same units, the trend may be estimated from the respondents and then this trend may be applied to the last actual response of the unit which is non-respondent in the current survey. Whatever the actual imputations procedures, they are clearly biased. The evaluation of these biases involves obtaining real data for the non-respondents and comparing the real data with the imputed ones. Of course, evaluation is no alternative to striving to contain the magnitude of non-response at a level where its potential impact on the estimates, with any reasonable rule of imputation, is minimal.

32. The only known unbiased way of dealing with non-response errors is to select a random subsample of non-respondents, enumerate them using specially intensive methods to secure responses, and to compile an unbiased estimate for the non-respondent group in the survey from this subsample. This is essentially what is done in the Job Vacancy Survey that was already referred to above. In this survey a random subsample of the mail sample is selected at the time when the mail questionnaires are sent out. At this time, of course, the non-respondents to the mail survey are not known yet. In spite of this, however, interviews are obtained from every unit in this subsample. At the end of the mail survey the interview schedules are matched with the mail survey schedules and are sorted into two categories: those interview schedules which are, in fact, re-interviews of units which completed the mail questionnaires and those

which relate to units that were non-respondent in the mail survey. The former group of interview schedules is used, as indicated above, to evaluate and correct the response bias of the mail survey. The second group of interview schedules is used to provide an evaluation of the impact of non-response on the mail survey results, as well as to provide an unbiased estimate for the non-respondents which is directly used in the compilation of the mail survey results.

33. The Job Vacancy Survey represents the most interesting example within the Canadian example of the measurement and correction of non-response errors. It is a new survey and our experience will be carefully watched and assessed with a view to its potential applicability to other surveys. Generally speaking, however, in most other surveys we do not have an unbiased method of dealing with non-response errors. However, as mentioned in section 4 in connection with the Labour Force Survey, we do compile regularly non-response rates and these are carefully watched. The level of non-response in our interview surveys is generally small: well below 5%. In mail surveys the level is significantly higher and it varies considerably from survey to survey.

34. The Dominion Bureau of Statistics is presently involved in a major drive to automate, as much as possible, the processing of surveys. The automation involves, among others, computer methods to identify non-respondents and to carry out imputations for them. This provides special problems and possibilities as far as non-response errors are concerned. The automation of imputations holds out a major possibility to compile early estimates on the basis of incomplete returns: the computer is capable of carrying out the imputations for a small or large number of non-respondents with equal ease. To the extent that these early estimates are based on incomplete returns, they are subject to larger non-response errors. However, since the publication of timely statistics is also one of the major objectives of statistical offices, there is considerable pressure to produce estimates as early as possible. Under these circumstances it is particularly important to have a continuing program which, though it may not estimate the total non-response error, it would, nevertheless, at least assess the increase in the non-response error that early estimates are subject to in comparison with the final estimates. Automation also provides some tools to reduce the impact of non-response errors: besides the power of computers which enable statisticians to design elaborate and involved methods of imputation, the computer can also be used to provide an early identification of what may be called priority non-respondents. These may be units for which, if they do not respond, the imputations would have an unduly large impact on the final estimates (such units may either be large companies or establishments whose statistics impact heavily on the current estimates, or medium-sized units for which experience indicates that it is particularly difficult to impute with reasonable accuracy due to their volatility). In the case of continuing surveys the computer can also indicate units that have been non-respondents in two or three consecutive surveys. For these repetitive non-respondents special effort may be mounted since the imputations, as mentioned above, are typically based on their last available response (hence it is particularly important to ensure that the last available response is not too far out of date). Various combinations of methods of identifying priority non-respondents are incorporated in the design of the automated processing systems of our new Retail Trade Survey, and experiments are planned in connection with other establishment surveys. In the case of the Retail Trade Survey our procedure calls for a telephone follow-up of the priority non-respondents, as well as a field follow-up in the case of any unit which has been non-respondent for three consecutive surveys.

35. A number of automated surveys are designed to get into production during 1970. It is too early, however, at this point to report on our experience with their evaluation from the point of view of non-response errors.

VII. Processing errors

36. Errors are, of course, introduced into survey data not only during collection but during processing as well. The theoretical considerations concerning distinctions between bias and variance and the measurement of each are essentially similar to those pertaining to response errors. The major difference concerns the fact that processing generally occurs at the headquarters of the statistical office and is, therefore, capable of being controlled. In fact, the process of control may also serve the dual purpose of evaluation. Standard techniques of quality control enable statisticians to measure the so-called "average outgoing quality" or put differently, the error of processing. This can serve the purposes of evaluation. In fact, in the case of large enough operations, it can serve the purposes of "process control", i.e. the identification of the worst clerks, coders, etc. for re-training or replacement. Quality control techniques will be applied to several operations in connection with the 1971 Census. On a regular continuing basis, they are applied to the key-punch operation.

THE NEED FOR INDEPENDENT CROSS-CHECKS
OF STATISTICS FROM DIFFERENT SOURCES.

THE EVALUATION OF THE ACCURACY OF
SURVEY RESULTS.

INTRODUCTION

1. It is not only important that data are made available to policy-makers and administrators, but it is also essential to ensure validity of such data. An estimate based on an efficient sampling design and having very little bias, if any, reasonably low sampling error and low non-sampling error can be considered to be a valid estimate. Non-sampling errors arise due to inadequate or imperfect specification of the data to be collected, lack of precise definition of the boundaries of area units, lack of qualified and trained investigators, non-response and recall lapse, deliberate fudging and defective data processing operations. Unless there are built-in checks at each stage, it is not possible to control such errors. These internal checks may take various forms such as (i) provision of internal consistency checks through proper designing of schedules, (ii) interpenetrating net-work of sub-samples (IPNS), (iii) rationalised supervision on a random basis. For imparting a certain measure of credence to the results, however, it is necessary to have cross-checks preferably by independent sources. Various measures have been adopted in our country to assess the accuracy of the data collected both through the survey methods as also through administrative channels.

2. In cross-checking, data from different sources should provide checks against one another. Unless one source can throw up information that can check against the other at different stages, and provide valid information for improving the system, the alternative data are just duplication and not cross-checks. In case the result of an independent survey is different, there is no means of objectively determining the validity of either of the series. Unless, therefore, there has been a deliberate attempt to design the surveys in such a manner that cross-checking is possible at various stages of the survey process, the purpose behind the survey as a cross-check is lost. This of course is not applicable to the situation such as, for example, the one in which the validity of production estimates are checked against consumption statistics. In comparing two series of data based on different concepts and definitions, it will be obviously necessary to adjust, if possible, the data for these differences before comparison is made, or keep in

view the essential differences in drawing conclusions where such adjustments are not possible.

3. It may be stated that with the ushering in of the era of planning in the fifties, there has been a growing demand for various types of data. Since data obtained as a by-product of administration failed to meet all the requirements, sample survey methods had to be increasingly used. Moreover, in a country like India arrangements had to be made for the collection of data in fields where physical records were conspicuous by their absence. The methods adopted were mostly those of interview through which events happening in the past one or two years or months or weeks were to be recalled. In quite a number of situations the very nature of the desired information was such as to produce biases leading to deliberate suppression or over-statement of information. In spite of these inherent difficulties, information on these subjects at the national, state or regional levels was considered essential. Information collected under such circumstances results in throwing up quantitative estimates which require to be used with adequate caution, and suitable measures to control the errors and biases become indispensable. The measures taken in this regard are reported in the subsequent paragraphs.

CROSS CHECKS

Interpenetrating Net Work of Sub-samples (IPNS)

4. The technique of selecting the sample in the form of two or more sub-samples, each capable of providing a valid estimate of the population parameter, has been in vogue in this country particularly in the National Sample Survey (NSS). The results of such sub-samplewise information can be used not only for obtaining margin of error but also for securing information on non-sampling errors, such as differences arising from differential interviewer bias, different methods of eliciting information, etc. However, in case non-sampling errors are of the same sign and similar magnitude, the IPNS design will fail to bring out the impact of the non-sampling errors. For example, we may consider the estimates of vital rates obtained through the N.S.S. Even though in most of the rounds, the sub-samplewise estimates of vital rates have closely agreed, it is contended that they are under-estimates. This technique has been extended not only to investigator differences, but also to agency differences. For example, the States in India participate on a matching basis in the work of the NSS. Thus two sets of sub-samplewise estimates are available, produced by two different agencies. It may happen that the separate sub-samplewise estimates of Central and State NSS closely agree among themselves but there are large differences between the Central NSS results and the State NSS results. In such a case the differences may be due to differences in the quality of the field staff of the two agencies or may have arisen at the processing stage. An investigation into these will be very fruitful to find out ways and means of improving the estimates.

Vital statistics

5. The events netted by the normal registration records are known to be seriously under-enumerated. As a short-term measure to provide reliable and quick estimates of vital rates at State and all India levels, a Sample Registration Scheme (SRS) of births and deaths has been initiated by the Registrar General. The SRS combines the advantages of both continuous (longitudinal) enumeration and survey procedures. Information on vital events in sample areas is collected continuously on a de jure basis by a locally resident enumerator through a built-in system and the continuous enumeration is supplemented by an independent half-yearly survey conducted by a supervisor. This half-yearly survey by the supervisor has been provided to ensure that the events which might be missed or erroneously included by the enumerator as part of continuous enumeration are detected and rectified. Likewise, the continuous enumeration record acts as a cross-check over the events listed by the supervisor during the half-yearly surveys. An independent cross-checking of SRS data has been undertaken by the NSS in West Bengal and Tamil Nadu. The sample villages covered by SRS have been completely enumerated by the NSS investigators in these two States. The results so far available in respect of West Bengal show that there are deficiencies in the enumeration of births and deaths by both NSS and SRS agencies. These results have indicated the lines of action needed for improving the investigation by both the agencies. The results of Tamil Nadu study are not yet available.

Production estimate of jute

6. One of the earliest attempts to cross-check the official estimates of production of an important crop was made when the official forecast figures of jute production in Bengal during 1944-45 and 1945-46 were compared with the data on production obtained through sample survey. A third estimate of production of jute was also available based on accurate trade estimates, but this could be compiled about 15 months after the harvesting season. In both the years mentioned above, the sample survey estimates were quite close to the trade figures whereas the official forecast figures were very divergent. The official estimates of area under jute though claimed to be based on plot-to-plot enumeration suffered from serious defects. Furthermore, for the estimation of yield rate used in the official forecast, the plots selected for crop cutting were not chosen on a random basis. The results showed that the official forecasts prepared in the traditional manner left much to be desired, and that the sample survey results provided closer estimates to trade figures. As a result, sample survey method has replaced the traditional method of crop estimation.

Production estimate of cotton

7. The official estimates of production of cotton are prepared and issued by the Ministry of Food, Agriculture, Community Development and Cooperation. These are based on the results of area enumeration by Patwaris (village revenue officials) and random crop cutting experiments conducted by the States. Estimates of production of cotton are also framed by the Trade, based on the reports of the net-work of their agencies in the cotton-growing areas. Staff and agencies employed by the Trade for this purpose are believed to have intimate knowledge of the local conditions and frequent contact with the concerned parties. They also keep in touch with market arrivals. On the strength of such information, the Trade estimates are prepared which serve as a cross-check over the official estimates. Estimates based on mill consumption, pressing and ginning returns are also available. These three sets of estimates have been found to differ and this has led to intensive investigations being undertaken in order to ensure the quality of the official estimates.

RATIONALISED SUPERVISION

8. The Agricultural Statistics Division (ASD) of the Directorate of National Sample Survey exercises supervisory control over the crop cutting work by the States. A part of the programme of inspection consists in randomised check over the field work of the State governments. A certain number of villages for each crop in each State is selected with equal probability from the list of villages selected by State Governments for crop cutting work. The main aspect of inspection of field work consists in visiting the selected villages on the appropriate day fixed for harvesting of sample fields, ensuring the accuracy of the conduct of crop cutting experiments and final weighment of green produce. The inspection has two objectives, viz., (a) serving as a cross-check, though not independent, over the State estimates; and (b) increasing the accuracy of the final estimates by way of improving the quality of the basic data.

ASSESSMENT SURVEYS

Post enumeration checks

9. The data collected in the decennial population censuses are affected by a variety of non-sampling errors as in any other survey, consisting of (a) coverage error and (b) content error. In the past two censuses, there had been post-enumeration checks (PEC), conducted by the census organisation itself, oriented towards measuring the accuracy of the census data in respect of the coverage error only. In the PEC of the 1971 census an attempt will be made to assess both the coverage and the content errors.

10. A quinquennial livestock census is carried out under the auspices of the Ministry of Food and Agriculture, the field work of which is organised through the staff of the State Governments. The census includes a built-in provision for rationalised supervision over the field work to improve the quality of the basic data. Besides, independent post-enumeration check surveys have been organised through the NSS to assess both coverage and content errors.

11. The ASD had organised a series of independent surveys to assess the additional area irrigated through minor irrigation projects. The Ministry of Food and Agriculture, which is responsible for the implementation of the programme of minor irrigation works, receives periodical progress reports from State Governments about achievement of these programmes. The ASD organised several surveys relating to masonry wells, pump sets, tubewells, tanks, etc., over a period of eight years in various States. In the first instance the detailed lists of works maintained at various administrative levels were examined. The second stage consisted in visiting the works selected and collecting, by spot-check and detailed enquiry, information on the extent of area irrigated by them. The results supplied useful information for improving the official records about the number of works constructed and area irrigated per work.

The need for independent cross-checks of statistics
from different sources. The evaluation of the
accuracy of the survey results

1. The need for independent cross-checks of statistics from different sources is a real one and has been recognised here since the Bureau came into existence. Owing to the fact that Lesotho does not have customs posts at her borders, our trade statistics have had to be collected by the round-about method of sending questionnaires to all possible importers and asking them to state quantities and values of imported commodities classified into the broad S.I.T.C. divisions. Figures so obtained are supplemented by those derived from a sample of gate checks to cover migrant workers and day-to-day shoppers (See 6 & 7 - Lesotho).

2. It has been possible in the case of a few imported commodities to conduct independent checks which have always yielded valuable information.

(a) Cigarettes and Tobacco:

Apart from information supplied by traders on the general questionnaire, major tobacco firms in the Republic of South Africa which are known to supply Lesotho traders are asked to give figures of their exports of cigarettes and tobacco to Lesotho customers and these have been found to be invariably higher than those obtained from our traders.

(b) Mineral Fuels:

Imports of petrol, oils, greases, paraffin etc. are also obtained separately from the major petrol companies (Shell, Caltex, Total etc) of South Africa to provide a check against those that are obtained from local importers. These have also been found to be higher than those obtained from Lesotho importers.

(c) Wool and Mohair Exports:

Wool and Mohair constitute Lesotho's major agricultural exports. Even long before the Bureau came into existence there were suspicions of understatement of quantities exported to evade levies on these items. No effective way of checking has as yet been introduced and for our statistical purposes we rely on figures of quantities sold at the South African ports (floors) of these commodities. The availability of the figures has been made possible by the wonderful co-operation of the South African Wool Commission and Mohair Board. These figures are always higher than those reported by Lesotho traders as their exports to South Africa despite the fact that all traders handling agricultural

produce are included in our enquiries.

3. In the cases of cigarettes, tobacco and mineral fuels perhaps the discrepancy between trade census figures and those obtained from South African firms can be accounted for by the insufficient coverage in our enquiries. This is brought about by the obvious deficiencies in the available lists of traders and the fact that smaller traders are not all covered; instead a small sample is selected. Non-response could also contribute towards the discrepancy. However, attempts are constantly being made to close whatever gaps we can and on the whole there is an improvement in our trade figures.

4. In conclusion, it should be pointed out that independent cross-checks of statistics from different sources have proved very valuable and the Bureau is always looking for possible chances of discovering more areas in which they could be applied.

5. The data collected in a survey could be compared with some generally accepted knowledge about the characteristics involved or their relationships. For example to evaluate data on population distribution by age groups, marital condition etc. as collected in a census of population, the following relationships are commonly used in quality checking.

- (a) Population distribution by age groups tends almost to follow a standard pattern.
- (b) The total number of married men should be almost exactly equal to the total number of married women. Only in polygamous and polyandrous societies will the two numbers be unequal.
- (c) Females generally marry at a younger age than males; accordingly in the younger groups one would expect to find more men than women single, and there would be but very few widowed or divorced people at these young ages.
- (d) It is also known that the death rates for men tend to be higher than those for women and that the remarriage rate for widowers is higher than that for widows, with the ultimate results that there are more widows than widowers in the older age groups. The older age groups also contain many more divorced persons than do the younger age groups, that is, in those countries where divorce is legally recognized.

N.B. (a) was quoted from "Techniques of Population Analysis" by G.W. Barclay.

(b), (c), and (d) were quoted from "Handbook of Statistics Methods for Demographers" by A.J. Jaffe.

6. As an actual example in Lesotho, in the analysis of the results of the 1966 Population Census of Lesotho, the age group 0-4 years was found to be consistently lower than the next higher age group, 5-9 years through all the nine districts and the country as a whole. Beside this age group,

some other age groups, for example 45-49 years, 65-69 years both for male and female were higher than one or two lower age groups. The census report recognized that these results were inconsistent with standard distributions, stated that this is largely due to the fact that the number of infants under the age of one year was under-enumerated by a considerable amount. For other age groups, it further stated that the age distribution of Lesotho Population showed a marked departure from the normal. The number of persons aged 33, 48, 66, 69, 86 were in all cases higher than the number of persons immediately below these ages. (See Table 1).

7. As to the other population characteristics and relationships mentioned in (b), (c), and (d), the ratio of married men and women in the results of census was calculated as 1:1.08 for the population of age higher than 15 years. If Lesotho were to be considered a monogamous society, then the results of the census for this characteristic would be quite an acceptable return. About marital status of females, in the age groups under 30-34 years, married female population was much more than male population. More widows than widowers in every age group, especially in the older age groups were found, and population of female in the older age groups was always having more persons than that of male, etc. These census returns were generally recognized consistent with the acceptable knowledge of population characteristics and relationships mentioned in (5) above. (See Table II).

T A B L E I

Population by Sex, by 5 years Age groups
(The 1966 Population Census of Lesotho)

AGE GROUP	MALE	FEMALE	TOTAL
0 - 4	62095	63341	125436
5 - 9	64796	64286	129082
10 - 14	62483	62086	124569
16 - 19	46747	49862	96609
20 - 24	34611	39291	73902
25 - 29	32106	33871	65977
30 - 34	30761	34069	64830
35 - 39	25523	25805	51328
40 - 44	21039	21937	42976
45 - 49	24596	24883	49479
50 - 54	17063	20211	37274
55 - 59	12242	13924	26166
60 - 64	9085	12315	21400
65 - 69	10220	16090	26310
70 - 74	4151	8164	12315
75 - 79	2251	3924	6175
80 - 84	1209	2452	3661
85 - 89	1650	3808	5458
90 - 94	276	785	1061
95 & over	230	636	866
Not Stated	2650	2110	4760
T O T A L	465784	503850	969634

T A B L E I I

Population Aged 15 years and Over, by Marital Status by Sex, by 5 year Age Groups (The 1966 Population Census of Lesotho)

AGE	MARRIED		SINGLE		WIDOWED		DIVORCED		NOT STATED		T O T A L		
	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	MALE	FEMALE	T O T A L
15 - 19	810	10879	45880	38751	3	72	10	97	1	5	46704	49804	96508
20 - 24	8989	29643	25380	8277	54	506	91	781	22	20	34536	39227	73763
25 - 29	21197	28736	10340	2870	170	1023	299	1141	16	13	32022	33783	65805
30 - 34	25275	28286	4454	1858	344	2237	576	1575	10	14	30659	33970	64629
35 - 39	22148	20674	2251	1122	365	2559	644	1348	6	13	25414	25716	51130
40 - 44	18531	16496	1314	830	479	3423	617	1098	5	4	20946	21851	42797
45 - 49	21762	16692	1081	783	820	6182	811	1135	8	8	24482	24800	49282
50 - 54	14909	11295	609	449	866	7614	610	751	-	9	16994	20118	37112
55 - 59	10629	6568	347	271	809	6552	390	450	7	5	12182	13846	26028
60 - 64	7741	4693	239	205	804	7071	261	286	1	4	9046	12259	21305
65 - 69	8427	4347	199	175	1282	11258	288	278	2	5	10198	16063	26261
70 - 74	3227	1593	92	85	739	6369	74	101	1	3	4133	8151	12284
75 - 79	1759	645	33	40	417	3196	34	36	-	2	2243	3919	6162
80 - 84	857	317	24	24	299	2088	23	20	1	2	1204	2451	3655
85 - 89	1134	339	17	25	476	3417	20	22	1	2	1648	3805	5453
90 - 94	163	58	4	9	102	714	6	2	1	1	276	784	1060
95 & Over	138	52	6	7	81	572	5	4	-	-	230	635	865
Not Stated	1549	926	892	646	60	444	37	52	108	38	2646	2106	4752
T O T A L	169245	182239	93162	56427	8170	65297	4796	9177	190	148	275563	313288	588851

THE NEED FOR INDEPENDENT CROSS-CHECKS
OF STATISTICS FROM DIFFERENT SOURCES. THE
EVALUATION OF THE ACCURACY OF SURVEY RESULTS

Introduction

1. The practice of the Department of Statistics is to maintain the statistical standards of its surveys as high as possible in terms of what is practicable, given the managerial and statistical resources available in the Department. It is recognised that one aspect to the maintenance of standards is the direct measurement of known or possible errors in statistical surveys and other exercises where a methodology is available for achieving this and the resources are available. It does seem however, that this approach to the maintenance of standards is currently hindered by the lack of a generally acceptable theoretical framework which can serve as a basis for assessing total survey errors. Furthermore, where methods for assessing errors are available for particular surveys they are frequently of a nature as to require relatively large resources.

General Comment of the Need for Error Evaluation

2. Quantification of survey errors, of course, provides an explicit indication of one aspect of survey standards and yields a valuable control over such standards. Every survey is subject to some form of error and the key factor in the approach to error control is not to attempt the impossible by trying to remove all errors, but to keep the errors down to a level where their size means that there is no significant risk of incorrect conclusions or decisions being made on the basis of the statistics. This approach is well understood in regard to the control of possible sampling errors in sample surveys, but is perhaps not as well appreciated in the field of non-sampling errors.

3. With the usual scarcity of resources available to a central statistical office in relation to its work programme the maximisation of statistical survey standards requires measurement of different types of survey error if managerial and statistical effort in controlling errors is to be well directed. This provides the basic validation of the increasing emphasis being given to measurement of survey errors.

4. So far as users of statistics are concerned, it is sometimes suggested that publication of studies of survey errors assists in the more intelligent use of survey statistics. While the Department of Statistics generally supports a policy of publication of survey methodological studies, the view is held that there are risks in such a course. Most survey errors

studies are of a technical nature, but are usually not capable of full understanding except by a professional statistician. Normally estimates of survey errors are subject to qualifications and care needs to be taken in the publication of error studies to ensure that these qualifications are heeded. It is not unknown for error estimates to be taken as precise estimates, with the possibility that after the corresponding corrections have been applied to the survey results these are then regarded as error-free.

Current Departmental Position

5. Under the departmental reorganization shortly to be implemented provision is made for ad hoc organisational units with specific responsibility for departmental-wide control of various aspects of survey standards. This will allow for a more systematic, larger scale study of survey errors and their control. Up to now, the approach to this question has been under the organisational and resource constraints prevailing, less systematic. The balance of this paper makes brief reference to particular cases of survey error evaluation undertaken by the Department.

Assessment of Sampling Errors

6. The quantification of possible sampling errors at the usual probability confidence levels for the main estimates obtained from the regular sample surveys conducted by the Department is well established. Regular short-period surveys of retail sales and stocks levels, manufacturers' sales and stocks and building activity are examples where possible sampling errors in the main national and regional estimates have been assessed.

7. Since considerable user interest normally lies in the inter-period movements shown by these estimates, in some cases more interest than in the absolute specific-period estimates, confidence limits for these inter-period movements are also quantified. Decisions on the range of detailed estimates from sample surveys are made partly on the basis of the results of these assessments.

8. There are a few examples of long-standing surveys in the field of incomes statistics conducted by the Department based on outside administrative records where only quasi-probability selection methods can be used because of limitation of access. As a result valid assessments of error possibilities are not possible, but departmental practice is to employ probability sampling methods wherever possible. It is planned that when greater use of computer methods is made in the administrative processing of income tax returns that full probability sampling will be possible in the area of annual income surveys.

Assessment of Non-Sampling Errors in Sample Surveys

9. The usual frame employed for regular short-period sample surveys is derived from corresponding censuses conducted at five-yearly intervals. These censuses are not carried out using the identical data collection and processing methods as employed in the sample surveys, however. When subsequent censuses are held and checks are placed on aggregated sample survey statistics which match the time specification of the census significant discrepancies are usually found.

10. Since the census cannot be taken as the equivalent full-enumeration survey to the sample survey (differing only in respect to the sampling fraction) the discrepancies are influenced by a combination of sampling and non-sampling errors, including any variation in the nature of the response errors in the two types of survey. The correct theoretical handling of the problem of reconciling these sample survey and census discrepancies is a problem the Department has not yet solved to its satisfaction. There is an increasing awareness of the importance of the reporting errors made by respondents in sample surveys as more recent work suggests that these are at least as important as the sampling errors in some cases. For example, the recent need to publish revised statistics of consumer spending in retail outlets obtained from the regular sample survey arose through discrepancies with the 1968 Census of Distribution statistics. A major component of these discrepancies has been evaluated as non-sampling errors of the response-reporting type, the sample survey being carried out by mail. The isolation of response errors was carried out by matching studies and follow-up inquiries.

Checking Methods in National Income Statistics

11. New Zealand's national accounts are at present prepared using as basic data income returns to the Inland Revenue Department. The position will alter when the New Zealand adaptation of the United Nations System of National Accounts is introduced, with its wider range of supporting surveys, but currently it is not possible to prepare national accounts from the expenditure side, because of the lack of a household expenditure survey and a survey of private non-profit organisations serving households.

12. However a cross check on the regular national accounting aggregates based on the production approach is obtained in those years when inter-industry input-output studies are prepared. It is found that input-output statistics, being drawn mainly from an extensive examination of production statistics, rather than an analysis of net earnings, tend to give higher values in the case of values added, depreciation, capital formation and for the residual category of consumption expenditure by persons and non-profit organisations.

13. It is recognised that while, on the one hand, it is probable that the annual national income estimates for New Zealand understate somewhat the correct economic position on both sides of the account, it is possible also that due to some double-counting of production, particularly in such fields as manufacturers' distribution expenses and road transport, there has been an inflation of net output in the most recent input-output study.

14. No exact reconciliation between the latest input-output study and the corresponding official national income statistics has been possible due to the difficulty of isolating all the sources of the discrepancies. However, the cross-checks which have been possible have been most useful in directing attention to areas of the national accounts where estimation methods may be in need of improvement. With the implementation of the S.N.A., of course, the input-output study will be integrated more fully with the preparation of the annual national accounts and more explicit cross-checks and reconciliations will be possible. A prerequisite to this is the inauguration of a central register of economic enterprises and the conduct of integrated economic surveys as a result.

Checks on the Census of Population and Dwellings

15. With New Zealand having no land borders with other countries and, as a result, full administrative control being possible above population movements into and out of the country, statistics of acceptable accuracy are available on external migration. Together with good quality statistics of vital events derived as a by-product of the compulsory registration scheme, these statistics make it possible to check census of population counts of population against estimates based on the previous census and increase in population over the inter-censal period. These checks have normally indicated that any relative changes in under-enumeration between censuses have been very small. The close administrative control over the census operation that is possible in a country with a relatively small population and where the level of public acceptance of the census is good also ensures that the absolute standard of enumeration is high.

16. The application of the check method to age and sex statistics derived from the census of population has indicated some anomalies in regard to counts of young children, but otherwise, as in the case of the total population, the level of consistency is statistically satisfactory.

17. The wider use of dwelling statistics from the associated census of dwellings, especially statistics of unoccupied dwellings, led to the conduct of a post-enumeration check on

the enumeration of unoccupied dwellings at the 1966 Census of Population and Dwellings. This check was carried out using random area-unit sample. The results indicated at the national level that the quality of enumerations was good, but that in some resort areas census field staff had under-enumerated certain classes of lower standard dwellings (known as baches) because of the difficulty of determining whether in fact they were still in (possibly periodic) use.

The need for independent cross-checks of statistics
from different sources

A. Statistics from different sources: need for independent cross-checks

The main job of the Bureau of Statistics being collection, compilation and presentation of statistical data in a comprehensive and methodical manner, it often becomes necessary to bring together data on the same characteristic originating from various independent sources. This is more so in a developing country where the degree of reliability is highly variable between sources of information.

2. Agriculture is the predominant sector of the economy in most developing countries and in this sector, the method of data collection is often the least uniform. It is most important to examine the reliability of data in agriculture from independent sources. The major export crops like cotton, coffee, sisal etc., are marketed through the respective marketing Boards or authorities. Apart from these offices, independent estimates of production and trade are also available from reports of Ministry of Agriculture and Trade Reports of Customs and Excise Department. In case of food crops like maize, cassava, bananas etc., the scope of comparison is more limited due to a substantial proportion being used for subsistence consumption. Even in the case of crops like maize and paddy, whose marketing is controlled by the National Agricultural Products Board, the official figures of marketed quantities do not cover the total transactions, particularly in years of food shortage. The regular sources of information, in case of maize and paddy for instance, are the reports from N.A.P.B. and Ministry of Agriculture. Ad-hoc household budget surveys from time to time also help to supplement the information on sales and subsistence consumption of food crops. It is often observed that the data from various independent sources do not closely agree with one another, and the Bureau of Statistics does examine such data which originate from different sources for consistency before actually using them for various purposes.

3. Similarly, in Manufacturing sector the growth in the consumption of electricity for industrial establishments is used for purposes of cross-checks of the growth of manufacturing output and sale of petroleum products in Transport sector.

B. The Evaluation of Survey Results

The number of countrywide surveys that have been conducted are only a few. Most of the ad-hoc surveys were restricted to small areas, or covering limited aspects of economic activities, and not much has been done for evaluation of the accuracy of survey results. In the work of the Population Census, 1967, it had originally been proposed to have an extensive post-enumeration check, but due to various difficulties, mainly of finance and manpower, the post enumeration check was restricted only to limited areas. Some efforts are being made to check the accuracy of the recently completed Household Budget Survey, e.g. cross-checks between income and expenditure, comparison of household characteristics against those obtained during Population Census, etc.

Item 13 - United Kingdom

THE NEED FOR INDEPENDENT CROSS-CHECKS OF
STATISTICS FROM DIFFERENT SOURCES. THE
EVALUATION OF THE ACCURACY OF SURVEY RESULTSI CROSS-CHECKS IN NATIONAL ACCOUNTS

1. One of the main examples in the United Kingdom of the use of material from independent sources to arrive at the same aggregate is in the estimation of the gross domestic product. At current prices, this is estimated independently from expenditure data and from income data. The expenditure estimate is itself built up from a wide variety of sources, particularly the estimate for consumers' expenditure. The income estimate is based primarily on data derived in the course of collection of Inland Revenue taxes. This covers data on wages and salaries, the profits of unincorporated businesses and the profits of companies. Some adjustment has to be made in each case to cover income outside the tax field and to adjust to national accounting concepts. The income of farmers is estimated separately, and the profits of public enterprises are taken from their own accounting data.
2. On an annual basis, the alternative estimates normally differ by less than one per cent of the total, and the sign of the difference varies from one year to another, so that over a period agreement between the two methods is even closer. Data by both methods are also produced quarterly. Here the differences in each quarter are frequently of about the same size as the annual discrepancies and are therefore relatively much larger. But there are inevitable differences in timing between data derived by the two approaches, which tend to cancel out and disappear over longer periods.
3. For the measurement of growth in real terms, the expenditure data are also expressed in constant prices. The deflation is carried out in as much detail as possible, partly by the use of price indices and partly by re-valuation at base-year prices of physical quantities. The implied overall deflator derived from comparison of the current and constant price expenditure figures is applied to total incomes, so that a constant price figure from this source is also available. A third and largely independent estimate of gross domestic product at constant prices is also available in index form. This is obtained by applying appropriate indicators of changes in output to the base year estimate of net output for each main industry group. The index of industrial production is used for the industries which it covers, namely, mining, manufacturing, constructions, and gas, water and electricity, and similar methods are applied to the various distributive and service industries, including public administration and defence. For agriculture, double-deflation of inputs and outputs is used. The year to year movements of this index of GDP at constant prices agree reasonably well with the alternative estimates from expenditure and income data, though here again the discrepancies tend to be relatively larger for quarterly figures than for annual ones.
4. Having obtained different estimates of GDP both at current and constant prices, the problem still remains of how they should be handled. One method which has been advocated but has not been adopted is that the income and expenditure estimates at current prices should be "forced" into agreement by some form of arbitrary or pro rata adjustment of the components

of one series or the other so that they produced the same total. Another possibility would be to average the two methods to produce a single best estimate of gross domestic product, showing half the discrepancy on either side. This is the method used in the Canadian accounts. The method adopted in the United Kingdom has been to show the whole of the discrepancy, termed the "residual error" as a positive or negative item on the income side. It has been explained that this is not because the income estimate is considered less reliable than that from expenditure but because this treatment makes it easier to present more detailed analyses of the various components of expenditure without the complication of having to show the residual error in each table. At constant prices, the practice has been to show each of the three estimates in index form, together with the average of the three as the best indicator that can be given of the rate of growth.

5. Although the presentation of two slightly different estimates at current prices and of three at constant prices can be embarrassing at times, and a nuisance in that it is not possible to give an unequivocal answer to questions about the level or rate of growth of GDP, this course is, nevertheless, considered preferable to publishing only a single estimate based on averaging or adjustment of the discrepant items. It is a constant reminder to users of the figures of the errors of estimation that are unavoidable in such a wide and complex field as national accounting, and at the same time it acts as a spur to the statistician to look at the figures with a critical eye and endeavour always to improve them so as to minimise the differences.

II CHECKS ON EXTERNAL TRADE FIGURES

Introduction

6. The notes below are concerned with the problems associated with the independent information available for an independent check on United Kingdom trade statistics and describes the checks that have been made on them.

Documentation Procedures

7. It will be useful to describe how the basic data are obtained for United Kingdom trade statistics. For the purpose of recording exports, traders are required to lodge statistical documents with HM Customs for all goods leaving the country, apart from a few exceptions, such as personal effects or goods of no commercial value. These documents require, inter alia, a description of the goods, their f.o.b. value and usually a record of quantity. For some categories of goods the documents have to be lodged before the goods leave, but for most a period of grace (six days after shipment) is allowed for the documents to be entered. A possible source of error in the collection of basic data for exports lies in the entry of incorrect values. But any failure by traders to lodge documents after the goods are shipped provides an additional source of error. The procedure for recording imports is different, since an importer must submit documents before goods can be cleared through Customs. Moreover, the import documents also record tariff details, hence they are particularly scrutinised by Customs Officers because of the revenue interest.

Independent Information

8. Whenever a ship or aircraft enters or leaves the country, the shipping company or airline is required to lodge with Customs a list of the

goods carried (the manifest). The manifest contains a broad description of the goods carried with some indication of weight, but no values are quoted. This information from the manifests cannot provide an independent control total for the value of goods traded; but it can, by a matching of the statistical documents against manifests, be used to check whether all the statistical documents necessary to record the movement of goods in and out of the country have been received.

Earlier Manifest Checks On Exports

9. Periodic checks of the export documents submitted against the goods described on the manifests have always been made by Customs. Before the war, about half of the manifests were checked, but after the war, the proportion examined declined to about one per cent. This reduction partly reflects the large increase in the total number of documents, since though the total checked fell because of shortage of resources, nevertheless Customs matched over 100,000 documents each year up to 1966. These earlier checks were of a fairly simple nature; however, they pointed to two possible sources of error. First, the failure to lodge a small proportion of the documents, implying a possible understatement in recorded exports; and secondly a compensating error due to duplicate documents being submitted for some goods. No firm conclusions could be reached from these checks. However, the possibility that exports were not comprehensively recorded was suggested by the identification of certain specific shipments for which declarations were found not to have been lodged. Therefore, a carefully planned investigation was instituted of the recording of exports in 1968.

The Manifest Checks On Exports In 1968

10. This inquiry is described in detail in the Board of Trade Journal of 10 September 1969, but a brief description is included here. The object of the inquiry was to identify any error in recording, and to measure its extent if it existed. A sample of sailings (or flights) for each port in the country was chosen, the items on the manifests were compared with the export documents received, and the reason for any discrepancies investigated. The exercise involved correspondence, often protracted, with traders to explain the apparent failure to lodge documents and to determine why some export consignments were apparently documented in duplicate. In the investigation, over 4,000 sailings, flights and crossings of the land boundary between Northern Ireland and the Irish Republic, involving nearly 100,000 manifest items were examined. It led to over 11,000 queries with traders. Some of the items listed on the manifests for which no document could be traced related to personal luggage etc. for which no document is required; but in the majority of cases queried either a document was supplied or a claim made that the document had been submitted. Claims that documents had been lodged were investigated further with the help of the additional information provided; in particular, checks were made against computer listings of all items included in the trade statistics up to two or three months after the date of sailing in case the documents had been submitted late. The net effect of these omissions and errors identified was an estimated under-recording of documents of between 3 and 4 per cent.

Over-valuation in Recorded Exports

11. Another aspect of errors in recording the basic data concerned the values declared on export documents which are required to be on an f.o.b. basis. Information available from material assembled for the purposes of exchange control was examined for shipments made for the most part in 1968. It was found that about one quarter of the documents were valued above fob (commonly cif) and about one tenth below (commonly ex-works); and the net effect was an estimated one per cent over-valuation.

Complete Manifest Check

12. The discovery of these errors in the recording of exports led to the introduction in August 1969 of a comprehensive checking of all manifests and a continuing sample check on the values declared on export documents. The complete check of manifests is a big clerical operation involving the comparison of about half a million documents each month with the manifests for all ships and aircraft departing. This is made possible because the statistical documents include not only the marks and numbers of the goods but also the date of departure and the name of the ship or flight number of the aircraft on which the goods depart. Hence by sorting the documents as they are received to bring together all those relating to a particular ship or aircraft it is feasible to match the goods declared on the documents with the items shown on the manifests. The matching operation is carried out several weeks after the departure of the goods to allow time for the documents to reach the Statistical Department of Customs, but there are often difficulties in performing the matching because of inadequate information on either the manifests or documents. Sometimes the goods depart on a ship different from the one stated on the export document and occasionally, particularly with ferry traffic, the goods leave on the ship stated but either earlier or later than described. In addition, the description of some goods in a ships manifest are incompatible with the same goods described on the documents.

13. There are also a proportion of documents which are lodged belatedly in the normal course of events, so that after the matching is finished some late documents are received. When therefore replies to the queries raised with the traders claim that the documents have been lodged, the verification of the claims requires further searches to be made amongst late documents, as well as among those received earlier but which were not matched at the first check.

14. Analysis of the information relating to the early period of checking confirmed the estimate of the level of under-recording derived from the sample investigation carried out in 1968-69. The results also showed that the Customs checking operations had succeeded in securing a marked improvement in the lodgement of documents. (At the time of drafting these notes, final results of the checks were not available and evaluation of the checking procedures was continuing).

Valuation Checks

15. An independent check on the values declared on a sample of export documents was initiated in September 1969 by HM Customs. In this exercise, requests for documentary evidence, such as the invoices, were sent to the traders responsible for the lodgement of a small random sample of export documents. This evidence was then compared with the value on

the documents to ascertain whether or not the correct f.o.b. value had been entered. The results of the first checks indicated that the previous estimate of a one per cent over-valuation was substantially correct. (At the time of drafting, the checks had improved the accuracy of the values declared and appeared to have removed the tendency to over-valuation; the checking was continuing).

Other Sources of Data

16. It is possible to make limited use of information about certain aspects of trade available from other sources - for example, the Ministry of Technology collect information about export deliveries from manufacturers. It was a query from the British Iron and Steel Federation about a change in the relationship between exports recorded in the trade accounts and weights of export shipments of steel reported by producers that was among the factors leading to the 1968 checks described above. There are however numerous difficulties, of timing, coverage, valuation etc. in reconciling figures derived from separate sets of records, although such comparisons are useful if they can indicate areas where further investigation may be fruitful. Even more severe difficulties arise when comparisons are attempted between trade recorded by two different countries; such difficulties are well known to statisticians and in general it does not seem profitable to investigate them closely, although there are instances where detailed investigations of discrepancies have led to discovery of errors of recording.

17. In the ordinary way systematic checks are not made on United Kingdom trade by comparing it with figures collected for different purposes, since the differences in the way two sets of figures are compiled may lead to greater discrepancies than those due to possible errors in the trade figures; some cases of special interest have been examined (e.g. exports of cars) and have confirmed this conclusion.

III COMPARISON BETWEEN DIFFERENT SOURCES OF EMPLOYMENT STATISTICS

18. In the United Kingdom, employment statistics are obtained not only from the censuses of population but also from the administration of the national insurance scheme. Under this scheme, it is compulsory for every employee to be insured against the risk of industrial injury. The great majority are also insured against the risk of unemployment, but certain groups (such as married women) can opt out of this part of the scheme. All, however, should be insured against industrial injury.

19. The contributions to the national insurance scheme are paid by purchasing stamps which are then affixed to cards, known as national insurance cards. These cards have to be surrendered once a year and exchanged for new ones. The exchanges of cards are carried out in a quarterly cycle. By counting the cards as they are exchanged, it is possible to obtain quarterly estimates of the total number of employees. By subtracting from this total the number of employees who are registered as unemployed at the local offices of the Department of Employment and Productivity, an estimate is obtained of the number of employees who are in employment. In the paragraphs which follow the estimates, made in this way by the Department of Employment and Productivity, are described as the "DEP estimates".

20. The Department of Employment and Productivity estimates of the number of employees in employment differ from the corresponding estimates obtained in the censuses of population. Part of this difference can be accounted for by some known deficiencies in the census estimates.

21. The data about economic activity in the 1961 census was obtained from a 10 per cent sample of all households in the census and was found to suffer from a slight bias. A post-enumeration survey by the General Register Office also suggested that there were about 230,000 females who had jobs but failed to declare them in the census.

22. The 1966 Sample Census was based on a 10 per cent sample of households. Owing to defects in the list from which this sample was selected, the estimates of the population of Great Britain based on the sample are believed to have been too low by about $1\frac{1}{2}$ per cent, though the under-enumeration was probably slightly more than this for economically active males and slightly less for economically active females.

23. The following table shows how the Census estimates, after adjustment for these known deficiencies, compare with the corresponding Department of Employment and Productivity estimates of civilian employees in employment, found by interpolation between the published March and June estimates:

		<u>Millions</u>	
	1951	1961	1966
<u>Males</u>			
Census	13.4	13.9	14.4
DEP	13.5	14.3	14.7
<u>Females</u>			
Census	6.5	7.4	8.3
DEP	7.0	8.0	8.6

24. It will be seen that in 1966 the Department of Employment and Productivity estimates of civilian employees in employment exceeded the adjusted Census estimates by about 300,000 males and 300,000 females. These are the net differences between the sizes of those groups of persons who are included in the Department of Employment and Productivity figures but not in the Censuses, and vice versa. The most important of these groups are as follows:

a. Full time students and school pupils who work in their free time. These are counted as employees in the Department of Employment and Productivity estimates if they exchange national insurance cards, but are counted as students (i.e. not as employees) in the census.

b. Persons who were out of work due to sickness in the week before the census. These would not be registered as unemployed and if they exchanged national insurance cards they would be counted as employees in employment in the Department of Employment and Productivity statistics. In the census they would be recorded as "out of work, sick".

c. Certain seasonal workers, and others who work at regular intervals, who exchange national insurance cards but may not have been at work in the particular week before the census.

d. Certain persons may describe themselves as employees in the census but may be classified differently in the national insurance scheme. For example, some persons have two jobs, one as an employee and the other as a self-employed person. Persons who have been sick for over a certain period may not be classified as employees for national insurance purposes but may still describe themselves as such in the census.

25. In view of the importance of reconciling the Department of Employment and Productivity and census estimates, it was decided to incorporate some special questions in the 1966 census of population and to design some special post-enumeration surveys. It was also arranged that the census in April 1966 and the regular Department of Employment and Productivity count of persons registered as unemployed in that month should be held on the same day.

26. In order to find out about the seasonal and irregular workers, persons in the census were asked whether they had a job at any time during the 12 months before the census; whether they had a job during the week before the census; and whether they had a job on the day of the census. People were also asked to state whether on census day they were registered as unemployed; seeking work but not registered as unemployed; unable to seek work because of temporary sickness or injury, waiting to take up a job which would start after the census day; and whether they were wholly retired or not seeking work for other reasons.

27. A post-enumeration survey, consisting of interviews with 5,000 respondents to check the accuracy of their replies, was carried out by the Government Social Survey Department. In a second post enumeration survey the 5,000 members of the sample were traced in the national insurance records and details of their contributions for a 3 year period were ascertained. From these records it was possible to assess whether or not each individual had exchanged a national insurance card during the appropriate period, and thus to estimate the proportions in the sample who had been classified differently in the census and Department of Employment and Productivity estimates.

28. In this way it was possible to make rough estimates of the numbers of persons in the categories mentioned in paragraph 7:

a. There were about 170,000 males and 130,000 females who were classified as employees in the Department of Employment and Productivity estimates but as students in the census.

b. About 130,000 males were classified as employees in employment in the Department of Employment and Productivity estimates but as "out of work" in the census. Of these about 60,000 were "out of work, sick" and 70,000 were out of work for other reasons. About 40 per cent of the latter were waiting to take up a definite offer of fresh employment.

c. There were also some irregular workers, perhaps including 50,000 or so males, who had jobs in the year before the census, and almost certainly exchanged national insurance cards in the years both before and after the census, but who were economically inactive in the particular week before the census.

d. Offsetting the above there were about 100,000 males who were counted as employees in employment in the census but not in the Department of Employment and Productivity statistics.

29. The above-mentioned groups explain 250,000 of the difference between the census and Department of Employment and Productivity estimates for males. In the case of females the position is less clear. There seem to have been more than enough seasonal and irregular workers to account for the difference between the census and Department of Employment and Productivity estimates. Against these there must be offset certain family workers, e.g. wives working for their husbands, who are not obliged to hold national insurance cards but who may have been described as 'in employment' in the census. There may also have been some females, working for less than eight hours per week, who were not insured for industrial injuries and so did not have cards.

30. The special census questions mentioned in paragraph 9 showed 221,000 males and 76,000 females who stated that they were registered as unemployed. The actual numbers on the register on the census day were 234,000 males and 65,000 females who were wholly unemployed, together with 7,000 males and 9,000 females who were disabled persons registered for sheltered employment only or persons not claiming unemployment benefit who were seeking part-time employment only.

IV EVALUATING THE ACCURACY OF THE POPULATION CENSUS OF GREAT BRITAIN: PRE AND POST ENUMERATION SURVEY

31. A programme to evaluate accuracy is an accepted part of a large and complex population Census of the type conducted in Great Britain. Such a programme has a dual function of determining the reliability of the data collected and of evaluating the methods of each full Census or test Census as a guide to future improvements. There are a number of cross checks on each set of Census data collected, but survey techniques are able to come closest to the causes of inaccuracy.

32. Self enumeration is the firmly established method of the population Census of Great Britain, used since 1841, and public participation has been legally enforced since the first Census in 1801. Census tests, however, are not backed by legal sanctions. Inevitably there has been a growth in the information required from the public, and successive Censuses have had more questions. Increasing complexity in the organisation of enumeration, and greater demands on the temporary Census field force have paralleled this growth. There is increasing scope for inaccuracy. Now that many Census customers are equipped with automatic data processing systems for the sophisticated approaches to information required by innovations in public administration, the accelerating growth in demand for more and better data may threaten the ability of a Census to collect accurate information. The point of collection is the weakest link in the whole system. The next few years will be a critical phase in determining whether established information gathering techniques and information systems in general can be developed to meet expanding demands.

33. The factual type of information gathered in the Census is commonly more defective than realised, and whilst the precision of any sampling employed can be reasonably determined, there is a basic need to understand the process of obtaining response to a form and a need to provide a scientific framework for this understanding. Use of the term "accuracy" implies that absolutely correct answers to Census questions can be found after a reasonable expenditure of effort. With the exception of a few very straightforward attributes, or conclusively documented facts like the date of birth of a person, the concept of identifiable truth has to be rejected. A reasonable substitute is to approach the same item by more rigorous alternative measurement, concentrating information gathering resources on a sample of the population.

34. Lack of accuracy in Census data is attributable to the variability of sampling, where employed, to "non-sampling" effects arising during collection, and to errors occurring in processing. Quality control of processing within the Census Office is outside the scope of this paper, and whilst survey checks determine whether a field sampling method has been correctly applied, questions of the confidence which can be placed in sample statistics are also outside the scope of this paper. What remains, the "non-sampling variability", is the most difficult to measure.

35. In discussing the non-sampling effects, it is convenient to distinguish "coverage" errors and "response" errors. Each necessitates distinct methods of cross-checking. The prime objective is to count heads and omission of persons by the enumerator or form-filler is an unrectifiable defect, but, even when persons are counted correctly, errors in response result from innumerable combinations of factors inherent in Census taking. Respondents may fail to understand the questions; respondents may be negatively motivated; respondents' memory may be faulty; or they may lack knowledge. Enumerators in their interaction with respondents and form are potential sources of error. The objectives of the evaluation programme in the British Census are thus to obtain knowledge of these errors, and to understand the causes of error, so that inaccuracy can eventually be reduced and, even if errors persist, bias can be eliminated from the statistics.

36. An optimum programme would mix long term research on methodology with short term 'ad hoc' studies aimed at specific problems. It would have two phases. Firstly the Census, or parts of it, would be simulated in a linked series of tests, in which alternative methods are evaluated for accuracy. Secondly, the Census proper would be monitored, with alternative approaches to the measurement of the data. Improving the accuracy of the Census seems to be an iterative process, perhaps akin to the development of some of the repeating sample surveys. Reality is somewhat different. The multiplying purposes of the Census, the relatively slow cycle of planning and completing the massive job, the context of rapid changes in policies and demands for statistics and of technical ability to deal with information, combine to make a cohesive overall strategy difficult to define.

37. Each component, for example, the wording of a question, which hypothetically contributes to the accuracy of the Census, is influenced by the other components, for example, a tendency for response to be to "key" words rather than a whole question. We do not have either a complete knowledge of the components or the relationships between them. The evaluation programme is therefore governed by the general action that improvements in any part of the system can be reasonably expected to improve the accuracy of the Census.

38. The pre or post Census survey is one of a number of ways of checking on the variation between the information collected and what is usually a theoretical "truth". At each enumeration there is a need to know the extent of error in the particular set of data and in what circumstances the error amounts to bias. Also at each enumeration there is a need to establish the cause of errors in response or coverage, as the basis of action for future improvements. Field survey, amongst the methods of cross checking, comes closest to establishing causes of error.

39. Although there was a precursor in 1951, pre and post enumeration survey has been effectively operative in Great Britain only over the past ten years. The scale of survey operations has grown rapidly since the first follow-up was undertaken in 1951. The 1951 survey was not planned until after the Census and dealt solely with new questions on "household arrangements", and it covered a mere 300 households. The first survey to be planned in advance of enumeration took place after the 1961 Census. It aimed to ascertain the completeness of coverage and the quality of response to Census questions. The survey field force was recruited from the temporary Census field force and thus did not consist of 'professional' interviewers. Six thousand interviews were conducted in households in the course of the check. The post enumeration survey for the 1966 10 per cent sample Census followed the two objectives set in 1961, coverage and quality, though there were special problems of the completeness of the sample frame. To improve the standard of the quality check a sample survey of 5,000 addresses was made by Government Social Survey interviewers on behalf of the General Register Office.

40. In the 1951, 1961 and 1966 Censuses field testing prior to enumeration was very limited in scope. But for the 1971 Census two large scale multi-purpose test Censuses, involving 120,000 persons in April 1968, and 530,000 persons in April 1969, evaluated the feasibility of proposals for 1971. Coverage and quality checks were vital parts of each test. In the 1968 quality check form-fillers were interviewed in the test areas, but in 1969, to ensure greater representativeness, a national sample of households outside the test areas were enumerated specifically for the purpose of the follow-up interview.

41. The small scale follow-up survey after the 1951 Census dealt solely with the question which asked households whether they possessed or shared five basic amenities. First results of the Census showed serious incompatibilities within households and between households in the same dwelling. Three hundred households, recorded as entirely lacking in water closets, drawn from six urban areas where housing was known to have a high incidence of inadequate facilities, were visited by local registration officers and it was found that only two thirds of these did, in fact, entirely lack water closets. The result could not be treated as a national estimate, but from the evidence it was postulated that error was the result of the omission of water closets entered from outside a dwelling, and subsequent Censuses have made specific reference to this category. The basic strategy of the 1961 statistical assessment, evaluating coverage and quality, has set the pattern for subsequent surveys and, where there were limitations, the need for the deployment of more specialised and extensive resources in the investigation of accuracy was indicated, if conclusive results were to be obtained.

42. The design of the 1961 post enumeration survey was a multi-stage sample of households, the households used for the quality check being a

sub-sample of the sample used for the coverage check. The size of the sample was determined by the need to give a statistically significant indication whether or not under-enumeration had exceeded more than 5 per 1,000 persons. A sample of 3,200 households was needed. Since the Census was the only sampling frame available, an area sample was necessary. The area sample was geographically representative, but there was no stratification to obtain more reliable information about the small proportion of the population where error was high. Thus, although the quantity of error was measured, the survey was less informative about the characteristics of error represented by only a few cases from the extreme of the clustered distribution.

43. In the first stage of the 1961 scheme a sample of 2,500 Enumeration Districts [the group of households allocated to one enumerator] were selected with probability proportionate to the expected number of households. The second stage selected a plot at random from a rough grid of 10 plots within each Enumeration District. Boundaries of the plots were adjusted to conform to observable features and further adjustment took place in the field to bring the number of households to about 20. In these plots enumerators worked very carefully over the "ground" counting the numbers of dwellings, households and persons, for comparison with the coverage of the original enumeration. [Some 146,692 people were counted in the coverage check and the final estimate of the net error in enumeration was -0.2 persons per thousand.]

44. For the interviews in households, which were expected to detect significantly more errors than the coverage check, an average of 3 households per plot were selected. Only a 10 per cent sample of households had been required to complete a full Census form, and all households in a plot answering the full form were included in the quality check. The other 90 per cent of households had completed short forms and one of those households in each plot was randomly chosen for interview. It was thus possible to compare the response to the questions common to both Census forms.

45. Enumerators in each area were recruited for the job of interviewing and making the coverage check in one plot each. The enumerators were given a brief training and full written instructions, but were not 'professional' interviewers in any sense. Interviewing was completed in the period one to three weeks after the Census, minimising memory problems and the effect of changing situations. Co-operation by the public in the checks was voluntary, but, coming so closely after the Census, response was very good. The interviewers used a questionnaire which, with different questions, asked for the same information as required by the Census form. The questionnaire contained a copy of the information previously given on the Census form and where there was a discrepancy an attempt was made to ascertain whether this was due to faulty question design, to inherent difficulty in the question, or in real uncertainty about the situation in the household.

46. Participation in the 1966 Census was limited to a 10 per cent sample of addresses, thus there was a particular need to evaluate the accuracy of the sampling frames which had been built up from the 1961 Census and other records. The coverage checks were carried out by Census Officers [temporary officers in charge of a Census District containing a number of Enumeration Districts], but the operation of the quality check by interview was taken over by the professional interviewers of the Government Social Survey.

47. The 1966 coverage check was carried out within two weeks of the Census and was a multi-stage area sample which gave each Census Officer a plot of about 15 households for thorough examination, where buildings, dwellings, households and persons were listed. Since the size of Census Districts varied considerably, the sample design was not the most efficient for estimating under-enumeration, but a prime consideration at the time was to share work equally between Census Officers. The addresses of 20,600 dwellings returned from the 1,300 check plots were compared with the sampling frames for those plots. The indications were that deficiencies of the 1961 Census records as a sampling frame resulted in a 0.75 per cent under-enumeration of the population of England and Wales. Subject to the limitations of the sample, there were indications of more serious under-enumeration in a few areas, pointing again to the need for methods of evaluation to take into account the clustering of coverage and response errors into certain groups in the population and in certain geographical areas. It was evident that one of the most serious deficiencies in the 1961 records was the failure by 1961 enumerators to list vacant premises, thus any becoming occupied by 1966 had no chance of selection for the Census sample.

48. The 1966 quality check was based on about 5,000 interviews conducted in 300 Enumeration Districts spread over 100 Census Districts, a quantity and pattern dictated by the number of interviewers available. The interview checked enumerators' work in covering the persons and households at each sample address, and discussed response to the Census form itself. It took some time to get essential information to the interviewers to enable work to begin, and interviews in households enumerated in April did not begin until June. Thus interviewer-induced errors were reduced at the cost of some increase in time-lapse errors. Analysis has been compartmentalised into Census "topics", and two examples, the first from the "housing" section of the Census form on a question asking for the number of rooms, the second on a new question asking for educational qualifications, are typical of the pattern of follow-up survey by interview and of the relationships to other cross checks.

49. After the 1966 Census, Government Social Survey conducted a special small scale survey in Glasgow in view of continuing doubt about the accuracy of response to a question in the Scottish Census asking household to count its rooms. Between the 1961 and 1966 Censuses there had been a change in the definition of a "room", but, even allowing for these changes, comparison between 1961 and 1966 aggregate figures showed differences which could not be accounted for by normal intercensal changes. Checks between 1961 and 1966 Census records for the same dwellings showed changes in the number of rooms returned. Neither of these checks, however, established the "correct" values or the causes of the variation. Thus a special intensive interview survey was undertaken in Glasgow for a small sample of hereditaments. There was a particularly good chance of aiming at absolutely "true" information as independent cross-checks were available. Glasgow is an area in which the rating records list the number of rooms per hereditament. The sample was drawn from hereditaments which had been included in the 1961 Census, in the 1966 10 per cent Census, and in a 1965 housing survey which provided a further comparative base. The interview technique was to ask a straightforward question on the number of rooms of the type an enumerator would have asked, and then to account for all space in the dwelling, culminating in a sketch. Finally the interviewer attempted to reconcile any discrepancies and give information helpful to hypotheses of the cause of variation. The results of the survey, pointing to the cause of error rather than quantifying error, supported the change from the 1961 definition of a "room" which depended on usage of the room and thus was unstable between different households occupying the same accommodation.

50. For the first time in the Census, in 1966, a question was included which asked for individuals' higher educational qualifications. There were about 4,000 acceptable qualifications and about 5,000 unacceptable qualifications with a very indistinct separation, and the question in the form could be framed only with general guidelines. A heavy and difficult coding operation was anticipated. It seemed that qualified respondents would have little difficulty with the question, although a form-filler was often replying for another qualified person. The interview worked through an individual's history of exam taking, and results were obtained for about 250 individuals in the sample who would or should have counted as qualified, and a net undercount of qualified persons of 7 per cent detected. The small size of the cross check in this topic illustrates the problem of checking in a multi-purpose survey an attribute with limited incidence.

51. The cause of the undercount appeared to be the omission of qualifications which were "not in use" and the confusion caused in respondents' minds by the stressing of "degree level" qualifications in the question, which resulted in the omission of some vocational qualifications. Although the overall conclusion on this question as asked in 1966 was pessimistic, and there was doubt whether such a topic could be covered by a single question, a number of recommendations for improvement have been incorporated in this question in the 1971 Census form. The question will ask for the institution awarding the qualification to clear up coding ambiguities; vocational and "unused" qualifications will be stressed; and the question will cast a wider "net" by asking for qualifications at a lower level than those required for tabulation.

52. Two subsequent interview surveys, in the 1968 and 1969 Census tests, have been carried out and many of the results of these surveys have been used in formulating the 1971 Census questionnaire. Following the pioneering and sometimes inconclusive 1961 post enumeration survey the intensive techniques of post enumeration interview have been improved and are now more acceptable as yardsticks for measurements taken in the Census. No Census question is outside the range of interview techniques, giving scope for far wider checks than are possible by matching to documentary sources. With only the Census proper as a trial of methods it remained difficult to incorporate the specific findings of survey work into the forward planning of a Census which differed markedly from its predecessor. This was one of the facts taken into account in organising a programme of multi-purpose test Censuses for the planning of the 1971 Census. The programme attempted to simulate Census conditions in which alternative procedures could be evaluated. Measuring the accuracy of each trial techniques has been fundamental. There have been, however, a number of problems in obtaining predictive results from the test Censuses. To achieve a realistic test of field organisation, it was decided to concentrate the tests in a few geographical areas purposively selected for the problems offered to enumeration. Results may have been unrepresentative on this basis. Moreover each Census was held simultaneously in all test areas and resources for field monitoring were extremely stretched and observation of certain phenomena rather limited. Public participation in each test was voluntary, and even with publicity on the aims of the test, there was little motivation to complete any difficult parts of the form and response fell through non-contact and overt refusal, most seriously in the groups and areas with expected high error rates. Significant sources of inaccuracy were under-represented and error rates had to be regarded as minimum estimates, although the compensating positive influence of legal sanctions and national participation have an unknown counter effect.

53. In the 1969 test Census post enumeration survey was conducted by Government Social Survey in a national sample of households outside the test Census areas. The aim was to obtain more representative reactions to the Census form used in the test, and to avoid any adverse local influences generated in the test areas. The sample of households was enumerated in the normal way before being re-visited for an intensive interview in which the information entered on the Census form was checked for accuracy and discrepancies investigated. A multi-stage sample was used, with selection proportionate to population, and there were about 1,350 successful interviews with form-fillers. Recommendations formed on the basis of the survey results have been incorporated into the 1971 Census at certain critical points.

54. In brief, although it has been extremely difficult to isolate all components in the error detected in the 1969 test Census, and projections to 1971 have been made with caution, the size and complexity of the test Census forms was shown to have reached the point where public ability to return accurate information was diminishing. In turn the need for increased assistance and checking threw a heavier burden on the temporary Census staff which threatened the efficiency of their work. Finer methodological tests tended to be submerged in this general pattern but on the evidence of pre Census survey, the complexity of the 1971 Census forms has been reduced as far as possible with the expectation that this will lead to a more accurate response.

55. The 1968 and 1969 Census tests have emphasized the need for a full programme to evaluate the accuracy of the 1971 Census. Until recently the organisation of interview surveys was separated from the administration and processing of the Census. But in May 1970 the new Office of Population Censuses and Surveys was established, bringing together the Census and the Social Survey into a single organisation. This change is expected to give a considerable fillip to the development of pre-and post-Census surveys in Great Britain.

STATISTICS OF THE BUILDING AND CONSTRUCTION INDUSTRY

1. While the statistician must agree with the economist and the administrator that the building and construction industries are not only a massive but also a crucial variable in the economy, he will not be slow to express his doubts when asked to provide them with data about it. Indeed the statistician's almost automatic aversion to this area has been so long ingrained, that it is now appropriate to review his fears and troubles. As compared with a farm, factory or shop, the building firm is a much more elusive thing. Even the larger units can operate over a very wide geographical area with a minimum head office establishment. The low ratio of clerical and administrative people in the labour force, means that these firms have very few resources with which to compile basic data for industrial censuses. Furthermore, there is an emphasis in the industry to concentrate what accounting skills it does possess on budgeting. The fact that the commodity produced is non-standard means that the firm itself has little need to compile and retain detailed figures for each job it does. Only when the final result conclusively exceeds the budget figure will it start analysing this final result. That there is so little concern with final costs means that an analysis of materials which is accepted for manufacturing statistics is very hard to obtain from building firms.
2. One of the most obvious effects of the organisation of the building industry with its small administrative units, is that, in some instances, these may be very hard to locate for census purposes. Although some of the larger firms do have distinct administrative offices and systems, the lower end of the spectrum provides the typical firm. At this end of the scale the differentiation between the independent tradesman and the wage earner can become somewhat blurred. In New Zealand a single individual can change between the two several times in a year, even assuming the two roles simultaneously or at least on different days of the week. For such a man even the Inland Revenue has difficulty extracting satisfactory figures and lacking the traditional aura of authority of the fiscal department the official statistician is even worse off. A significant segment of this industry is composed of independent tradesmen who operate with a minimum of administrative organisation and this is one of the main troubles of the statistician.
3. The question that the foregoing immediately raises, is just how important is it to obtain cost analyses and other data from this section of the industry; in other words these difficulties must be put into perspective. In the case of New Zealand two immediate complications are apparent. Firstly, of all the industries, building and construction presents the most difficulty when calculating net output. Not only is this figure wanted for the output approach to national accounts, but it is also necessary as the basis of productivity studies. The need to obtain it is

aggravated by the fact that the ratio between net and gross output cannot be taken as constant. In no other industry do technological differences in both time and space play such havoc with this ratio. Such things as prefabrication in building and mechanisation in earthmoving make it essential to calculate net output as well as the gross figure.

Secondly, the lack of organised information on costs in this industry is very embarrassing when constructing input-output tables and similar models. It may be that certain of the service industries do have this data gap, but they are not big intermediate users of materials and fuels as is building and construction.

4. While the first set of problems associated with this industry arise from the difficulties of obtaining internal data such as labour force and an analysis of costs from the industry, the second set of problems could be said to be external. They have to do with defining the boundaries of the industry and the gross value of its output. It is true that the 1968 version of the ISIC widened the industry to include ancillary units of other enterprises engaged in construction. At this stage this piece of wording in the new ISIC is somewhat ambiguous, but wording alone is not responsible for the lack of definitiveness.

5. Quite large construction jobs are undertaken by government and local authority departments, both trading and non-trading as well as by private firms not primarily engaged in construction. In the case of trading concerns such as railways, electricity supply and property management, a large building and construction labour force is necessary to maintain their current operations. This force can however be switched and often expanded to undertake large capital works. New Zealand interprets the rather limited exposition of the new ISIC, "Construction Work for the Parent Enterprise" to include only those forces engaged in capital work, in this phrase construction is synonymous with capital. To separate current operations would not give a true picture of the parent enterprise and its activity.

6. No doubt such units are a significant portion of Major Division 5. However, the problem is whether or not the parent enterprises can and are willing to separate these activities. It is true that many private firms with construction staffs try to write off as much new work undertaken by them against current expenditures for taxation purposes. This presents problems to the statistician but they are not major. The most acute difficulties in this area are not here or in central government but in the local authority sector. Territorial authorities, that is boroughs and counties, in particular, are responsible for much building and construction work. Their accounting systems however allocate this to departments. Their councils are concerned with how much is spent on each activity rather than just how it is spent. Thus these accounting systems do not distinguish between work done by their own staff, and work done

by outside contractors. A very similar situation exists with the Ministry of Works who allocate expenditure to separate projects or separate districts. This is in fact what the local or the national politician wants, and this is what his salaried accountant gives him. The statistician is third in order of priority. It should be mentioned here that a special form designed for government departments and local authorities was used for the 1968-69 Census of Building and Construction and produced a better result for these bodies than was the case in 1963-64. As yet this improvement has not been fully analysed, but it is most probably due to the accompanying instructions emphasizing the need for estimating when exact accounting figures were not available.

7. If the total value of self construction by government and local authorities can be taken as the sum of costs, this is perhaps the easiest calculation of gross output for the industry. In the case of those jobs which are done on contract or other profit basis there are difficulties in the statistical recording of the gross value of work done in any period in particular the financial year. These difficulties are rendered more acute because of the importance of these statistics in the estimation of gross capital formation. This is due in the main to the paucity of accounting data mentioned earlier.

8. Until recently the main source of statistics on new buildings and alterations and additions to existing buildings in New Zealand was permits to build issued by local authorities. These give ex ante the cost of the building. The inaccuracies of this are obvious. Actual cost can differ from and indeed usually exceeds budgeted cost. This can only partly be explained by price rises and changes in design. The Department of Statistics has to rely on permit values which are not collected under the Statistics Act 1955. There is really no guarantee that they cover the whole cost of the job, or that coverage is even uniform from job to job. To remedy this the Department of Statistics instituted in 1965 the Building Activity Survey whereby the builders taking out all large permits and a random sample of small permits are asked at three-monthly intervals for their estimate of the value of work done during the quarter.

9. To some extent the Building Activity Survey is dependent on the permit system operated by local authorities. However, with all the drawbacks to building notification, this aspect of the industry is better recorded than the non-building aspect. Here the local bodies are not interested in anything like a permit system. True, most of this work, roads, aerodromes, bridges, dams and the like are constructed on behalf of government, central and local. For this major portion some figures can be taken from public accounts. However, this is really second hand data. It is not likely to be standardised between the various accounts and certainly not in a form compatible with production statistics.

10. The question then is not only how a census of this sphere of construction be conducted but also what questions

should be asked. There should of course be the usual questions on materials, employment and the like, but the main emphasis should be on measuring the real output, both gross and net of this industry. The fact that its products are the most heterogeneous imaginable makes for the main difficulties. It has been suggested that some sort of input measure, such as employment or machine-hours is the most promising solution. While this may be so the fact that the industry is capital intensive and is rapidly becoming more so, means that any one input will increase less than output during any period. At the present time the most appropriate questions for this part of the industry would be those relating to types and horsepower of machinery together with running hours. This would at least form the basis of output indexes.

This is still the problem, partially unanswered, of how to cover every unit in this field and how to ensure that the various works departments of government and local bodies present their accounting data in a form meaningful for the statistician and economist.

STATISTICS OF TOURISMA PROPOSED SYSTEM FOR COLLECTING MIGRATION
STATISTICS IN EAST AFRICAI. INTRODUCTION

1. For clarity, the system described below is generally outlined as it would apply to Kenya. The system would, it is felt, be equally applicable in Uganda and Tanzania - read "Uganda" or "Tanzania" for "Kenya" below. Comments applicable to an East African system are made where appropriate.
2. Put briefly, the system suggests the use of a single, double sheet, form. Tourists would fill in the top sheet on their arrival in Kenya and hand it in, and retain the bottom sheet to hand in on their departure. The details initially filled in on the top sheet would be automatically copied on to the bottom sheet by using carbon or pressure sensing paper. Kenya Residents would, whether arriving or departing, fill in only the top sheet, and hand in both sheets immediately. A rough outline of the suggested form is attached to this paper. (See Appendix)

II. DETAILED DESCRIPTION OF THE FORM

3. The form would be passport sized or smaller and composed of two sheets on different coloured paper. The back of the front sheet and front of the back sheet would be treated for "pressure-sensing" - so that writing on the front of the front sheet would be copied on the front of the back sheet. This would make it unnecessary for tourists to have to write basic information such as their name, passport number etc. twice. As a (messier) alternative to using pressure sensing paper, a carbon backed front sheet or even an interleaved carbon sheet could be used.
4. The fronts of the front and back sheets would be identical, providing space for official stamps etc., questions for the traveller to complete, a declaration by the traveller (possibly unnecessary) and coding boxes (if desired). The back of both the sheets would provide space for Kenya residents to fill in details of dependents travelling with them. The back of the back sheet would, have two date stamp boxes for recording travellers' entry/departure dates.
5. Questions Asked
As well as the usual questions asked for Immigration purposes, the form asks for:-
 - (a) Country of Residence - this replaces the previous question (on Kenya forms) "Country of Original Departure/Country of Final Destination". It is felt that these previous questions caused considerable confusion among travellers filling the forms and that the replies received were not always accurate. "Country of residence" is a much more clearly understood concept, and will probably correspond

in most cases to Country of original Departure or country of Final Destination. And it will enable new and important information to be collected on tourism in East Africa generated by East African residents. The information obtained from the "Country of Final Destination" question, on whether a traveller was leaving East Africa or not may be gathered more simply, if required, by asking an additional question, as at (d)

- (b) Number of Persons (including yourself) travelling on your passport - this is an "Immigration" question, and Kenya residents would be required to fill in names of dependents on the back of both the sheets. (But its inclusion on Tanzania and Uganda forms would ensure, comparability between all three countries statistics, since all travellers would be counted as in Kenya now, and not just all passports).
- (c) Principal reason for Entry/Departure - these 6 purpose categories, taken in conjunction with the Nationality and Residence questions, and whether a person is entering or leaving Kenya enable travellers to be allocated into the various Immigration Categories used in Kenya's migration statistics, (and enable follow-up of returning Kenyan University students).
- (d) Are you now leaving East Africa? - this would enable the "Visitors Departing Overseas" series of statistics to be continued, and, on an East African basis, give total visitors to East Africa from overseas. These could be compared with the visit statistics which counting Immigration forms gives, to arrive at an estimate of how many journeys between the three countries an average visitor makes. To obtain slightly more detail on this, the question might be phrased:

"Is your destination Uganda ; Tanzania
or elsewhere, outside East Africa, ?," "although
for most non-air travellers this will be self-evident from
the Date/Place stamp on the form.

- (e) Additional Questions could be added to the form. It should however be borne in mind that :
- (i) space is limited, and the fewer questions there are, the more clearly the form may be laid out;
 - (ii) part of the object of introducing the new form will be to speed tourist processing; the fewer questions there are, the quicker this can be. There are many questions it would be useful to ask of departing tourists especially, but in many cases the required information could be obtained more efficiently and in more detail by a sample survey rather than by asking everyone.

III. USE OF FORM AT IMMIGRATION POSTS

6. Procedure would be slightly different for Kenya Residents and non-Kenya Residents.

(a) Persons Arriving in Kenya

- (i) Kenya Residents ("Returning Residents") and persons intending to take up residence in Kenya would fill in the front sheet, and return the whole form to the Immigration Officer. The front sheet would be used for Immigration Entry Statistics, and the rear sheet for Immigration records.
- (ii) Non-Kenya residents would fill in the front sheet and hand the whole form to the Immigration Officer, who would date-stamp the rear of the back sheet, secure it in the visitor's passport, and retain the front sheet for Immigration Entry Statistics, (and Immigration records, possibly).

(b) Persons Leaving Kenya

- (i) Kenya Residents and anyone else not having a back sheet already in their passport (e.g. because it was lost, or because they had arrived in the country before the new system was introduced), would fill in the front sheet (including the question on leaving East Africa) and hand in the form. The front sheet would be for Immigration records and back sheet (with arrival date filled in on the back for non-Kenya residents who had lost their previous back sheet - this date will be in their passports) for Immigration Departure Statistics (and Immigration records, possibly).
- (ii) Non-Kenya Residents already having a back sheet in their passport would simply complete the question on leaving East Africa, sign the declaration and hand in the back sheet, which would be used for Immigration Departure Statistics (and Immigration records, possibly).

(c) Role of Immigration Officers

- (i) at Entry desks, Immigration Officers would have to determine whether travellers were Kenya residents or not and return the back sheet or not accordingly, having date stamped the front and rear sheets. If a serial number for forms was employed (see below) this would also be entered in the traveller's passport, together with the usual date stamp, permits etc.,
- (ii) at Departure desks; the Immigration Officer would simply detach the back sheets from the passports of travellers having back sheets, ensure that the leaving East Africa question was completed, and date stamp the "out" space on the rear of the back sheet, and the

travellers' passport. All other travellers would be required to fill in a new front sheet. For Kenya residents the "out" space on the rear of the back sheet would be stamped, and the whole form handed in. For Non-Kenya residents the front sheet would be discarded, to avoid confusion, the "Out" space would be stamped, the "In" space completed with the date of entry (as in his passport), the serial number of the new form amended to correspond with that of the previously lost rear sheet, as entered in his passport (if applicable) and the form handed in.

IV. USE OF SERIAL NUMBER FOR CONTROL

7. It would be possible for each form (front and back sheets) to be marked with a unique serial number. (A six element number AA0000 - ZZ9999 would provide nearly $7\frac{1}{2}$ million combinations, which could be increased by transposing letter and figure groups). This would give two advantages for control purposes:-

- (a) when forms were issued to border stations, a record could be kept of the blocks of numbers issued, and checked against the numbers of forms received at Immigration Headquarters to ensure that no piles of completed forms had been overlooked at the border station. Clearly this check could only operate for "front sheets", since "back sheets" might turn up at any exit point in the country.
- (b) It would eventually be possible for Immigration Headquarters to check that for every completed front sheet received, they also eventually had a completed back sheet, indicating that the visitor had left the country.

8. However,

- (a) check (a) could be performed almost as well without knowing the serial numbers of forms, by simply recording how many forms were despatched, how many returned and how many retained in stock,
- (b) check (b) might require disproportionate numbers of clerical staff for the gains achieved,
- (c) printing costs would be increased by the addition of serial numbers, and
- (d) the entering of serial numbers on passports would be time consuming for Immigration Officers. On balance therefore, this may not be worth adopting - but there is of course no reason why it should not be adopted by some and not all East African countries.

V. INFORMATION CODED / TABULATED

9. The following items of information from the form would be coded:

(a) Front Sheet (arrivals)

- (i) Entry Code (?)
- (ii) Date of Entry.
- (iii) Place of Entry (also gives mode of transport)
- (iv) Nationality.
- (v) Country of Residence.
- (vi) Males travelling on passport.
- (vii) Females travelling on passport.
- (viii) Purpose of Travel.

(b) Back Sheet (departures)

- (i) Departure Code (?)
- (ii) Date of Entry) enables length of stay to be computed
- (iii) Date of Departure) (days or nights, as agreed)
- (iv) Place of Departure. (Also gives mode of transport)
- (v) Nationality
- (vi) Country of Residence
- (vii) Males travelling on passport.
- (viii) Females travelling on passport.
- (ix) Purpose of travel.
- (x) Leaving East Africa?

10. It would also be possible to code passport number for each form, this enabling various control functions to be carried out, and the production of itinerary figures etc. on an East African basis. But this would increase coding/punching work by around 50 per cent, and would be probably uneconomical.

11. It could be a matter for individual countries to decide in how much detail to code certain items of information (e.g. purpose of travel, nationality, residence, place of entry/departure) but maximum gains from co-ordination would be achieved only if a certain minimum level of standardisation were imposed. It would be possible to incorporate identification of forms as originating from Kenya, Uganda or Tanzania in the Entry and Departure codes - say using 1, 2, and 3 as representing Kenya, Uganda and Tanzania entries respectively, and 7, 8, 9 as the respective departure codes. It would also be possible to combine the "place of entry/departure" and "Leaving East Africa" codes to ensure that impossible combinations (e.g. leaving East Africa from Uganda via Malaba) were ruled out.

12. Detailed work on codes used need not however be outlined here - it would clearly follow any general agreement on the system (s) to be adopted.

13. Data Processing, Even with standardised codes there would be scope for flexibility in data processing. Coding could be done on the forms themselves, or on coding sheets, and the information transferred on to punched cards or paper tape, to be processed by computer or counter/sorter. (Except that a counter/sorter would not cope with the computation of length of stay, and a manual computation of the coded information would therefore

be needed).

14. Tabulations As with work on codes, there is no need to go into details here. Clearly a large number of useful tabulations could be produced by cross-relating the information which it is suggested should be coded. Assuming that basic and comparable information on total entries and departures by point of entry/departure, purpose of visit etc. as tabulated by each country and made available to the others, it would be for each country to decide in what further detail other information was analysed.

VI. TECHNICAL POINTS.

15. For completeness, a number of technical points, as they would affect Kenya, are set out here, since they may also be relevant for Uganda and Tanzania.

- (a) The present Immigration date/place stamps are too large for use of the suggested smaller forms. They would have to be made smaller. It is suggested that the new stamps should retain their distinctive different shapes for entry and exit - oval and triangular, and that the "In" and "Out" boxes on the rear of the back sheet of the form should be similarly distinguished. Also the new stamps might incorporate the code number of the relevant entry/exit point.
- (b) Whether information is coded on the forms or on sheets, and card punched or paper tape punched will largely depend on whether it is permitted to move forms from the Immigration Headquarters to the Punch Room in the Treasury Building.

If it is permitted, then coded forms plus card punching will probably be the most efficient system; if not, there will be a choice between coding sheets, coded in H.Q., and card punched in Treasury, or coded forms, paper tape punched in H.Q.
- (c) It will be necessary to clear up technical printing problems for the forms - similar forms for Library purposes have however been produced in Kenya.
- (d) Computer programming work will have to be put in hand well before the new forms are introduced.
- (e) Immigration Officers will have to be briefed in advance on the operation of the system.
- (f) Outlying border posts, airline operators etc., would have to be supplied with forms in advance of the starting date.

VII. CONCLUSION

16. The system outlined in this paper has been under consideration for some months, and it is believed that it has considerable advantages over the existing Kenya system, not least in that it cuts down form-filling by half for tourist visitors while still enabling more useful information to be collected from them. It is however, still only a suggested system, and would be open to modification, the more so if such modification would ensure compatibility of Immigration data among the East African countries. It is intended that the system should go into operation in Kenya (and hopefully elsewhere) on 1st January, 1971. It is therefore essential that agreement on all substantive issues should be reached at an early date, to allow time for the many purely technical issues to be resolved by the end of the year.

OUTLINE OF SUGGESTED IMMIGRATION FORM
FRONT SHEET (Pink ?)

	Front		Back
(DATE/PLACE STAMP)	REPUBLIC OF KENYA IMMIGRATION DECLARATION FORM. to be completed in accordance with the Kenya Immigration Regulations. (Tick boxes <input type="checkbox"/> as appropriate)	R.No:..... Code:..... Sec:..... <div style="border: 1px solid black; padding: 2px; text-align:center;">Immig. Officer</div>	<u>KENYA RESIDENTS ONLY:</u> Dependents travelling on your passport:- (1) Mrs./Mr./Miss (2) Mr./Miss (3) Mr./Miss (4) Mr./Miss
SURNAME..... OTHER NAMES..... PASSPORT NUMBER DATE and PLACE ISSUED..... NATIONALITY..... COUNTRY OF RESIDENCE..... NUMBER OF PERSONS (Including yourself) TRAVELLING ON YOUR PASSPORT: MALE..... FEMALE..... (Kenya Residents enter details overleaf)			
CONTACT ADDRESS IN KENYA: (Arrivals only)			
PRINCIPAL REASON FOR ENTRY/DEPARTURE: Holiday Visit <input type="checkbox"/> Business Trip <input type="checkbox"/> In-transit Stop <input type="checkbox"/> University Study <input type="checkbox"/> Permanent Immigration/Emigration <input type="checkbox"/> Start/End Employment or School <input type="checkbox"/>			
ARE YOU NOW LEAVING EAST AFRICA? (Departures only) Yes <input type="checkbox"/> No <input type="checkbox"/>			

Front		Back				
BOTTOM SHEET (Blue ?)						
AS FRONT ABOVE		<u>KENYA RESIDENTS ONLY:</u> Dependents travelling on your passport:- (1) Mrs./Mr./Miss (2) Mr./Miss..... (3) Mr./Miss (4) Mr./Miss				
		<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th style="width:50%;">DATE IN</th> <th style="width:50%;">DATE OUT</th> </tr> </thead> <tbody> <tr> <td style="height: 40px;"></td> <td style="height: 40px;"></td> </tr> </tbody> </table>	DATE IN	DATE OUT		
DATE IN	DATE OUT					

935

14(b) - NEW ZEALAND

STATISTICS OF TOURISMVisitors

1. Basic data on overseas visitors comes from arrival and departure declarations completed by all 'longer staying' visitors (that is excepting crews and migrants). However those visitors arriving and departing on the same vessel or aircraft (principally ships' through and cruise passengers) are not at present required to make such declarations, and only numbers, port (sea or air) and arrival month are recorded. Data from arrival and departure declarations is analysed in two publications

- the Government Statistician's "Statistics of Population, Migration & Building"
- and the Tourist & Publicity Department's "New Zealand Visitor Industry Statistics" (copy accompanies)

Major deficiencies in the present data are

- no actual length of stay data according to purpose of visit
- no data on state/province of residence (country only is known)
- no data on visitors route to/from New Zealand
- no distinction between those 'tourists' visiting New Zealand on a paying holiday and those coming for a social purpose (e.g. wedding)

These deficiencies are remedied in new declarations which it is hoped to introduce after 1 April 1970.

Visitors' spending

2. Trading banks make statistical returns to the New Zealand Reserve Bank of all deposits recognised as coming from overseas visitors' spending: in the accounts of overseas exchange transactions these are classed as travel receipts (international fare payments go into a transport category). The sum is nowhere claimed to be an exact measure of all visitors' spending and because some actual expenditures are not banked is probably a significant under-statement.

3. Although some previous survey attempts at assessing visitor expenditure have been made, the first substantial effort

was made in 1968-9. The method used was personal interview, to a set questionnaire, of shipboard passengers, and of air-departing overseas visitors (in airport departure lounges just before flight departure). The airport section of the survey was of five whole weeks spaced to give whole-year coverage (some 90% of all 'long-stay' visitors - that is those not 'through or cruise passengers' - arrive and depart by air). Although the principal aim of the exercise was to get spending patterns of major groupings, amplification of the sample results will yield a measure of overseas visitors' expenditure. The survey was designed, commissioned and analysed by the Research Section, Tourist & Publicity Department.

Visitors' accommodation usage

4. A similarly operated survey is operating from November 1969 to October 1970 to determine patterns, by location and accommodation type, of overseas visitors' accommodation usage. As 'riders' to the main purpose there is coverage of visitors' main vacation activities and their judgments of 'critical' areas of hotels (reception, restaurant etc., tying their judgment to the named hotel 'which most suited their needs'). The survey is designed, commissioned and (will be) analysed by the Research Section, Tourist & Publicity Department.

Accommodation

5. Licensees of liquor-licensed accommodation make returns of room and bed occupancy to the Licensing Control Commission. From these returns the Tourist & Publicity Department compile a report on the accommodation including an inventory of acceptable standard (hotel and motel) accommodation. The first such report was for 1967-8.

6. In 1966-7 a sample survey of motel occupancy rates was carried out by Tourist & Publicity Department, in association with the N.Z. Motel Federation Inc. In 1969-70 the Federation, advised and aided by the Tourist & Publicity Department, is running another more extensive survey.

7. Attitudinal and Propensity to Travel Sample Surveys are carried out (Tourist & Publicity Department) in the major market areas, Australia and U.S.A., as the need is anticipated.

14(d) - INDIA

PROJECTIONS OF POPULATION FOR
NATIONAL AND SUB-NATIONAL
AREAS, INDIA

1. Introduction: Prior to the advent of the Five Year Plans projections of population was more or less an intellectual exercise of individual experts. In the official circles even as late as 1956-57 inter-censal estimates were made in the Ministry of Health or the Office of the Registrar General mainly for purposes of computing the vital rates for all India and the States based largely on simple mathematical methods like the arithmetic, geometric or the modified geometric method. It was only at the time of the preparation of the Second Five Year Plan that need was felt for projections of population not only for the next 5 years but for next 15 or 20 years so as to build up a plan perspective. In 1958 an Expert Committee on Population Projections under the chairmanship of the Registrar General, India, was set up by the Planning Commission to make a set of projections of population for all India and the States for adoption uniformly for all official purposes.

2. The Expert Committee was faced with a paucity of data which precluded the possibility of basing the future assumptions on the components of population growth on firm quantitative foundations. The vital registration statistics were very unreliable. The data on inter and intra-State migration were confined only to the place of birth data of the 1951 Census. Only the official Life-tables based on the various censuses provided a broad schedule of average inter-censal death rates which when related to the appropriate inter-censal growth rate, gave the associated birth rates. These Life Table rates indicated clearly a falling trend in mortality and a sustained high level of fertility. It was felt that the falling trend in mortality would also persist in the future, if the available international experience is any indication.

3. The Public Health programme of the First Five Year Plan and improvements in medicines indicated that mortality would decline considerably from the then currently high level of about 27 per thousand population. But with regard to fertility, which needs individual effort to arrest itself, the situation was nebulous. The family planning programmes of the first two Plans were rudimentary and it would have been unrealistic to expect them to have had any deep impact on fertility. The growth rate of 1.9 per cent and a birth rate of about 39 per thousand thrown up by the preliminary estimates of the 14th Round (1958-59) of the National Sample Survey which was for the first time based on a large sample, gave an indication of the trend of

population growth subsequent to the 1951 Census. Based on these considerations an assumption was made with regard to future trend in mortality and related with three alternative assumptions on future course of fertility to give three sets of population projections up to 1971 by the component method. These were respectively called 'high', 'medium' and 'low' projections. The set of medium projections was recommended for official use.

4. With regard to the States, no usable data were available regarding intercensal migration. The movement of population between Pakistan and India was considerable even after 1951 but no estimates of this flow could be made. In the circumstances the ratio method was utilised for obtaining the projections for the States. Realising the limitations of the ratio method viz., (i) the growth rate is assumed to be uniform for the region and the sub-regions, (ii) the cycles for pro-rata adjustment by rows and columns involve laborious calculations which distort the internal consistencies particularly when the sub-region is rather small in size and (iii) the arbitrariness involved in assuming that the rate of change in the proportion of State population to the national population would taper off to zero in thirty years, the projection was made only upto 1971 and confined to sex and the broad age groups of 0-14, 15-59 and 60+. The Union Territories by virtue of their being very small in size posed a serious problem. The projections for many of the Territories were open to severe criticism as not reflecting a balanced trend of population. With all these limitations, it should be mentioned that these projections for the first time provided a possible order of magnitude of the population of the respective sub-regions which could be adopted on a uniform basis by the planners at the Centre and the States/Union Territories. These projections were released in March, 1959, for use in the preparation of the Third Five Year Plan.

5. Revised all India projections - 1961: With the release of the 1961 Census provisional totals the projected figure for 1961 was found to fall short of the Census count by about 3 millions and a revision of the projections was considered desirable. There was, however, not much improvement in the limited data situation. The revision was, therefore, made at the national level only with the 1961 provisional totals as base and with slightly revised assumptions on fertility.

6. Revised projections - 1964: The 1961 Census final population totals as well as the age tables 1961 and Life Tables 1951-60 were released by 1963. By that time the vital rates computed from the data collected in a few rounds of the NSS were also available. The clinical approach to the family planning programme was reviewed and an extension-oriented programme of taking it to the door steps of the households was initiated with a substantially increased

outlay. This necessitated a further revision of the extant projections. The age distribution of the 1961 Census remained essentially the same as in the earlier Censuses. The NSS data showed constancy of the birth rate with a decrease in the death rate. The family planning programme also, if at all, could not be expected to have had an impact on the fertility before 1966. An examination of the available data on cause of death also showed that mortalities are still high enough to expect a steadily falling trend in mortality in the future also in the States. Revised quinquennial series of projections of population by age and sex for all India and States were brought out by April, 1964.

7. All India projections: Three sets of alternative projections were made upto 1981 for all India based on a unique assumption on future reductions in mortality and three alternative (High, Medium and Low) assumptions on future trends in fertility. Migration was ignored since international migration had so far been negligible and any likely immigration rate in future of persons of Indian origin from other countries could only be insignificant compared to the rate of national increase for the country as a whole. The medium set was recommended for official use. The various fertility and mortality assumptions made in 1964 for all India projections are given in Statement 1.

Statement 1: Alternative assumptions with regard to trends in general fertility rate (g.f.r.) and expectation of life at birth (e_0^o) during 1961-80

Projection	g.f.r. in 1961-66	Percentage decline in g.f.r. during			e_0^o during 1951-60 centred at 1956	Annual increase (in years) in e_0^o during	
		1966-70	1971-75	1975-80		1956-70	1971-80
Assumption I (High)	195	0	5	15	(M) 41.6 (F) 40.6	0.9	0.75
Assumption II (Medium)	195	5	10	20		0.9	0.75
Assumption III (Low)	195	10	15	25		0.9	0.75

The projected figures of population of the country in 1966, 1971, 1975 and 1981 obtained according to different assumptions are given in Statement 2.

Statement 2: Different sets of projections

Projection	Projected Population in millions			
	1966	1971	1976	1981
Assumption I(High)	495	564	644	725
Assumption II(Medium)	495	560	630	695
Assumption III(Low)	495	555	616	667

1961 Census count: 439 millions

In view of the fact that the Medium projections under Assumption II had been recommended for official use and that the projections of total population under the three alternative assumptions did not differ materially, particularly up to 1971, it was found expedient to work out the detailed classifications by age and sex for assumption II only. As a consequence the Statewise projections discussed in the next section were also worked out for the medium set (Assumption II) only.

8. Projections for States: With regional differentials in the growth rate being a proved fact, and with the availability of birth rates and mortality rates for the States associated with the official Life Tables, it was not found realistic enough to project the population by the ratio-method. It was therefore decided to use the component method for projecting the population of the States. This was found feasible because a preliminary study of the migration(place of birth) data thrown up by the 1961 Census showed that, but for the State of Assam whose population growth was likely to be affected appreciably by large scale in-migration from Pakistan(which could not also be assessed in any objective way), the overall effect of migration on State growth rates had so far been very small. Any reasonable guess about the future magnitude of migration being extremely difficult it was felt that the margin in the rate of natural increase assumed would take care of it as much as possible.

9. The States were first grouped into three categories according to the level of the e_0^0 during 1951-60. The projected annual increase in the e_0^0 was so fixed that States with relatively lower expectation of life at birth

14(d) - INDIA

at the base period have a higher rate of increase so that the Statewise mortality differentials would be wiped out in the future. Care was also taken to see that the average expectation of life at birth obtained from aggregating the State's experience would broadly be compatible with that obtained directly. As stated earlier migration with regard to States was ignored and the population of States were projected by the component method on the assumptions regarding future trends in fertility and mortality given in Statement 3. The birth rate and e_0^o calculated by the 1961 Census Actuary for individual States based on the Quasi-stable population approach were adopted to fix the corresponding base-line general fertility rates and e_0^o .

Statement 3: Assumptions with regard to trend in fertility (g.f.r.) and mortality (e_0^o) during 1961-81 for the States of India.

Group of States	1951-60		g.f.r.	Percentage decline in g.f.r.			Annual increase in e_0^o in years	
	e_0^o			1966-70	1971-75	1976-80	1956-70	1971-80
	males	females						
I. Kerala	49.5	47.1	173	5	10	20	0.8	0.65
Maharashtra	46.1	44.3	191					
Punjab	49.9	45.0	230					
Rajasthan	49.2	44.3	210					
W.Bengal	44.1	44.5	219					
II. Gujarat	40.8	39.2	220	5	10	20	0.9	0.75
Madhya Pradesh	41.2	40.0	203					
Tamil Nadu	40.7	38.8	149					
Mysore	41.2	39.2	193					
Orissa	40.7	41.1	186					
Uttar Pradesh	39.4	38.3	201					
III. Andhra Pradesh	37.8	36.0	179	5	10	20	1.0	0.85
Assam	31.6	37.0	258					
Bihar	37.4	37.8	203					

The figures of population projection for different States obtained on these assumptions are given in Statement 4.

Statement 4: Statewise Projections of Population

Group*	States	Population in thousands as on			
		1st March			
		1966	1971	1976	1981
I	Kerala	19137	21701	24422	26829
	Maharashtra	44931	51000	57471	63242
	Punjab	12906	15111	17438	19660
	Rajasthan	23257	26822	30653	34219
	West Bengal	39945	45801	52251	58325
II	Gujarat	23619	27093	30879	34368
	Madhya Pradesh	36624	41490	46807	51672
	Tamil Nadu	36640	40034	43563	46467
	Mysore	26463	29832	33508	36850
	Orissa	19587	21992	24571	26879
	Uttar Pradesh	82364	92378	103184	113087
III	Andhra Pradesh	39602	43920	48597	52789
	Assam	13711	15994	18633	21220
	Bihar	52063	58762	66283	73230

*The States were divided into three groups according to the level of e_0^0 for the base period 1951-60.

10. Projections for Union Territories: With regard to Union Territories which have individually small populations the use of the component method was found impracticable. Apart from the Census counts no other data were available. It was, therefore, found expedient to project their populations mathematically by analogy with an adjoining State. The rate of change of the growth rate for States adjoining the particular Union Territory in question obtained from the State projections already made was assumed to apply to the initial growth rate of that Union Territory. In the absence of requisite data on sex-age distribution the smoothed sex-age distribution obtained for 1961 were projected to future years in analogy with the adjoining States in the same way as the growth rates were projected for obtaining total populations.

11. Adjustment of marginal totals: As a result of independent sets of projections for all India and States and Union Territories there arose a problem of tallying the totals. Though the sex distribution of the total population for all India and States were tallied on a pro-rata basis, the age distribution was left unadjusted as the differences were mostly well within 1%.

12. Projection for individual ages, in the educational age-range 5-24: To meet the requirements of the educational planners, the projected population by individual ages for the age-range 5-24 was also computed by applying Sprague's Multipliers to the five yearly group totals. Use of the Sprague Multipliers had resulted in certain abnormalities in the single year population figures for some of the States and Union Territories. Specifically, (i) the single year age populations did not come out as expected in the descending order in some cases and (ii) the progression of the cohorts from age 'x' in one year to age 'x' + 5' in the next year of projection was not smooth in certain other cases. Such abnormalities were generally observed in the ages 5-9 particularly in Union Territories. As far as practicable some smoothing was done to tone down such disparities. An indication was, therefore, given that it was preferable to combine these data into appropriate age groups for planning purposes so as to minimise the abnormalities.

13. Rural-Urban projections: There has been a persistent demand from the Planners for the rural-urban break-up of the population projections. A sub-group set up by the Expert Committee went into the question exhaustively and came up with a rural-urban break-up upto 1981 by quinquennial years. The basic difficulty was that the definition of urban areas had undergone substantial changes over the years and there was reason to expect that even in their application in the

field a large amount of subjectivity had come into play. The projection on economic considerations was not also found feasible because of lack of reliable data on the likely trends in development of different areas. Though the place-of-birth data gave limited information on rural-urban migration, it was not found sufficient to set up reasonable assumptions on migration trends. On the other hand, it was found that the towns common to the four censuses during 1931-61 formed over 95% of the urban population in most cases. The proportion of the population of common towns to total population of the State for the 4 years 1931, 1941, 1951 and 1961 when plotted indicated an unmistakable rectilinear trend. A straight-line was, therefore, fitted and extended upto 1981. The proportions drawn from this fitted line were then adjusted proportionately to conform to the total urban population of 1961. The projection was done in terms of two 'size classes' and four 'functional classes'. Though six population size classes are available from the censuses, it was considered expedient to group them into two large classes, namely (i) towns with population 50,000 and over and (ii) towns with population below 50,000. The four functional classes were (i) manufacturing, (ii) trade and transport, (iii) services and (iv) agriculture. The Statewise projections for any particular year were aggregated to give the all India figures so as to avoid pro-rata adjustments.

14. Two specific situations, one involving the toning up of the projected trend based on past data and the other involving toning down of the trend are worth the mention here. In the case of Assam, Bihar and Orissa the 1961 proportions of urban to total population were very much below the corresponding all India level and the past trends in them were so static that projecting them beyond 1961 lead to situations contrary to the stated objectives of the Five Year Plans namely to obviate regional disparities in the country. This objective would imply among other things, promotion of greater mobility of labour between different parts of the country and launching of schemes of migration and settlement from more to less densely populated areas. Hence on the basis of a few trial calculations the projected proportions were inflated by 25% so that the adjusted proportions would increase by about one percentage point per quinquennium and that the average quinquennial growth rate would not be unduly large.

15. The rapid growth of the urban population of Delhi during the immediate past decade was toned down a bit on the assumption that though the tempo of urbanisation of Delhi would be kept up in the future years also, with the concomitant shrinking of rural Delhi there would still be some parts of Delhi which would preserve their rural character even by

1981. With regard to Union Territories, because of their small populations, their aggregated total was first projected as in the case of the States and then apportioned among the constituents according to the distribution of population in 1961.

16. Sex and age distribution of Rural/Urban population: The size of the rural male population was first calculated. In the absence of further firm data on trend in the sex pattern of movement of population from rural to urban areas State by State, realistic assumptions regarding future trends in urban sex-ratio became difficult. On the other hand, since the rural component is large, it was reasonable to assume that the trend in the rural sex-ratio would largely be similar to that of total population for which projection by sex was already available. Accordingly the proportion of rural males in the total rural population was projected for each State quinquennially upto 1981. For this purpose it was assumed that the ratio of the proportion of the rural males to rural total population and the proportion of all male to total population of State would remain the same as in 1961. The other components were then obtained by differencing.

17. The age distribution of the rural-urban population was obtained only for the broad age-groups 0-14, 15-34, 35-59 and 60+. The projections were made by a suitable application of the ratio method. The assumption made was that the disparity in the age composition of total and urban population in 1961 will be eliminated in about 100 years in order to allow for the changes in the urban age composition in comparison with that of the total during the period 1961-81. This was done separately for urban males and females and the corresponding rural figures were obtained by a simple process of subtraction.

18. Labour force projection for all India: Making use of the rural-urban population projections and the working force participation rate derived from the 1961 Census and making some quantitatively arbitrary assumptions, based largely on intelligent guesses, on the future trends in the age-sex pattern of participation rates quinquennial labour force projections upto 1981 were also calculated. In view of the progress of industrialisation and urbanisation, population redistribution and expansion of educational facilities visualised in the Five Year Plans it was found reasonable to assume that the labour force participation rates would, at least qualitatively, show a decreasing trend over the future years in the case of both rural males and females. In the urban areas the participation rates for males in the years 15-34 were expected to decline progressively since a large number of them might be reasonably expected to avail of the expansion in educational facilities. The

increasing tempo of economic development in conjunction with the upward trend in age at marriage could be expected to draw an increasing number of females in the folds of the urban labour force. Though these assumptions were qualitatively convincing, it was recognised that under the present circumstances of limited or no data, certain amount of arbitrariness was unavoidable while quantifying these assumptions for projection purposes.

19. Revision of projections: In 1967, a working group was established to review the extant projections in the light of the increased allocation for family planning purposes in the Fourth Plan, to interpolate annual estimates and to extend the projections upto 1986. The working group felt that with the family planning programmes not having any substantial impact so far, there was no likelihood of the projections being drastically changed at least upto 1971. It was, therefore, thought that revision of the projections was rather premature at present and might well be taken up after the 1971 results are released.

20. Annual estimates of population for 1951-81: The annual population totals were computed for all India by the application of Newton's divided difference formulae to the Census and projected population figures for the period 1951-1981. These all India estimates were then broken up by States and by age/sex by the Water's II Method.

21. Prospects for the future: The foregoing paragraphs show that a number of methods like (a) component method, (b) ratio method, (c) analogy method, (d) apportionment method, (e) extending the past trends into the future by curve fitting and (f) interpolation for obtaining individual year populations from the quinquennial projections have been used according to circumstances. In all these cases though reasonable qualitative assumptions could be made, more often than not quantification had unavoidably resulted in certain amount of arbitrariness and intelligent guesses.

22. With the release of the 1971 Census results, attempts will be made to review the existing projections and revise them in a big way. Though in projecting population below the level of the State limitations of data might still preclude the use of improved techniques of projection in the near future, it is expected that the projections at the national and State levels could be built up on a sounder basis. The 1971 Census will for the first time present data on migration based on 'place of last residence' and 'duration of residence at the current place of residence'. These data on inter-State migration will help in building up for the first time reliable life tables for the States and in framing suitable assumptions on future migration trends for application of the component

method of projection more rigorously at the State level. Even at the district level migration data could be built up under the special inter-censal programmes of tabulations. But the limitations of age data will exert a greater influence on the accuracy of the projections by sophisticated methods at the lower levels like a district or a town.

23. As has been usual in the past the Life Tables of the 1971 Census will no doubt be built up mainly on the basis of Census age data. However, a welcome addition will be the data on age-specific mortality rates for the various States from the Sample Registration Scheme of the Office of the Registrar-General for a series of years since 1966. The various rounds of the National Sample Survey could also provide schedules of age-specific mortality rates by the States. These data could provide valuable independent cross checks for the preparation of life tables and would help in building up realistic future trends in mortality for purposes of Statewise projections.

24. For want of reliable data on age-specific fertility rates, the overall general fertility rates have been made use of in the earlier projections. With the data available for a series of years from the Sample Registration Scheme and the National Sample Survey, it may be possible to build up realistic schedules of age-specific fertility rates and fix up related future trends for purposes of all India and Statewise projections.

25. The earlier projections had always faced the lack of data with regard to the impact of the family planning programmes on fertility in its various aspects. During the last few years the Ministry of Health have been attempting to build up a series of data on the progress of the various family planning programmes and have also initiated studies on evaluation of the effect of family planning on fertility. These data are expected to provide a mechanism by which realistic assumptions could be made on the future trends in fertility for purposes of population projections.

26. As has been the usual practice in earlier Censuses smoothed age distributions will be made Statewise. In the 1971 Census, single-year age data will be prepared on the basis of a 20 per cent sample. Since the sample size is large and since five-yearly age data based on the entire population will also be prepared at the time of preparation of the Primary Census Abstract, it is expected that no particular difficulty will be encountered in preparing smoothed series of Statewise age data.

27. Of late computer facilities for tabulation and analysis of data are being made available on an increasing scale. It is proposed to utilise the electronic data processing system for projection work so that a variety of assumptions on the components of population growth could be tested out for setting up a range of variation of the future population.

POPULATION PROJECTIONS IN IRELAND

14(d)-Ireland

1. Population prognostication in Ireland poses some interesting problems because of several unusual demographic features. It may be of interest to mention these briefly before describing the population projection work at present in hands.

2. The first and most important problem is in regard to emigration, which, for many years past, has been almost as important a factor as mortality in determining the level of the population. Year by year statistics of emigration are not, however, available because of the absence of formalities in travel between the Republic of Ireland and the United Kingdom and the high number of passenger journeys into and out of the State in relation to the total population; for example the 1966 population total was 2.9 million while the number of passenger journeys by sea and air out of the State in 1967 was 2.5 million. Inward movement was almost as high. The only firm figures are for net emigration for the intervals between censuses of population, obtained by comparing natural increase figures with changes in population as shown by successive censuses. In the intercensal period 1961-1966 net emigration was estimated at 81,000. Table 1 shows birth, death, natural increase and net emigration rates for a number of intercensal periods and it will be seen that for two of these periods the net emigration rate exceeded the death rate.

TABLE 1 - AVERAGE ANNUAL RATES OF CHANGE IN POPULATION, NATURAL INCREASE AND NET EMIGRATION PER 1,000 OF AVERAGE POPULATION AND NUMBER OF FEMALES EMIGRATING PER 1,000 MALES, 1926-1936 TO 1961-1966.

Intercensal Period	Change in Population	Births	Deaths	Natural Increase	Net Emigration	Females Emigrating per 1,000 males
1926-1936	-0.1	19.8	14.3	5.5	5.6	1,298
1936-1946	-0.4	20.5	14.6	5.9	6.3	662
1946-1951	+0.4	22.1	13.5	8.6	8.2	1,365
1951-1956	-4.3	21.3	12.1	9.2	13.4	817
1956-1961	-5.6	21.1	11.9	9.2	14.8	935
1961-1966	+4.6	22.0	11.7	10.3	5.7	1,143

3. Table 1 also shows the number of female emigrants per 1,000 males and this figure also shows wide fluctuation varying from 662 females per 1,000 males in the period 1936-46 to 1,365 in the following period 1946-51. Despite this, it is remarkable that over an extended period the numbers of male and female emigrants are almost equal. In the period 1926 to 1966, net emigration in the case of males was 492,000 and in the case of females 473,000. In a sense, emigration is even more important than mortality in Ireland in determining the level of the population since the groups mainly affected by emigration are the young adults. The age composition of emigrants may be inferred by comparing the expected number of survivors at each age from the preceding census with the actual numbers at the most recent census. In the case of the under five years age group the survivors are calculated from the number of births registered in the State in the preceding five years. These results, for net emigration during the period between the 1961 and 1966 Censuses, are shown in Table 2. In this period, which was one of relatively light emigration, the age groups mainly affected were 15-19 years and 20-24 years. The inward movement in the 0-4 year group and the 70-74 year group will be noted. The age structure of emigration is illustrated in Appendix A which shows the percentage of the population at each year of age emigrating within the next five years, based on the 1961-66 experience.

TABLE 2 - HYPOTHETICAL AND ACTUAL POPULATION BY AGE, 1966.
(Thousands)

Age group	Males			Females		
	Without migration during 1961-1966	Actual	Effect of migration between 1961 and 1966	Without migration during 1961-1966	Actual	Effect of migration between 1961 and 1966
0- 4 years	154.9	161.3	+ 6.4	148.6	154.6	+ 6.1
5- 9 "	152.7	152.4	- 0.3	146.8	146.5	- 0.2
10-14 "	146.7	145.6	- 1.0	140.4	139.8	- 0.6
15-19 "	147.9	133.3	- 14.6	140.3	126.0	- 14.2
20-24 "	119.8	94.7	- 25.1	113.2	90.6	- 22.6
25-29 "	80.0	75.0	- 5.0	77.3	74.3	- 3.0
30-34 "	71.8	73.8	+ 2.0	72.7	72.8	+ 0.1
35-39 "	74.5	76.5	+ 1.9	76.9	77.8	+ 0.9
40-44 "	80.5	81.0	+ 0.4	84.2	82.4	- 1.9
45-49 "	83.1	83.6	+ 0.6	84.1	82.9	- 1.3
50-54 "	85.9	84.3	- 1.6	83.2	80.6	- 2.6
55-59 "	76.9	75.8	- 1.1	72.4	71.2	- 1.1
60-64 "	62.3	61.6	- 0.6	63.5	62.2	- 1.3
65-69 "	55.1	55.0	- 0.1	60.2	59.2	- 1.0
70-74 "	40.3	42.2	+ 1.9	44.1	47.9	+ 3.8
75-79 "	30.3	28.7	- 1.6	36.0	34.1	- 2.0
80-84 "	15.9	15.8	- 0.0	20.0	19.8	- 0.2
85 and over	7.4	8.1	+ 0.7	11.5	12.2	+ 0.8
All ages	1,486.0	1,449.0	- 37.0	1,475.3	1,435.0	- 40.4

4. The inward movement in the 0-4 year group although surprising, appears to represent genuine immigration as the coverage of birth registration is believed to be practically complete. Presumably these children form part of the families of returning emigrants. From the 1961 Census inquiry on Birthplaces one can see that 8,627 children aged 0-4 enumerated in that Census had their birthplace outside of Ireland, tending to confirm that this inward movement represents immigration. The explanation of the immigrants in the 70-74 age group is the return of emigrants from the United Kingdom and the United States to retire in Ireland.

MARRIAGE:-

5. By European standards marriage rates in Ireland have always been very low. The crude marriage rates in Ireland and in a few other countries are compared in Table 3.

TABLE 3 - CRUDE MARRIAGE RATES PER 1,000 POPULATION IN IRELAND AND CERTAIN OTHER COUNTRIES.

*Ireland 1921-30	4.8
*Ireland 1931-40	4.9
*Ireland 1941-50	5.6
*Ireland 1951-60	5.4
Ireland 1961	5.4
" 1962	5.5
" 1963	5.5
" 1964	5.6
" 1965	5.9
" 1966	5.8
" 1967	6.1
" 1968	6.9
England and Wales 1966	8.0
Scotland	8.1
France	7.0
Sweden 1967	7.2
Spain 1967	7.2
United States 1966	9.5

*Annual average.

6. From this table it can be seen that marriage rates in Ireland were very low up to recent years, since then there has been a dramatic increase, the crude marriage rate rising from 5.8 per 1,000 population in 1966 to 6.9 per 1,000 in 1969.

TABLE 4 - PERCENTAGES OF WOMEN SINGLE IN CHILD-BEARING AGE GROUPS IN IRELAND, 1946-1966, AND IN VARIOUS COUNTRIES ABOUT 1961.

Country and Year	Age Group							Average* Duration of Celibacy
	15-19	20-24	25-29	30-34	35-39	40-44	45-49	
Ireland 1946	98.4	82.5	57.5	38.8	32.1	27.6	26.3	18.2
Ireland 1951	98.9	82.3	54.4	36.4	28.5	26.7	26.3	17.7
Ireland 1956	99.1	81.3	53.3	33.3	27.4	23.6	24.6	17.1
Ireland 1961	98.9	78.2	45.1	29.6	23.5	22.0	22.1	16.0
Ireland 1966	98.4	74.8	37.8	24.1	21.2	19.5	20.4	14.8
Albania 1955	75.4	23.8	7.7	3.8	2.5	1.7	1.5	5.8
Austria 1961	94.0	58.0	24.2	15.9	14.4	12.4	12.1	11.6
Czechoslovakia 1961	91.3	32.8	10.0	6.9	6.5	6.2	6.2	8.0
Denmark 1960	95.1	45.9	14.5	9.4	8.2	8.1	9.4	9.5
Germany (BRD) 1961	94.9	54.6	20.8	13.3	12.7	10.8	9.3	10.8
Greece 1961	94.2	65.3	34.7	18.2	10.1	8.2	n.a.	n.a.
Hungary 1960	85.3	31.4	11.2	8.5	7.6	7.4	7.3	7.9
Netherlands 1960	96.3	59.4	20.5	12.1	10.4	10.1	11.0	11.0
New Zealand 1961	91.6	40.5	12.4	7.8	7.3	7.4	8.3	8.8
Norway 1960	95.2	49.8	17.8	11.6	10.9	11.2	12.9	10.5
Peru 1961	83.4	44.2	25.0	17.8	13.2	14.6	14.5	10.6
Poland 1960	90.9	40.3	17.3	14.0	15.1	19.0	24.3	11.0
Portugal 1960	95.3	62.1	32.0	21.8	18.2	16.9	15.9	13.1
Spain 1960	97.9	73.4	34.9	20.1	15.3	15.5	14.6	13.6
Sweden 1960	97.3	57.5	20.7	11.9	10.0	9.5	11.0	10.9
United Kingdom 1961	93.4	42.0	15.7	10.9	9.8	9.5	10.5	9.6
United States 1960	83.9	28.3	10.5	6.9	6.1	6.1	6.5	7.4
Venezuela 1961	77.3	41.8	26.8	22.1	21.3	24.1	24.6	11.9

*The average duration of celibacy is defined as 5 times the sum of a row divided by 100.
n.a. = not available.

7. The pattern of low Irish marriage rates is illustrated also by Table 4(1) which shows the percentages of women single at ages 15-49 years in Ireland and a number of other countries.

8. Another unusual feature of Irish marriages has been the number who emigrate directly after marriage. Such information became available for the first time in 1957 when a question as to future permanent residence after marriage was introduced on the marriage form. Since 1957 as can be seen from Table 5 the percentage of such marriages has been fairly stable at between one-eighth to one-seventh. However an important component of this movement is formed by persons who have already emigrated and return for the marriage. In 1967 about two-thirds of these grooms and one-third of the brides gave area of residence before the marriage as outside the State.

TABLE 5 - TOTAL MARRIAGES, MARRIAGES WITH FUTURE PERMANENT RESIDENCE ABROAD 1957-68.

YEAR	(1) Total marriages	(2) Marriages with future permanent residence abroad	Col. (2) as a percentage of Col. (1)
1957	14,657	2,064	14
1958	15,061	2,247	15
1959	15,420	2,173	14
1960	15,465	2,395	15
1961	15,329	2,266	15
1962	15,627	2,037	13
1963	15,556	1,949	12
1964	16,128	1,995	12
1965	16,946	2,004	12
1966	16,849	2,272	13
1967	17,788	2,242	13
1968	18,993	2,293	12

9. It should be mentioned that there is no legal provision for divorce in Ireland, one factor, at least, tending to simplify Irish population projection work.

FERTILITY:-

10. The low marriage rate in Ireland is counteracted by a very high fertility of marriage so that the crude birth rate is high by European standards; in fact in 1966 the Irish crude birth rate was fifth highest in Europe, coming after Albania, Iceland, Northern Ireland and Portugal.

11. The very high fertility of marriage in Ireland by comparison with England and Wales may be illustrated by Table 6, which gives legitimate age-specific fertility rates. In the 15-19 age group the Irish rate was 33 percent above that in England and Wales, in the 20-24 age group 73 percent higher, while the excess in the Irish rate rose steadily with increasing age.

TABLE 6 - LEGITIMATE BIRTHS PER 1,000 MARRIED WOMEN IN EACH AGE GROUP IRELAND AND ENGLAND AND WALES.

Age of mother at maternity	Ireland 1966	England and Wales 1965
15-19 years	639.2	481.3
20-24 "	484.5	280.7
25-29 "	369.6	201.5
30-34 "	280.9	108.8
35-39 "	186.0	51.2
40-44 "	70.7	13.3
45-49 "	5.8	1.0
Total 15-49 years	195.8	105.1

12 Following this brief review of the main demographic characteristics, the various assumptions and methods used in the demographic projections are considered in greater detail.

MORTALITY:-

13. Projected trends from the 1951 and 1961 Life Tables are used. The factors $\frac{L_{x+5}}{L_x}$ from the 1951 and 1961 Life Tables are compiled for each age x , that is the proportion of males (or females) at each year of age who will survive five years. Call this $S_x = \frac{L_{x+5}}{L_x}$. These proportions show a rise, substantial for the younger ages, between 1951 and 1961. It was assumed that these figures would continue to rise following a hyperbolic curve, rising asymptotically to 1. Values of the factor $\frac{L_{x+5}}{L_x}$ are obtained from this curve as appropriate for the intervals 1966-71, 1971-76, 1976-81

EMIGRATION:-

14. As already explained it is not possible to work in terms of gross emigration and immigration since there is no information on these separately even for total figures, much less for single years of age. All the assumptions and calculations are, therefore, in terms of net emigration only.

15. Two methods of incorporating emigration assumptions are being tested. The first is to assume total net emigration figures for five-yearly periods e.g. one pattern might be 1966-71, 90,000; 1971-76, 65,000; 1976-81, 40,000; 1981-86, 25,000 etc. The assumptions here are, however, to a large extent arbitrary since the level of emigration cannot be directly controlled. On the other hand a factor clearly linked with emigration, namely the rate of creation of new jobs can to a large extent be influenced by Government action. The second method of dealing with emigration is to make assumptions regarding future employment (consistent with national development plans) and participation rates, in 1971, 1976, 1981, 1986 etc., deduce the corresponding future populations and hence derive the net emigration in 1966-71, 1971-76 etc. which would result in these future populations.

16. Considering first the simpler method of specifying the net emigration for five year periods, assume that net emigration in the period 1966-71 will amount to 90,000. It is assumed that the total net emigrations of males and females are equal.

17. As already mentioned there is a special problem in the case of infants under 1 year. It appears, from comparing census of population and birth registration results that there has been a substantial immigration of infants at this age, amounting to about 2,000 per annum. To allow for this it is assumed that the number of baby immigrants will equal 2,000 per annum equally divided between males and females since no information is available on the trend in such infant immigration. Thus, if net emigration in a five year period amounts to 90,000, then the net immigration under 1 year is 10,000 while the net emigration at ages over 1 year is 100,000.

18. As regards ages over 1 year, the allocation of net emigration over the individual years of age is done in the following way:-

Comparing the 1961 and 1966 Census results by ages, the proportion remaining in the State from age x in 1961 to age $x+5$ in 1966 is calculated for each age (for each sex separately). Call this R_x .

$$\text{i.e. } R_x = \frac{P_{x+5}(1966)}{P_x(1961)}$$

There are two sources of decrement here, death and emigration. Assuming no emigration the proportion surviving would be S_x (already referred to) and the value would be for the mid-point on the curve between 1961 and 1966. If one calculates $M_x = \frac{R_x}{S_x}$, the M_x represents the proportion surviving five years if there was emigration but no deaths. A jagged set of values of M_x is obtained on the 1961-66 experience. A smoothed set of M_x 's is calculated using a summation formula. This smoothed set of M_x 's form the basis of the emigration assumptions. The emigration rate, $Q^2_x = 1 - M_x$. All the Q^2_x are scaled up or down by a constant factor K to produce the desired total emigration. The total net emigration at ages over 1 year, assuming 1961-66 rates apply, is given by the formula

$$\sum_{1}^{\infty} (1966) \frac{Q^2_x (1 - \frac{1}{2} Q^1_x)}{P_x} \quad \text{where } Q^1_x \text{ is mortality rate for persons aged } x \text{ years.}$$

For the period 1966-71 for males, this figure worked out at 52,000. The assumption we are aiming at is 50,000 i.e. half of the 100,000 mentioned earlier. By multiplying the Q^2_x 's by the factor 50,000, (= K) the net emigration for males over one year is adjusted to 50,000. Then the survivors

in Ireland in 1971 out of the population aged x in 1966, i.e.

$$P_{x+5} \text{ (1971) is } P_x \text{ (1966) } S_x (1 - KQ^2_x).$$

In this way the number of males and females at each year of age over five years is calculated. Persons aged under 5 years in 1971 would not, of course, have been born in 1966 so that to calculate these, assumptions about marriage and fertility rates must be made. For each subsequent 5 year period the adjusted net emigration rates (i.e. KQ^2_x) are scaled up or down by a common factor to yield the predetermined aggregate net emigration.

FERTILITY:-

19. The basic data entering into the projections of numbers of births is that derived from the

-Fertility of Marriage inquiry, at the 1961 Census of Population. The principal results of that inquiry is a classification of married women by age at marriage and duration of marriage. No such data were available from the 1966 Census, but it was possible to make use of the information on births registered in the period 1961-66.

20. The "Age at marriage x Duration of marriage matrix" was adjusted to October, 1963, i.e. the mid-point of the 1961-66 interval using data on marriages registered during the period. From the 1961 Report on Vital Statistics it proved possible to estimate fertility rates specific for duration of marriage and age at marriage. These rates were adjusted so that when multiplied by five and applied to the October, 1963 "Age x duration matrix" they gave the number of births registered in the intercensal period 1961-66.

21. The main part of the work in projecting the number of legitimate births in each five-year period consisted of constructing this "Age at marriage x Duration of marriage matrix" for the mid-point of each five-year interval. This was done, by the computer, by applying suitable survivorship factors to the matrix for the period five years earlier. Thus the 0-4 years duration group became the 5-9 year group and so on. Age at marriage, of course, remained unchanged.

MARRIAGE:-

22. It was necessary now to insert the line of the matrix for duration of marriage 0-4 years, i.e. those marrying in the preceding five-year interval. This presented a difficulty however. Marriage rates should be expressed in terms such as "marriages at age x, divided by single population aged x". The estimates of population at each year of age, at this stage in the computer storage, were total figures only and not divided into single and married. The input to the computer had, therefore, to be converted to marriage rates per unit of total population. Having examined the trends in marriage rates per unit of single population in each age group, it was decided to project these linearly. Corresponding rates per unit of total population in each age group were calculated manually from these, and these figures formed part of the computer input.

23. At this stage account had to be taken of the complication mentioned above that in approximately 12 per cent of all marriages the area of future residence of the couple was stated to be outside Ireland. The foregoing calculations were, therefore, carried out excluding such marriages and the rates entering into the projections are, in fact, what might be called "marriage, stay at home, rates".

24. Having obtained the "Age at marriage x Duration of marriage matrix", fertility rates were applied to these to give the legitimate births in the five year period. Illegitimate births were assumed to equal 2.2 per cent of legitimate births over the whole period of the projections. Total births were divided into male and female in the ratio 1.056 to 1.

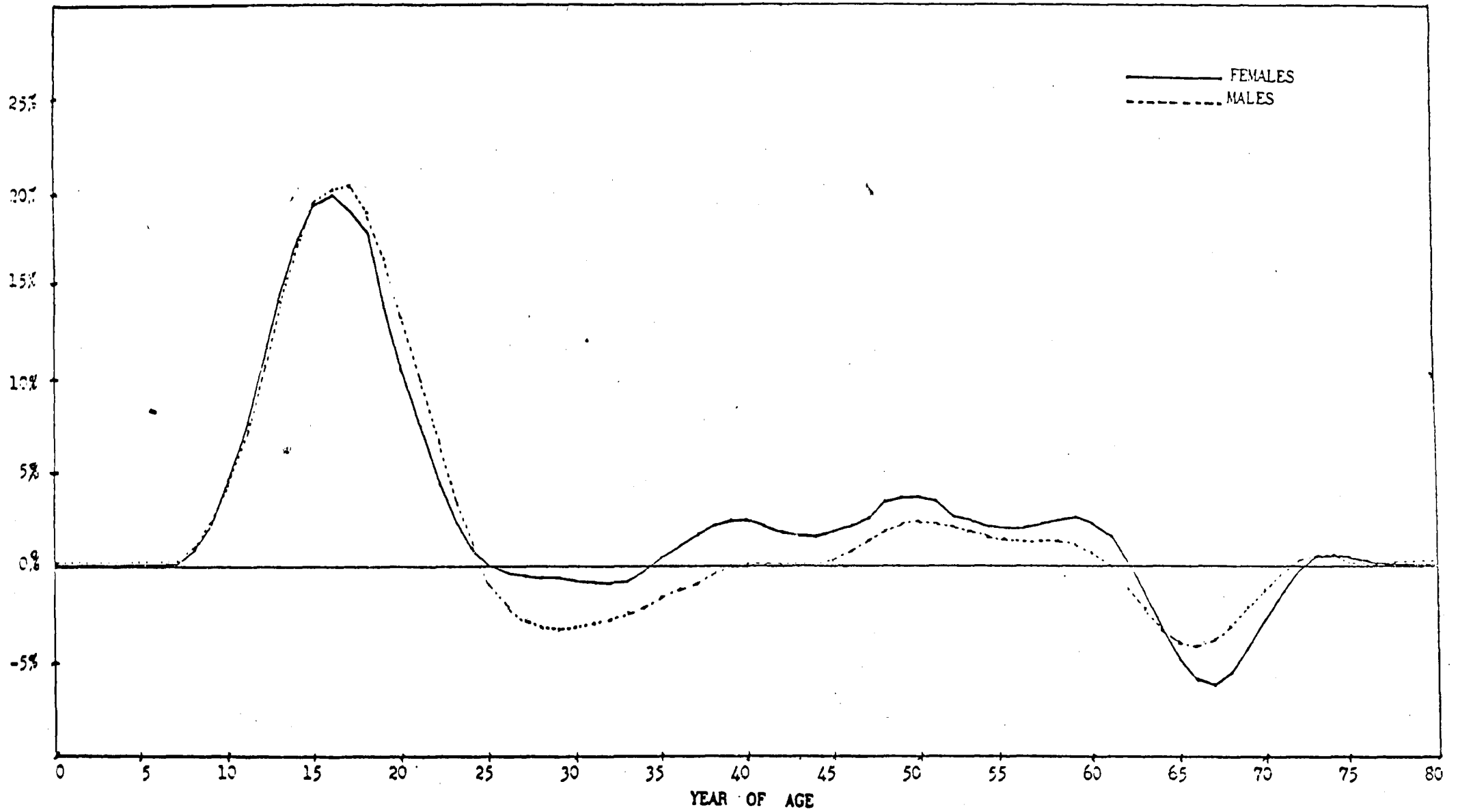
25. The system is designed so that the assumptions as regards emigration and fertility can be inserted on cards separate from the main programme. Thus these factors can be varied easily and projections on any desired assumptions obtained. Fertility can also be increased or decreased by a fixed amount, over the period, on an "age x duration" basis.

Emigration, second approach.

26. Here, the assumed total labour-force in the years 1971, 1976, 1981, 1986 etc. form the input to the computer, together with assumed labour-force participation rates by five-year age groups for these years. As a first step, for the 1966-1971 interval, the computer calculates projected populations in 1971, by age groups, on two emigration assumptions, namely nil emigration and 100,000 net-emigrants for the five-year period, using the programme already described. It then calculates the total 1971 labour-force corresponding to each of these assumptions. By interpolating between these two labour-force figures, a net-emigration figure is found corresponding to the desired 1971 total labour-force. As labour force decreases linearly with increasing emigration, simple linear interpolation is sufficient here. Having obtained the net-emigration figure to give the desired 1971 labour-force, the 1971 total population and population by ages are obtained. Thus, to get from the 1966 population by ages to the 1971 population by ages, three projections are required involving, (a) nil emigration, (b) 100,000 net emigrants and (c) a level of emigration calculated by the computer to produce the desired labour force. By a similar procedure projections for subsequent five-year periods are obtained.

APPENDIX A

Percentage of Population at each Year of Age Emigrating within Five Years
(Based on 1961-66 Experience)



14(d) - NEW ZEALAND

BRIEF HISTORY OF OFFICIAL POPULATION
PROJECTIONS IN NEW ZEALAND

1. The first population projection to be published by the Department of Statistics was a provisional one giving the total number of persons in the total population (Non-Maori and Maori). Its base was the mean population for the year ending 31 March 1953, and a constant annual rate of increase and a constant inflow of migrants was used in conjunction with survival factors.
2. Subsequent sets of component projections, of which there have been five, used a component method. These were carried out using projection points five years apart, until the projections with base 31 December 1963, and subsequently 31 December 1967, were calculated using computer facilities; allowing projection intervals to be one year apart.
3. Birth rates specific for age of mother were used in the construction of the projection with base population 31 December 1952, but for all later projections, birth rates specific for age and marital status of mother were used, with net marriage rates to project the proportion of married women at each age.
4. Mortality rates have been taken from the latest life tables available. In all projections no attempt has been made to project trends in the net inflow from migration, but average levels of net inflow based on past experience were used.
5. All published projections have been for the combined population, apart from those with bases at 31 December 1952 and 31 December 1968 when separate projections were carried out for Maori and Non-Maori population.

Economic Significance of Population Projections

6. Planning for economic and social development has become firmly established as a role of government in most countries of the world. For rational and effective planning a comprehensive statistical information system is required. The first requirement is an accurate quantitative description of the current economic and demographic situation. The next requirement is a number of statistical projections on the basis of existing policies and trends and of known developments.
7. One of the most basic statistical projections is the population projection. Population determines both the supply of labour available to produce goods and services and the potential number of consumers of those goods and services. Both the supply of labour and the demand for consumer goods and services vary with the age, sex, marital status, level of education, income etc. of the population. Thus if the projected population is classified by all these relevant variables it is possible to project the supply of labour and the demand for goods

and services, e.g. dwellings, electricity generation, and consumer durables such as cars and telephones, quite realistically.

8. While demographic factors have a significant effect on the demand for goods and services, the economy also has a significant effect on demographic factors. Labour force participation, fertility and nuptiality have been shown to be related to income at both the micro and macro level. Similarly the demand for consumer goods and services is related to income. These relationships are not easy to measure and tend to be overshadowed by demographic and social factors in many cases. However, with the current emphasis on economic model building attempts must be made to take this feedback effect into account.

An important use of demographic projections is in the planning of social services. In a welfare state like New Zealand, the government has to be aware of changes in the structure of the economically and socially dependent population (the young, aged, handicapped, etc.) and to face up to the problems of continuing expansion, especially in the population at the younger ages, if it wishes to provide an equitable distribution of income (in benefits, for example) and facilities to the different sections of the population, and to formulate policies to continue these provisions in the future.

The Concept of a Projection

9. A statistical projection uses the most accurate statistics available on the current situation and specific assumptions about the future course of those factors which influence the variables of interest, to calculate projected values of them. The uncertainty of future economic and social developments means that forecasting must always be largely subjective, and it is this that distinguishes forecasts from projections. Because the role of the Department of Statistics in New Zealand must be an objective one the department limits itself to the making of projections based on current and past trends or, where the assumptions have a normative or policy basis, to the making of forecasts for planning organisations using assumptions defined by them.

The Information Base

10. If timely and realistic population projections are to be made historical data are needed on trends in fertility, mortality, nuptiality and a projection base that are both timely, comprehensive and accurate.

11. The timeliness of demographic projections is more important for short-term than long term projections. Attention needs to be given to the timeliness of the processing of census and other statistics, the calculation of the projection base and

14(d) - NEW ZEALAND

the calculation and publication of projections. The efficient use of computer facilities plays an important part in carrying out these tasks more quickly.

12. The most important factor in obtaining accurate historical data and an accurate projection base is the existence of a computerised demographic data system that incorporates an accurate method of calculating intercensal estimates of the population classified by the relevant demographic variables.

13. In New Zealand the department has relied on a standard elementary method of calculating intercensal estimates classified by age, sex and race that can be handled using a desk calculator. However, in order to incorporate variables such as marital status, duration of marriage, children ever born etc. into population projections, it is realised that it will be necessary to incorporate these variables into intercensal estimates as well. It would also be valuable to have these estimates available for every quarter of the calendar year, and to increase their accuracy. As a result the department is carrying out research to produce intercensal estimates that are classified by these variables and are more accurate and available for more points in time.

14. A final method of increasing the information base is to obtain more information from basic statistical record documents. Research is needed on the design of suitable questions for incorporation in basic statistical and administrative documents, and the integration of this data into projection methodologies. Information is particularly required on fertility, labour force participation, internal migration and household formation.

Realism of Population Projections

15. Once a projection has been made, it is important to evaluate its performance to see which assumptions are proving realistic and which unrealistic. Research can then be directed to the areas where it is required. Generally, demographic projections made by the department have agreed with actual experience fairly well, but of course discrepancies of some magnitude are bound to occur. An analysis of the sources of "errors" in New Zealand population projections, made over the past decade where error in this context means the difference between a projected value and the actual population, showed that "errors" in the projection of fertility and migration contributed about 45 percent each and mortality about 15 percent to total "error". (Strictly speaking, there is no such thing as an "error" in a projection, except those of an arithmetical type). "Errors" in the projection of marital status contributed to the "errors" in fertility but also affected the projection of labour force and households.

14(d) - NEW ZEALAND

16. Migration is one of the most variable elements of population growth but is the most difficult to project. It is subject to purely economic conditions, but it is difficult to quantify the relationship.

17. Fertility is more amenable to demographic analysis and it is hoped that by increasing the number of relevant variables, such as birth order of child, parity of mother and birth interval etc., more realistic projections will be possible in the future.

18. Mortality contributes little to the "error" of New Zealand projections. While at the youngest ages some improvement in projection realism could be had by projecting trends in death rates at older ages recent increases in male mortality in New Zealand (as in some other countries) have shown that such projections would be hazardous. For this reason projections of mortality have not been made for use in New Zealand population projections.

Labour Force Projections

19. Manpower planning requires projections of the labour force in total, and classified by industry, occupation, educational attainment, level of skill and by region on both the demand and supply side. In this way imbalances between supply and demand for the total labour force or a particular classification of it can be ascertained.

20. The Department of Statistics in New Zealand has made projections of the total labour force supply, using demographic techniques, since 1959. These projections have tended to under-estimate the participation at middle ages (especially of married women) and over-estimate the participation at young ages. Methods of taking into account education enrolment trends at young ages and such variables as the presence of children under five years of age which affects the participation of married women are currently under study.

21. The projection of occupational supply is only partly a demographic exercise. Estimates of entrants (transfers from other occupations, migrants and entrants from educational institutions) are depleted by estimates of deaths and retirements and removals for other reasons. However, it is a complex task requiring a detailed analysis of each component of change for each occupation, consistent definitions of occupation, consistent definitions of occupation and educational attainment, and detailed data on occupational mobility, migration and the output of educational institutions. It is therefore not surprising that while projections have been made for specific occupations such as doctors and teachers, no comprehensive and integrated set of projections has ever been prepared for New Zealand.

14(d) - NEW ZEALAND

22. The projection of the demand for labour is an exercise in economic techniques. In New Zealand, the most recent approach has been to adopt a stated rate of growth of real gross national product as a target, and then to relate industry output to total output using input-output analysis. The industry outputs were then converted to employment levels using projected values of productivity and average hours of work.

23. From projections of industry demand it is possible to obtain projections of occupation demand by the use of an industry-occupation classification matrix. In New Zealand the application of these techniques has been hampered by the lack of use of a consistent classification of industries and occupations in labour force statistics, census statistics and national accounts. As a result discussions are not being held for the purpose of having the International Standard Industrial Classification and the International Standard Classification of Occupations adopted uniformly in all official employment and earnings surveys.

Household Projections

24. Projections of families and households are useful in economic and social planning in view of the many types of goods and services for which households and families are units of consumption. Projections of consumption of dwellings and of consumer durables such as telephones and cars, can be made from these basic projections of households and families.

25. The projection of families is basically a demographic exercise and can be done straightforwardly as a supplement to the making of the population projection. The main component of variability is of course the marriage rate. In order to project marriages and proportions married use can be made either of nuptiality tables, with the calculation of the effect of marriage rates, death rates and migration rates on proportions married separately, or use can be made of the within-cohort change in proportions married as derived from census of population data. The latter method has been used in New Zealand but because it does not analyse the effect of marriage, mortality and migration separately, research is proceeding on the use of nuptiality tables in projecting marital status and marriages.

26. Derivation of projections of households from the projections of families is in fact a projection of dwellings because each head of household represents one occupied dwelling. Thus the projection of households should ideally involve a consideration of economic factors such as the cost of housing, the availability of loan finance and the income distribution. Usually, however, the quantified relationships of such variables are difficult to determine and instead the projections are based solely on demographic considerations. Head-of-household rates

from a previous census of population (i.e. the proportion of the population in each age, sex and marital status category who are heads of households) are applied to the projected population classified by age, sex and marital status to obtain a projection of heads of households.

27. Projections of consumer durables can be made on the basis of these household projections. The actual requirements of consumer durables must in addition take into account destruction rates (e.g. loss of houses due to fire, demolition and other causes) and the number of durables that are not being used (e.g. unoccupied houses). Political or administrative considerations may also have to be taken into account (e.g. the maintenance of certain standards of housing). Projections of families and households are also of use in the planning of social security provisions.

Regional Projections

28. There is a growing demand for regional projections to supplement the national projections. Projections of regions are required for the planned allocation of resources within the country and are needed particularly for planning urban development programmes.

29. In New Zealand, because of lack of statistical resources regional projections have not been made by the national statistical office. However regional projections of population in total numbers using historical trends and growth rates modified by information about the locality itself and proposed developments within it have been made by the Ministry of Works, Town and Country Planning Branch. These have been done for local authority areas - counties, towns, cities etc. However, there is considerable demand for projections that are on a sounder methodological basis and that are also classified by age, sex and marital status so that they can be used to derive projections of labour force supply and households by regions. As a result the Department of Statistics has initiated research into regional mortality, fertility and net migration flows (using residual methods) in order to make regional projections by the component method. In addition, there will be a question in the 1971 Census of Population questionnaire on internal migration which should aid understanding of the factors influencing gross migration flows.

PERSONNEL TRAINING

1. Over the last few years the New Zealand Department of Statistics has faced a gradual rundown in its numbers of qualified and experienced senior and intermediate staff. This has occurred at a time when increasing demands on the service it can give have come from both government and private sectors. The synthesis of this demand is contained in the Technical Committee's Report on Statistics in which a wide variety of extensions to statistical services have been asked for and in which considerable centralization within the national statistical office has been suggested. On the other hand the university system in New Zealand does not produce academically qualified statisticians but rather academics qualified in a variety of useful disciplines such as economics, mathematics, geography, sociology, commerce and arts. The Department is thus faced with a continuing problem of organising its own staff training scheme and coupled with the big intake of staff to meet modern commitments with a crash programme to produce practical statisticians at short notice.

2. With the help of the State Services Commission (the personnel department of New Zealand's State Services) and the tertiary and university educational authorities, the Department plans to overcome these difficulties in a variety of ways.

Training Needs

3. The Department sees its training needs under the following headings:
- (a) Pressurised training of the personnel at the intermediate level and senior level. These officers will come from a variety of government agencies but will not necessarily have had any wide statistical experience.
 - (b) Training of a limited number of young cadets entered directly from high school and usually having completed School Certificate or University Entrance. Average age is about 17 or 18 years.
 - (c) Training of under-graduates and graduates in departmental methods. These officers are usually in their 20's on their arrival in the Department.
 - (d) Training of other entrants in older age groups. These personnel frequently have only limited educational qualifications and are engaged to carry out, in the main, the bulk of processing of input data before it is punched for electronic data processing.

- (e) Upgrading the knowledge of statistical techniques of younger intermediate staff who have had relatively little formal education in this field.
- (f) Disseminating specific departmental techniques and policies among all staff, particularly at the intermediate and senior levels.

Meeting the Training Needs

4. The Department recently increased its strength of staff devoted directly or indirectly to training. One officer is now responsible for organising training, dealing with educational problems of staff, assisting with training, etc; another officer is responsible for the development and maintenance of training manuals; and a third officer engaged full time as a tutor in statistics. With this group, supplemented by assistance with lecturing by senior officers and by formal courses at tertiary educational institutions, the Department expects to be able to meet the extraordinary training needs which have arisen.

Training of the intake of Senior and Intermediate Officers

5. As these officers will be experienced and trained in general clerical investigatory work or executive methods and common governmental procedures, they will require in addition:
- (a) to acquire background knowledge of this Department and
 - (b) in many cases, to acquire a knowledge of statistical techniques.

Both of these needs will be met by full or part-time courses utilising, mainly, the services of the Tutor in Statistics, at least in the early courses. It will be necessary to persist with further courses at later dates in order to bring the general standard to a sufficient level. For the first courses, use is being made of the ECAFE "Manual on Training of Statistical Personnel at the Primary and Intermediate Levels", published in 1963, and its Supplement "Training Notes and Exercises", published in 1964. This Manual will be supplemented by lectures from experienced officers in selected statistical techniques. Course participants will generally have had previous 'on the job' courses in supervisory and management techniques; where these service-wide courses have not yet been undertaken the new recruits to the Department will be given the appropriate courses by the State Services Commission.

Training of Young Cadets

6. The Department has not been recruiting and retaining enough school leavers and training measures provided for them no longer meet the emerging requirements of the developing work functions of the Department. To ensure that the operational statistician at middle and senior levels ultimately has a reasonably

extensive theoretical grounding in the increasingly technical field of statistics, the Department will sponsor a planned scheme of training and education lasting four years, for suitable school-leaver recruits to the Department. It is expected to accomplish this in cooperation with the Technician's Certification Authority, which provides a course and Certificate in Statistics. At present this course is beamed primarily for statistical technicians working in the natural sciences and biological fields. By adaptation of curricula for some subjects and introduction of some additional options an appropriate course has been arranged that more appropriately meets the needs of statistical work in the economic and social fields. We hope to attract school leavers with endorsed School Certificate or lower sixth leaving certificate in at least English and Mathematics.

7. The formal training course would be supplemented by a period of initial practical training in the first year, by controlled working arrangements in small teams in selected branches during the second year, working arrangements as individuals in three to four representative sections in the Department in the third and in the final year work experience would be specialised in the field most appropriate to the trainee's aptitude and the Department's need.

8. In the first year of the scheme formal educational arrangements will require six hours of day-time lectures at the appropriate tertiary educational institutions. The practical training in this initial year will be based on a modification of the ECAFE course mentioned above with some exposure to work environment for case study activities and introduction to basic techniques in selected subject matter survey sections. In subsequent years four hours per week will be devoted to day-time lectures at the educational institute; this will be supplemented by additional lectures in the trainee's own time. Besides the arrangements for practical training mentioned in the previous paragraph throughout the second, third and fourth year trainees will be required to attend special departmental courses covering such topics as sample surveys, preparing statistical publications and reports, index number theory and practice, and demographic statistics.

9. The aim of the traineeship is to produce a qualified technician statistician. In broad detail the scheme has been conceived:

- (a) to give proficiency in the techniques which the trainee is required to use;
- (b) to give a general appreciation of other techniques used in work which impinges on that which the trainee is doing;
- (c) to give him the necessary theoretical background to his work;

- (d) to develop methodical work habits, accuracy and speed.

In short, to fit him for and to interest him in general survey operation and statistical analysis and interpretation work in the Department. On successfully completing the four-year training and education programme the trainee would be granted an officially recognised certificate to this effect. Initially about ten traineeships would be given each year.

Training of Under-Graduates and Graduates

10. In addition to this scheme post-primary school leavers with the required entrance qualification would be given time off on full pay to attend university for an approved degree course leading to qualification in an appropriate discipline such as economics, mathematics, sociology, geography or commerce. In these cases practical training paralleling that given to technician trainees would be given by the Department.

11. Training needs of graduates (or near graduates) are of a specific nature. They are generally recruited to undertake specialised research and development duties in economic, social or demographic statistics. While well qualified in their respective disciplines they are relatively inexperienced about the practical conditions surrounding the market place and industrial activity. In some cases they also need special tuition in advanced statistical techniques such as sample survey design, multiple decrement tables, demographic techniques, index number theory and national accounting concepts. These lacks are supplied by controlled reading courses, lectures by senior officers and by practical Training Sessions organised by the Tutor in Statistics.

12. Where the discipline in which the person is qualified is of less detailed application to the Department's needs it is planned to supplement the person's knowledge by a diploma course at the University of Wellington. This will supply a general course in economic and social statistics and will also have application outside of the Department. Plans are under way to commence the first course in 1971.

Training of other entrants in older age groups

13. Experience has shown that a worthwhile contribution to data preparing for computer and other inputs and for arranging of data for publication can be made by the older age groups who are entering the labour force for a second time or who have changed recently to clerical and allied duties from another form of occupation. Training needs for this group are less diverse and within these restricted areas of statistical processing intensive training programmes are required if the entrants are to become fully productive in a reasonable period of time after joining the Department. Various departmental courses are planned; they will be organised by the training officer and for the most part carried out by the Tutor in Statistics assisted where appropriate by senior officers for specific lessons.

Upgrading Knowledge of Statistical Techniques

14. In the dynamic situation surrounding statistical work and its applications there is a continuing need to expand and revise our statistical knowledge. It is planned therefore to hold from time to time refresher courses to meet the varying needs of junior, intermediate and senior personnel. Some of these lectures will be given from departmental sources, others of a managerial or specialised management services nature will be given by the State Services Commission specialists and others by invited guest speakers or by attendance at conferences and seminars organised by professional or technical organisations. Within the limits of financial resources allocated by central government for these purposes attendance at international gatherings is one of the many appropriate vehicles for this type of updating of statistical knowledge. Within the national scene the Department recently organised a special seminar to disseminate knowledge of the new system of national accounts to both statisticians and users of statistics. Other similar gatherings will be held in the future.

Disseminating Departmental Techniques and Policies

15. With the increasing emphasis on an integrated and coordinated system of statistical services without duplication or omission in its classifications and concepts it is of paramount importance that standardisation of departmental techniques is brought about and that continuing checks are made on the uniform application of these techniques and of departmental policies.

16. The vehicle for achieving this cohesion is a system of staff training manuals of instructions on statistical procedures, standards, concepts and classifications. This system is soon to be updated and the regulatory function of the Department strengthened so that the application of these standards can be checked continuously.

17. Much still remains to be done to prove the success of our revised training proposals. With the enlarged intake of staff to meet the urgent demands for a modern statistical service for indicative planning and for economic and social policy needs these new training proposals must be implemented with urgency. We are confident that the training programme will be successful.

PROBLEMS OF MAINTAINING HIGH STANDARDS
OF SURVEY METHODOLOGY, ANALYSIS
AND COMPILATION TECHNIQUES

INTRODUCTION

1. In recent times scientifically planned sample surveys have come to be used for collection of statistics in many fields. Considerable experience has been built up in this country in the past two decades in the planning and organisation of such surveys and in the development of efficient procedures of data processing and analysis. Some of the problems experienced in India in connection with designing and conducting large scale sample surveys are described in this paper.

SURVEY PLANNING

2. Surveys are conducted in India by direct physical observation as well as by the interview method. The mailed questionnaire system is not feasible as the standards of literacy are still not adequately high. Hence substantial resources have to be earmarked for the collection of data by observation or personal enquiry. The data requirements have to be clearly visualised in a comprehensive manner and concepts and procedures developed in such a way that they can be clearly understood by various types of staff who are engaged for the work of primary data collection. In a large country like India, with a wide spectrum of geographical and climatic features as well as historical and socio-economic development, it is obviously difficult to evolve concepts and procedures which will be uniformly appropriate throughout the country. For instance, the concepts of unemployment and under-employment in rural areas where agricultural employment is markedly seasonal will have to be quite different from those adopted for urban areas. The procedures for land utilisation survey have to be different for areas which are cadastrally surveyed and those which are unsurveyed.

3. For the purposes of planning and evaluation, bench-mark information on a number of key indicators of the economy is usually required for the lowest administrative (or geographical) units. Limitations of resources often preclude any detailed geographical or classificatory breakdown of information for all levels through sample surveys. The choice of levels for which information is required and the precision with which estimates are to be obtained have, therefore, to be balanced with the availability of resources.

One way of obtaining usable estimates for lower levels could be by pooling the data collected for two or three successive years, if the annual variations could be ignored. This procedure was adopted in the land holdings surveys undertaken by the Indian National Sample Survey (NSS) during 1960-61 and 1961-62 to obtain better State level estimates of distribution of holdings. To give another example, in the Sample sector of the Annual Survey of Industries the annual sample covers each year an independent and randomly selected one-third of the total number of factories in the sector and the entire sector is covered in the course of three years. This will facilitate pooling up of the results to obtain estimates for all industry groups right down to the lowest digit of the standard industrial classification, while the annual results will be given up to the 3-digit level.

4. It often becomes a matter of considerable debate as to whether survey on a particular subject should be carried out year after year or whether a periodical enquiry would suffice. In the context of some socio-economic enquiries it has been suggested that a detailed enquiry on a large sample need be made only periodically for obtaining reliable bench mark information, while for studying changes over years, a thinner annual enquiry might be adequate. Such a scheme may, however, imply uneven work load from year to year, and hence uneven staffing pattern. In a permanent survey organisation with more or less constant staff resources, it is not possible to organise periodic recruitment of a large number of qualified staff at short notice, only to be disbanded after the survey period. It will, therefore, be advisable to develop an integrated programme in such a way that bench mark surveys on different subjects may be spread out evenly over a few years with simultaneous thinner enquiries for the remaining subjects. This will necessarily call for long-range programming.

5. As India has a predominantly agricultural economy the socio-economic activities of the rural population have a seasonal character so that it becomes important to collect information for different parts of the year. To get at the total annual picture for a household, it is desirable to canvass information for the same set of sample households at different points of time. In socio-economic enquiries by interview, this procedure implies repeated visits to the same set of informants, and is observed to result in informant-resistance. The NSS gets round this difficulty by canvassing information for different sub-periods of the year from different sets of independent sub-samples so that valid estimates are available for each sub-period as well as for the whole year.

6. Considerations of economy of operation lead to the use of a common set of primary sampling units (PSU's) for different enquiries. In crop yield estimation surveys, for instance, instead of choosing a separate sample of villages for each crop, a common sample list is drawn up. Experience has shown that multi-subject surveys undertaken on a continuing basis will be generally more economical as compared to a series of individual enquiries on different subjects, provided the latter are not so numerous or diversified as to affect the quality of data. This economy becomes important in developing countries on account of paucity of trained personnel.

7. A multi-subject survey imposes restriction on the sampling design which of necessity will be a sort of compromise design for getting a common sample. However, when the optimum design for a particular subject is very different from the common design, as it happens sometimes in practice, the advantage of the integrated approach can still be realised by reinforcing the common sample by a special sample in strata which are known to be areas of concentration for a specific subject.

8. Multi-subject surveys are usually based on a stratified multi-stage sampling design. From the point of view of design, it will need to be examined whether the multi-subject enquiries may be made with the same set of ultimate sampling units drawn through a common sample design or whether the ultimate units should be drawn separately for different subjects from a common sample of PSU's. The main advantage of the former will be the flexibility introduced in the tabulation for cross-classifications and improvement in quality of data through cross-checks and consistency checks among different items of information. As against this, there is the risk that the common design may be inefficient or the sample size inadequate for some of the subjects and collection of data on many subjects from the same units may lead to poor quality of data on account of respondent and investigator fatigue. Sampling efficiency will be enhanced by using appropriate frames for selecting different sets of secondary sampling units for different enquiries within a common set of PSU's. For instance, for a survey of land holdings, a list of households operating land classified according to broad sizes of holdings would be a suitable frame within sample villages. Similarly, for survey of small scale manufacture, households engaged in such enterprises could be listed.

9. A self-weighting sample design is often used for selection of ultimate sampling units within PSU's, for facility of tabulation of data in multi-subject surveys. While this design obviates use of weights at the PSU level

in the estimation procedure, it poses problems of uneven work load on the field staff as the number of ultimate units will differ from PSU to PSU. In the NSS, the workload is kept within limits by allowing the field staff to survey a smaller number of ultimate units in special cases. The self-weighting feature is, however, restored at the scrutiny stage by repeating the requisite number of filled-in schedules. The consequent loss in precision would be small since it involves only the secondary stage units. It may also be noted that though canvassing time may be reduced in the field, desk work at the scrutiny stage is increased.

10. When the number of ultimate sampling units to be chosen within a PSU is very small, it will not be possible to achieve a strictly proportionate allocation among different sub-strata. A device used in the NSS in such cases is to arrange the ultimate units sub-stratumwise, but in one continuous serial order for the entire PSU and to draw samples by systematic random sampling.

SAMPLING FRAME

11. One of the main requirements in sampling is the availability of appropriate sampling frames, with information on suitable auxiliary characteristics to permit stratification, if necessary. The decennial population census in India fortunately provides as a by-product a ready frame of small geographical units (villages or urban enumeration blocks) into which the country is divided. The frame is used for sampling for socio-economic and other enquiries. The census frame is, however, found to become out of date within a few years and special efforts are to be made to update the frames especially in the urban sector where peripheral changes take place continually as a result of urban development.

12. In surveys involving geographical units, information on stratum size can be built up from a frame of smaller units like villages or larger units comprising administrative groups of the smaller units (e.g. tehsils). It has been noticed curiously enough that the figures of geographical area and population derived from the two frames may not agree. This discrepancy may be due to a number of causes. The two frames may have been compiled by different agencies or may not refer to exactly the same point of time or the definitions used in the two frames may not be quite identical. Even when the two frames are constructed by the same agency, there may still be some discrepancy; for example, the village population figures may be those of a

preliminary count while the tehsils or district figures refer to the final count. There may also be clerical and printing mistakes in the available frames.

13. India has a unique advantage in having a detailed system of land records in rural areas in the major part of the country which provide an excellent frame for surveys in the field of agricultural statistics, particularly crop surveys. Detailed cultivation accounts are maintained plot by plot and accurate survey maps are available at the village level. In areas where cadastral survey is not made, however, recourse has to be taken to other means such as ad-hoc compilation of lists of plots, cultivators or households. Even where village level records are available it is usually very time consuming to compile a frame of cultivators through the village records, as the actual cultivator of a plot may either not be shown in the records or may be different from the owner of the plot. Nor is a frame of agricultural holdings readily available in India, as a complete census of holdings has not been undertaken so far and there is no uniform system of maintaining upto date lists of holdings. Alternative methods have, therefore, to be adopted for collecting data on agricultural holdings. The frame adopted in the surveys of land holdings carried out by the NSS in 1954-55 and 1960-62, was a list of households stratified according to the size of holding ascertained approximately while compiling the list of households.

14. In the industries sector, a complete frame of industries is available in respect of factories which are registered under the Factories Act. These are the units employing 20 persons or more without using power or 10 persons or more using power. An annual survey of such industries is carried out by the NSS on a complete enumeration basis for larger units and on a sampling basis for the rest. For the unregistered units, however, there is no list readily available, and sample surveys which have been undertaken to cover this sector have used the household approach. Proposals are under way for extending the scope of complete enumeration, at least for listing purposes, to cover units in the non-registered sector employing five or more persons.

15. Systematic collection of statistics regarding distributive trade has been hampered mainly due to non-availability of a complete frame. Though many of the trading establishments are covered under some licensing scheme or the other, or under sales tax administration, the criteria for licensing and the turn-over limits for levy of sales tax vary considerably in different States, thus making a uniform frame difficult. Furthermore, aggregation of results of surveys if carried out in different States raises problems of classification

and comparability. A substantial part of trade belongs to the unorganised sector, not covered under licensing, etc.. Some efforts have been made in the NSS to cover these through the household approach.

16. In some cases problems of classification lead to incomplete frames. For instance, it is difficult to decide if a small establishment is a trading unit or a manufacturing unit as both activities are merged in one unit and an ad-hoc definition has to be adopted. Similarly, in the light of Indian conditions, restaurants used to be classified under services rather than trade or manufacturing, though in conformity with the revised system of national accounts, these are now grouped along with trade.

DATA COLLECTION

17. In large scale surveys where the work of data collection is spread over a vast area and involves a large number of investigators, some steps in the procedure of selection of samples have necessarily to be left to the field staff, particularly when the frames are to be compiled for the secondary and subsequent stages of sampling. The criteria for identifying the units eligible for sampling and the selection procedures have, therefore, to be particularly simple, sometimes even at the risk of reduced statistical efficiency. For instance, in the annual crop surveys, selection of fields within a village for a particular crop is generally made by first selecting the requisite number of fields from all fields in the village instead of making a separate list of plots growing the crop. If any of the selected fields is not found to grow the crop in question, it is replaced by the next field growing the crop (following a serial order). This procedure will make the probabilities of selection unequal for different fields (especially when the crop is sparsely grown), which fact is usually ignored in the estimation procedure.

18. To maintain a good standard of methodology, it will be necessary to give intensive training to the field staff entrusted with data collection. Training of large numbers of staff involves a good deal of organisation and will have to be phased over more than one level. It is necessary to make special efforts to ensure that the prescribed concepts and field procedures are clearly understood by the investigating staff. This becomes particularly important in socio-economic enquiries by the interview method, where the field staff have to elicit information from informants, a majority of whom are illiterate.

19. One major difficulty in canvassing information by enquiry is the fact that the questionnaire approach with a list of standardised questions will not be very useful where many of the respondents are illiterate. The schedule approach is, therefore, preferred and the investigator has to develop his own technique to elicit the requisite information. This becomes a great burden on the investigator particularly when he has to master the concepts and definitions relating to a number of subjects.

20. Supervision of data collection is indispensable in assessing and controlling non-sampling errors and adequate arrangements should, therefore, be made for efficient supervision. The ratio of supervisory to primary staff has to be fairly high in the developing countries, as on account of inadequate facilities, a considerable time is spent on traveling and communication, thereby reducing effective inspection time. The supervision should also be so organised on a random sub-sample of units that the data for inspected units could be subjected to a rigorous analysis with a view to drawing conclusions about the extent of error and calibrating the survey results if necessary. This aspect of rationalised supervision on a statistical basis has not probably received the attention it deserves.

21. Both for reasons of economy and closer understanding of the subject matter field, data collection is sometimes entrusted to investigators drawn from the concerned agencies (generally belonging to government departments) and they are required to work on surveys on part-time basis, in addition to their regular departmental duties. The technical control for the survey may, however, vest in a different organisation (e.g., the State Statistical Bureau). This arrangement often proves to be unsatisfactory as supervision becomes ineffective.

22. Collection of current data should, as far as possible, be supported by actual physical observation or measurement, whereas retrospective information will have to be elicited by interviewing informants or consulting records. In India the bulk of the people, particularly in rural areas, operate at a subsistence level and detailed records or accounts are not maintained. Farmers conduct transactions partly in kind. Decisions on agricultural operations are not necessarily based on economic considerations. Farmers are often not cognisant of the quantitative aspects of the production process, such as daily yield of milk or the cost of feeding the stock. Consequently, in order to minimise serious non-sampling errors, considerable attention has to be paid to the collection or at least verification of data by direct physical observation by trained staff, using appropriate sampling

procedures, such as double sampling. Thus land use has to be verified by measurement, crop yields have to be observed through crop cutting experiments involving actual weighment of produce, milk yields have to be noted by actual weighment and investigators have to be stationed among selected cultivators to maintain farm accounts in cost of cultivation studies.

23. In surveys where information is obtained through oral enquiry of respondents, the schedules have to be drawn up carefully to achieve a balance between the advantages of maximum detail and the disadvantages of a long and unwieldy schedule which is a strain on both the investigator and the respondent. In the context of multi-subject household surveys, it is to be considered whether individual subject schedules which give subject information in depth should be canvassed from different sets of households or whether essential information on all aspects of the household economy should be collected through an integrated schedule canvassed for each selected household. The latter is advocated for providing internal consistency checks and for ensuring better quality data on the whole. The integrated approach also introduces flexibility in tabulation through suitable cross-classification of results.

24. Pre-testing of survey schedules is an important exercise which must precede the finalisation of survey schedules, concepts and procedures. This aspect, however, does not appear to receive due emphasis under the plea of lack of time or paucity of resources.

25. Securing the cooperation of the respondents for eliciting information by the interview method and of the farmers in case of agricultural surveys fortunately does not seem to present a serious difficulty in India. The reliability of information is, however, likely to be affected by the fact that in rural areas, investigation has to be made in the presence of a number of villagers who invariably collect at the spot and the informant may not feel free to give correct details. The generally low level of literacy also adds to the difficulties of eliciting information.

26. The choice of an appropriate reference period has been a problem from the point of view of collection of realistic and meaningful data. Thus in urban areas data on consumption expenditure could perhaps be ascertained monthly, as most people would be on jobs carrying monthly salaries, but the rural population might be able to give information by agricultural crop seasons only. Some items of expenditure are incurred on an annual basis. In a multi-subject survey, adoption of

different reference periods for different items may prove an undue strain in investigation. Longer reference periods are known to result in recall-lapse in getting information on vital events. Experience has shown that in general a moving reference period of the appropriate duration is preferable to a fixed reference period.

COMPILATION AND ANALYSIS

27. The processes of compilation and analysis have to be preceded by a thorough scrutiny of the data collected in the field. In large scale surveys and multi-subject enquiries, the data are so voluminous that a question might be raised as to how best the scrutiny should be arranged. In case the processing is done centrally, problems of communication with the field might necessitate a field scrutiny in regional centres to remove the inconsistencies before the data are passed on to the processing unit. The processing unit has also to undertake its own scrutiny of data before they are accepted for processing. A general question might, therefore, be posed as to whether there is a necessity for such double scrutiny and whether such scrutiny should be 100 per cent or on a sample basis. The level of staff who should do the scrutiny and the need for a standardised scrutiny programme are also relevant issues to be considered.

28. A common problem faced at the processing stage relates to missing entries. If some items are left blank by the investigating staff in a schedule having many items, it becomes difficult to get the requisite information from the field. The difficulty will be greater if the commencement of processing is delayed. It may be worthwhile to study the effect of alternative actions which can be suggested in such cases on the cost of tabulation as well as on the efficiency of estimation. A parallel problem arises when some of the figures entered in the schedules appear to be doubtful or abnormal. There is no clear cut basis for deciding as to which figures are abnormal. Similarly, possible effect of the action to be taken when there are non-responding units is not always clear for different situations and sampling designs. A question can be raised as to whether coding should be left to the field staff or done at the processing stage; whether quantitative data for classification should be obtained in terms of codes for pre-specified classes; whether it is a good procedure to ask the investigating staff to record answers to questions by checking one of the alternative box spaces provided; how the different procedures would affect accuracy of data and cost of collection.

29. For speedy tabulation of voluminous data, mechanical data processing is preferred. Over the last decade, mechanical tabulation has been increasingly used. But there

has appeared some imbalance of resources for various stages of tabulation with the result that the completion of tabulation gets delayed. Attempts are being made to sort out this difficulty which is essentially a problem of resources.

30. In multi-subject enquiries, mechanical tabulation is usually resorted to for detailed cross-classification of data and as a matter of course, all items are included in the programme for the mechanical tabulation. However, there may be some items which can be tabulated more economically and expeditiously through manual tabulation (such as items which require only a single effective run of punched cards). It is felt that it may be possible to save substantial machine resources and tabulation time by diverting such items to manual tabulation. The decision to process manually the individual slips for rural persons in the 1971 Population Census is an instance in point.

31. Due to inadequacy or imbalance of resources, arrears of tabulation tend to accumulate. Priorities have to be disturbed in order to process data of more recent surveys. Dislocation of priorities also gives rise to problems of proper scrutiny of tabulated results. An important aspect of data collection and tabulation which affects the progress of processing deserves mention here. In large scale surveys which are repeated year after year, in order to facilitate the building up of a time series, data are collected every year on a large number of items and a set of identical tables are generated. Where paucity of tabulation resources is felt acutely, it should be a matter for serious consideration whether there is a genuine need for continuing enquiries for all items. Apart from some basic tables required for constructing the time series, other tables to be generated could possibly be suitably phased over different years so that a different set of tables might be produced each year, thus reducing tabulation requirements to manageable levels.

32. Data collected in large scale surveys are voluminous and it takes considerable time to bring out all the results envisaged in the tabulation programme, especially when the resources are in short supply. The user (generally the administration) would, however, need quick information on some of the key items for policy decisions. Much of the data would also lose their value unless the results could become available without delay. There is need, therefore, for a special programme of quick tabulation for extracting in advance summary information on selected items of importance, on a sub-sample basis if need be.

33. For general improvement in collection techniques as well as quality of statistics, it is essential that the data collected should invariably be subjected to economic analysis and the results of analysis fed back to the planning agencies. However, inadequacy of tabulation resources has generally come in the way of detailed analysis of survey data. Use of electronic data processing equipment which will facilitate complex types of statistical analysis, especially for research and development of techniques, though increasing, is as yet limited. The existing conventional unit record machines are being found to be increasingly inadequate. Replacement of old machines present considerable difficulty. Staff resistance to automation is also a complex problem.

34. To save time, routine statistical analysis is often reduced to the simplest form. Decisions have to be taken on the stages of computation which can be given up in order to simplify the operations without affecting the overall results. Analysis of variance is made only periodically in some established surveys like crop estimation surveys. It seems desirable to compute the conventional analysis of variance and estimate the errors at least once in a few years. The design of independent and interpenetrating sub-samples (due to Mahalanobis) which is adopted in the NSS, permits easy calculation of relative variances based on sub-sample variations, irrespective of complications in sampling and estimation procedures. When the sub-samples are surveyed and processed by different sets of field processing staff, a broad check on the quality of data becomes available.

35. Mention must also be made of the fact that the importance of graphical analysis does not seem to be stressed adequately in all surveys. In this connection a reference might be made to the fractile graphical method of analysis of NSS data developed by Mahalanobis⁽¹⁾. The use of concentration curves to study the distribution of income, expenditure, landholdings, and similar characteristics over various sections of the population is another feature of graphical analysis deserving special mention, on which aspect some interesting results were presented by Mahalanobis and Lahiri⁽²⁾.

(1) P.C. Mahalanobis (1960) "A method of fractile Graphical Analysis", *Econometrica*, Vol. 28, 2

(2) P.C. Mahalanobis and D.B. Lahiri (1961) "Analysis of Errors in Censuses and Surveys with special reference to experience in India", *Bulletin of International Statistical Institute*, Vol. 38, 2

36. It often happens in large scale surveys that the results reveal certain unusual features or apparent inconsistencies. For a proper interpretation of such results, a few supplementary case studies or type studies would be very useful. The need and importance of such studies which will permit a deeper probe into the results cannot be too strongly emphasised. However, in practice, this aspect tends to get ignored.

14(f) - NEW ZEALAND

THE PROBLEM OF MAINTAINING HIGH STANDARDS
OF SURVEY METHODOLOGY ANALYSIS
AND COMPILATION TECHNIQUES

SOME PROBLEMS BEARING ON THE MAINTENANCE
OF THE STANDARDS OF ITS STATISTICAL
SERVICE BEING FACED BY THE
DEPARTMENT OF STATISTICS

Introduction

1. Every central statistical office faces problems in maintaining satisfactory standards in the provision of statistical services, these problems being generally classifiable as either managerial or statistical. Some problems are of a type which are met with generally by all statistical offices, irrespective of the resources available in relation to the economic and social populations under survey or the range of statistical services being provided, and some are specific to the office under consideration. Rather than attempt a comprehensive review of all types of problems of maintaining the standards of statistical services, this paper confines itself to a brief discussion of the special problems being faced by the Department of Statistics in keeping its statistical services up to the standard demanded for policy review and decision-making purposes of central government, the indicative planning organisation,* and other organisations justifying official statistical servicing.

Characteristics of the New Zealand Official Statistical System

2. An appreciation of the problems being faced in New Zealand requires, as a preliminary, a brief explanation of the type of statistical service currently being provided.

3. Official statistical services within the New Zealand Government are organisationally centralised within the Department of Statistics to the extent that between 80 and 90 percent of all such series are handled by the department. In spite of the smallness of the economy (total population 2.8 million) and its relatively high dependence on overseas trading, with exports dominated by a narrow range of primary products, a high standard of living is experienced. The institutional framework regulating

* See the New Zealand paper on provisional agenda item 5(a).

political and social behaviour of the community is well-developed along the western pattern and a large measure of awareness exists of the utility of quantitative statistical descriptions and analyses of economic and social affairs as a required information base for public decision-making.

4. The demand for statistical services has led, over the years since official statistical work commenced in the middle of the nineteenth century, to the rather unco-ordinated development of a wide range of subject-matter surveys and analyses, supplemented with a range of basic national accounting statistics to give the broader macro description of economic performance. The regular preparation of a wide range of subject-matter statistics has posed major statistical and managerial problems in the past for the department because of its relatively small staff. In spite of significant increases over the last 20 years the current staff establishment (including all classes of machinist and statistical coding staff, as well as professional staff such as survey statisticians, economic statisticians and theoretical statisticians) numbers only 400.

5. One aspect of the administrative solution of these problems worth noting has been the necessity among trained statistical staff of avoiding over-specialisation in the range of statistical work being handled. This has had both advantages and disadvantages, of course, the chief advantage being the resulting versatility of staff and the wider statistical perspectives which are developed. However, as discussed in the next section, the stage has been reached where the disadvantages now greatly outweigh the advantages.

Resources required for current Statistical Developments

6. The New Zealand paper on provisional agenda item 5(a) refers to the 1969 review of the adequacy of the present statistical system in relation to the information needs of the indicative national planning machinery and also of central government. The report of the Technical Committee on Statistics included a schedule of major new statistical developments which need to be implemented if the inadequacies are to be eliminated. These developments include the establishment of an integrated register of all economic enterprises and the conduct of all economic statistical surveys in terms of this frame; the setting up of an organised file of statistical data, both aggregated and unaggregated, in computer storage, so as to permit ready retrieval of selected data and the linking of cross-source data; the introduction of a comprehensive set of national accounts based on the United Nations System of National Accounts; the introduction of a system of fine geographical referencing of elementary units of statistical

14(f) - NEW ZEALAND

data as the basis for both achieving consistency in areal analyses in different subject-matter surveys and meeting the demand for small-area statistical analyses; and the automation of routine survey clerical work, such as questionnaire addressing for mail surveys, the recording of survey response and the control of follow-up action.

7. These and other needed statistical extensions are, in the main, such as to require a fairly substantial minimum complement of professional staff in order to achieve any significant progress towards implementation and herein lies a problem which did not occur to the same extent with the typical subject-matter survey which formed the key element in a developed statistical system up until recently. With such surveys the "variable-resource" staff requirement conditional on the size of population under survey tended to greatly outweigh the "fixed-resource" staff requirement independent of survey size. This fact allowed the smaller central statistical offices to provide a statistical system, not without considerable managerial and statistical effort of course, fully comparable with that given by larger offices. A small nucleus of short-supply professional statisticians could handle survey designs and the variable-staff requirement demanded by the operation of the surveys, did not have to be so intensively trained. Coupled with the versatility of its key professional staff engendered by the provision of a comprehensive, if relatively unco-ordinated, statistical service this fact permitted a central statistical office like the Department of Statistics to match fully the range of statistical services of larger offices.

8. With the introduction of computer processing into statistical offices and the statistical extensions which have directly or indirectly followed, the situation has gradually become more pronounced where much greater specialisation in professional staff has been necessary and a considerable increase in man-years of professional staffing input is required to carry out minimum development work for a new application. For example the staff resources required to devise the clerical and computer system for a computerised statistical data base is generally not a function of the volume of data generated by surveys, but the range of surveys to be accommodated. The variable resource requirement is now to a large extent the computer central processor and storage capacity, rather than the former less intensively-trained survey staff. The recruitment of the required core of professional statistical staff is the crux of the problem here, as in other examples which could be elaborated. The smaller the country the smaller the overall pool of such persons.

9. This fact has been recognised by Government in New Zealand in the magnitude of extra staffing resources to be allocated to implement the recommendations of the Technical Committee on Statistics, and the revised statistician salary scales introduced. It is doubtful, however, even with the extra staff and the provision of a higher proportion of professional statistical staff that the problem will be easily overcome. Unfortunately, in the case of some desirable extensions there has been a lack of even initial development work, because of the shortage of professional statistical staff. Widespread recognition of the situation now exists and there is a strong determination to catch up on development work so as to provide a statistical service of a standard that latest statistical technology permits. A major managerial and statistical challenge is therefore facing the official statistical system in New Zealand at the present time.

10. A significant factor in meeting the challenge will undoubtedly be the use of the technical statistical work on statistical standards done by statistical offices of the various international agencies, the participation in international statistical conferences and seminars, and exchange of experience with other countries facing similar problems or initially more advanced in development work. This course will obviously avoid duplication of theoretical and practical development work on points of fundamentals, even if New Zealand conditions require special modifications of general standards. This applies to such widely varying exercises as the introduction of the United Nations SNA and the setting up of a computerised statistical data file.

11. On the other hand the development programme on which we are embarking will undoubtedly lead to methods that will have some interest for other countries faced with similar problems related to limited absolute statistical resources for official statistics.

Effective Use of Computer Processing

12. During nine years' experience of the use of computer processing in the Department of Statistics new type problems have had to be faced by survey statisticians. Such problems are related to the frequent communication gap between professional computer staff and those statistical staff whose work is dependent on computer servicing, due to lack of knowledge of the other side's technology and requirements.

13. Any organisation using computer processing must give considerable emphasis to education for its staff in computer concepts and the implications that computer methods have in

14(f) - NEW ZEALAND

regard to changed work procedures in clerical processing systems which interlock with computer systems. Experience indicates, however, that, short of a large scale regular interchange of computer and non-computer staff (seldom achievable in practice), adequate communication can be a continuing problem.

14. Consequently, when statisticians are not able to participate in the computer systems design stage of setting up a survey processing scheme, the devising of computer systems that allow statistical staff adequate control over the accuracy standards of survey processing is a matter of great difficulty. The use of statistically trained computer systems planners who act as intermediaries between statisticians and computer systems and programmers is a course currently being implemented by the Department of Statistics now that computer processing is being converted from an internal computer installation to a government computer centre. In the past the difficulty of imparting the wide range of training required, the administrative difficulty of transferring key staff and the reluctance for staff to move their work area have been obstacles. The use of ad hoc planning and systems committees representative of all parties concerned is another solution but is often very unwieldy and slow-acting.

15. As well as in the area of setting the optimum depth of statistical processing in relation to the technical and economic characteristics of the computer installation to be employed, the lack of computer expertise possessed by survey statisticians (and vice versa) has led on occasions to difficulties in formulating comprehensive checks on the completeness and accuracy of survey input data (after transference to a computer input medium) which is the basis for the processed survey statistics. It is not unusual for data errors at this stage to produce consistent, seemingly error-free final processed statistics. Even enhanced "statistical sense" checks applied to processed results can fail to detect data input errors based on input data gaps or duplications. The critical importance of incorporating every check procedure at this phase of the computer system, including independently producing unit-record counts, is one of the lessons emerging out of nine years of the duplication of computer methods in the department.

16. The important positive contribution that programmed validity and consistency checks can make to insure error-free unit records for computer processing of survey results is now widely understood. Substantial relaxation of clerical checking of primary survey data is quite possible once experience has indicated that the edit programme incorporates a sufficiently comprehensive range of checks. The devising of these checks, of course, again depends on the ability of the survey statistician and the system analyst/programmer to communicate effectively.

17. Problems of maintaining statistical standards in relation to computer processing can assume a new character when the development of computer systems and computer processing for a central statistical office are handled by a separate computer centre servicing a number of organisations. This is currently a problem facing the Department of Statistics. The establishment of a specialist statistical control section answerable to the central statistical office but located in the computer centre to insure satisfactory scheduling, preparation of computer run instructions, checking of input and output data, and enforcing procedures designed to insure maintenance of confidentiality guarantees given to survey respondents, is essential.

Confidentiality Guarantees and Statistical Standards

18. It is accepted by official statisticians that the granting to survey respondents of guarantees that confidentiality of supplied data will be maintained is a key factor in achieving co-operation and the desirable level of accuracy in reported data. In addition it is important that not only must confidentiality guarantees be given and adhered to but they must publicly seem to be adhered to. It is essential that the central statistical office avoid procedures that might give the impression (wrongly) that individually supplied data are being made available outside the statistical office.

19. With the wider recognition of the value of management information systems and the consequent demand for as much detail as possible in officially provided statistics, the experience of the Department of Statistics is that increased pressure is being brought to bear to release details of survey frames (names and addresses) with basic classifications of survey units (form-type, industry classification, etc.) as a basis for outside surveys, whether of a statistical or non-statistical nature. Coupled with a certain degree of non-statistical scepticism about the importance of the confidentiality issue this pressure is causing the official statistical system in New Zealand to currently re-examine to what extent confidentiality procedures are in need of revision. While nothing can be done which would lead to a deterioration in statistical standards due to loss of public confidence the extent to which certain classes of basic information obtained from survey respondents might be released selectively, say to appropriate administrative departments of government (e.g. Department of Agriculture in the case of statistical data supplied by farmers) is in question.

14(f) - NEW ZEALAND

20. A detailed discussion of the statistical, administrative and legal factors involved cannot be presented here, but the increasing importance of this problem to the department is illustrated by the fact that refusal to release basic reference data from surveys can lead to a demand for the transfer of statistical subject-matter sections to the relevant administrative department, with the possibility of statistical inefficiencies resulting. The aim of organisers of regional or special-purpose computerised data bases to adopt classifications of business and social units consistent with those used by the central statistical office is another example where pressure for release of what is currently regarded as confidential data is mounting. The final solution to this problem has yet to be formulated in the New Zealand setting.

Proposed All Nigeria Urban Household Budget Survey

Background:

1. The first Urban Consumer Surveys in Nigeria were conducted in 1953 in Lagos, Ibadan, and Enugu. Although at various times since then, consumer surveys have been conducted in some eighteen or more towns throughout the country it has not been possible to produce a national cost of living index because the towns covered were not included as units in a national survey but as individual urban areas. Consequently what we have is consumer price indices for individual towns or groups of towns. Because of the restricted coverage of previous surveys and lack of simultaneity in their execution, the surveys have not been as useful as they could have been. This Household Budget Survey is being planned to avoid these defects.

Purpose and Coverage:

2. The main purpose of the survey is to provide information about the patterns of expenditure required for the weights in the construction of an urban consumer index. It will also provide as full information as possible on levels and standards of living and provide data required to estimate and analyse income, expenditure, and savings. In research and economic studies, the survey will equally yield data for the calculation of say demand elasticities, and in the estimation of national income and especially the final consumption components of the expenditure on gross domestic product.

3 It is planned that the sample should be sufficient to allow publication of separate results for :

- (a) The whole country;
- (b) Each one of the state capitals;
- (c) Each of the states of the Federation and
- (d) The whole of the rest of the urban sector in each state.

As there are 12 states, this implies 24 distinct reporting domains. Further, as the Western State has a very much larger Urban population than others it is felt that a third reporting domain should be created in this state consisting of towns of over 100,000 population (other than the State Capital, Ibadan) in the 1963 Census. This, in effect, means 25 reporting domains on the whole.

4. It is clear that a Survey of income and expenditure pattern of the Community without relating it to specific income group or groups, will certainly yield misleading results. In the circumstances, it is suggested that one index would be required for each relevant income group in each of the reporting domains. The "middle income" group would be represented by households whose head has a declared annual income ranging between £N450 and £N1,200 (corresponding roughly to the professional, executive, managerial class), while the "lower income" group would be those households of which the declared income of the head falls below £N450. Probably the middle income group would not be sufficiently numerous outside the State's Capitals to enable results to be published for this group in the "other Urban" domain of a state. An additional index would, therefore, be published for the whole urban sector in each State for each income group, and here the small number of middle income households found in "other urban" sector of the state would be included in the total. The arrangement is shown

schematically below:

<u>State Capital</u>	<u>Other Urban</u>	<u>Whole State, Urban</u>
Middle Income A	B(not published)	Average A, B
Lower Income C	D	Average C, D

A, B, C, D are the four indices for the appropriate income groups and reporting domains. The averages would, of course, be appropriately weighted to take account of the frequency of the households in the different categories. A national urban index would also be published for each income group by weighting the State indices by the State urban population.

Sample Design:

5. The basic design is a two-stage sample, stratified at the second stage with unequal sampling fraction in the strata. In the first stage, Enumeration Areas (E.As) will be selected. A reliable frame of Enumeration Areas is not available at the moment. However, the Federal Office of Statistics is currently planning to demarcate the whole Country into mutually exclusive enumeration areas, each consisting of about 500 people. The Survey cannot, therefore, start until the completion of E.A. creation in the urban sector. The second stage selection will require the listing of the households in each sampled E.A. since there is no available document embodying such households. This is then followed by the stratification of the listed households into three distinct strata according to the occupation of the heads of the households.

<u>Stratum</u>	<u>Occupation of the household head</u>
I	Professional, executive, managerial
II	Other employed persons
III	Self-employed and others

Sample Size:

6. In any multi-stage sample survey such as the one being planned the sample size depends on the desired level of precision as well as available resources. In terms of precision, the 1960 Urban Consumer Survey in Lagos shows that the estimates derived from it have relative errors less than 5 per cent for a sample of 500 households. It follows therefore that in each reporting domain we need at least 500 households to achieve a 5 per cent precision for most of the estimates.

7. Determination of the first stage sample size is very difficult since no estimates of the between and within Enumeration Areas variances are available. However, from operational point of view, it is thought that in each reporting domain a team of 4 enumerators will be able to cover 36 Enumeration Areas each consisting of 5 selected households in a year. It follows, therefore, that in order to cover not less than 500 households in any reporting domain, we need at least 3 teams. Thus in each State Capital (except Lagos and Ibadan) the sample size will be 3 teams X 36 blocks X 5 households = 540 households among 108 Enumeration Areas. In Lagos, we hope to use 6 teams; in Ibadan, 5. Thus total sample size for the State Capitals will be:-

Lagos	1,080 households
Ibadan	900 "
Other State Capitals	540 " each
Total Sample Size is therefore 7,380			
Total Enumerator - years =			164
Total Supervisor - years =			41

8. Sample size in the other - towns stratum will be the same as in the standard state - capital sample i.e. 3 team-year covering 540 household-months, in each state. As far as possible, in each selected E.A. the following sample should be selected:

2 households from stratum	I
2 " " "	II
1 household " "	III

If it is not possible to find more than 1 household for stratum 1, then the numbers to be taken will be 0, 3, 2. If it is not possible to find enough households for stratum II, then the number is made up to 5 by taking more from stratum III. Thus the sample will always be 5 per E.A. This will lead to sampling fractions very roughly in the ration 5: 2: 1 in the 3 strata respectively. The total sample size in the other-towns strata will be

Western State I	540 households
Western State II	540 "
Other States	540 households each of the 11 states.
Total sample size	=	7,020 households	
Total Enumerator-years	=	156	
Total Supervisor-years	=	39	

Field Organization

9. It is important to note here that occasional surveys can no longer meet the need for the essential information required; so vital sample surveys of this nature call for a more permanent organization and continuous operation for fieldwork. Survey experience over a long period no doubt generates expertise which continues to be available on a continuous basis. It is with this in view that the Federal Office of Statistics has just created a separate unit charged with the organization and control of the fieldwork. With the present structural set up of the office, therefore, the entire organization of data collection falls within the ambit of the schedule of the Surveys Field Organization of the office. This unit controls a large field staff or well over 400 (for both rural and urban surveys) stationed in various parts of the country.

10. Data are to be obtained partly by interview and partly from records kept by each head of household. One interviewer is to be allocated to each block, for a period of 35 days during which period he/she makes daily visits to each sampled household. The first five days, of course, will not be

taken into account in the analysis as experience shows that figures collected during such period are not usually reliable.

11. There are two constraints on the selection of E.As. in the Other Urban sector. If they are selected without regard to their situation in the particular town, there would often be no more than one enumerator in a town so that supervision would be hampered. Thus E.As. would be clustered in multiples of 3 in each town, so that the team will not be split between more than one town. At the same time, it would often be necessary to move the team from one town to another every month. This, no doubt, will present a problem of accommodation and it is felt that such conditions would be intolerable for enumerators. We have, therefore, assumed that, for the convenience of enumerators they must be assured of a stay of at least 6 months in each town. One team will cover 18 sample blocks in this period, or 90 sample households. Thus we assume that the selection of E.As. will be such that a multiple of 18 E.As. is selected in each town. Since the total No. of E.As. to be selected is 108 (i.e. 540 divided by 5) there is no doubt that this will severely restrict the efficiency of the sample in many states. However, there seems to be no alternative. Towns will therefore, be selected by systematic sampling with probability proportional to the 1963 Census Population of the town, and each selected town will be allocated a sample of 18 E.A. (If a town is selected twice it will have 36 sample E.As. etc.) In the selected towns the required number of E.As. is then selected exactly as in the state - capital stratum.

The survey will be a quinquennial enquiry and the first in the series is scheduled for 1974/75 by which time complete normal conditions and consumption patterns will have been fully re-established in the Country.

Costing

12. The cost of most Socio - economic sample surveys depends principally on the sampling design as well as on the cost of hiring or acquiring the human and other resources required for their execution. To this end, alternative sampling designs have also been proposed bearing in mind the desired precision to be achieved.

13. The first alternative is a design that will allow for the calculation of the following indices.

- | | | |
|-------|--|-------------------|
| (i) | National Index | with 1% precision |
| (ii) | State Index | with 3% precision |
| (iii) | State Capital Index | with 5% precision |
| (iv) | Other towns in the State with 5% precision | |

In this case, each State (except Western State) is divided into two strata:-

- (a) State Capital
- (b) Other towns in the state

Western State will, of course, be divided into three strata viz:-

- (a) The State Capital
- (b) All towns with population over 100,000 minus State

Capital

(c) Towns with population less than 100,000

There will, therefore, be 25 strata in all. At second stage selection, 540 households are to be randomly chosen in each stratum (except Western and Lagos State Capitals, in which 900 and 1,000 households are selected respectively). It is estimated that the total cost of this operation will amount to £N105,000.

14. The next alternative would be to report a single figure for each state, without distinguishing the capital from other towns. Here, a state forms a stratum. 540 households are to be selected in each stratum (except West and Lagos where 900 and 1,080 households would be randomly selected). This implies a design to allow for the calculation of the following indices:-

- (i) National index with 1.5% precision
- (ii) State index with 5% precision

15. A third possibility still would be to limit the enquiry to state capitals only. In this case each state capital will form a stratum and all other towns in the country are to be lumped together into one stratum. This means a design to allow for the calculation of the following indices:-

- (i) National index with 1.5% precision
- (ii) State Capital index with 5% precision

The cost is estimated to be about £N55,000.

16. Then the last scheme, which is the one the Federal Office of Statistics is about to implement will provide the same information as obtainable from the first at a lower but still acceptable precision. The design is expected to allow for the calculation of the following indices:-

- (i) National index with 1.5% precision
- (ii) State index with 5% precision
- (iii) State Capital index with 6% precision
- (iv) Other towns in State index with 9% precision

Stratification as already described is just the same as we have in the first alternative except that in each state capital, 360 (instead of the usual 540) households are selected (540 in Ibadan, 720 in Lagos) and 180 households in each of the other strata. The cost is estimated at about £N70,000.

17. This paper has only attempted to outline the tentative plan for its first nationwide household budget survey for 1974/75. It is hoped that by that time another national population census would have taken place and that the population would have settled down after the mass movements of population during the civil war. The gap between now and 1974/75 will also allow enough time for the consumption patterns to return to normal in the war affected areas of the Federation. The sampling scheme will be reviewed from time to time before the survey year in order to take account of any significant shifts in the population.

A NOTE ON
DISPARITIES IN THE LEVEL OF LIVING

By P. C. MAHALANOBIS

1. *Levels of living of households*: A good deal of information on household consumption expenditure was collected by comparable methods in the National Sample Survey (NSS) in the form of 'rounds' of survey, year by year, for many years. The aggregate value of consumption per person¹ of a household may be used as an approximate indicator of the level of living of the household. For any given round of the NSS, it is possible to arrange the sample households in ascending order of the aggregate value of the total per capita consumption expenditure per month²; such ordering would correspond approximately to increasing levels of living. The sample households can then be divided into a suitable number of groups of equal size³ from the bottom. If, say, ten equal groups each of ten per cent (i. e., ten decile groups) are used then the bottom ten per cent would represent the poorest section of the people, and the top ten per cent the richest households. A comparison of the consumption in physical terms, or in money terms, of particular items, or of all items of goods or services, for the ten different decile groups would supply information about disparities in the level of living between the ten decile groups.

2. *Comparison over time*: Detailed information on the consumption of a number of goods and services has been tabulated for seven rounds of the National Sample Survey. It is possible, in principle, to examine whether there has been any change, over time, in the distribution by size of particular items of consumption, during this period. In the case of expenditure on consumer goods and services, such comparisons over time would be meaningful only if all prices had remained the same or had changed in the same proportion. In fact, prices had changed continually. Conversion of all these prices into a standard constant price involves great technical difficulties; adequate information for this purpose is also not available. No attempt has, therefore, been made to study the distribution by size of expenditure at constant prices. One way of getting over the difficulty, at least partially, of making adjustments for changes in prices would be to consider, not the *actual* expenditure of different fractile groups but, the *relative share* of expenditure.

3. *Concentration curves*: By plotting on a graph paper the cumulative percentage shares of consumer expenditure against successive decile groups, it is possible to obtain the concentration curve for the distribution of consumer expenditure over fractile groups. If the distribution is equal, that is, in the purely

¹ Or, in ascending order of aggregate consumption expenditure per household as a whole. Results valid for the two types of ranking would be valid for ranking by any scale of adult equivalents.

² In the National Sample Survey estimates are usually given for a 'standard month of 30 days'.

³ That is, containing an equal number of persons of the estimated population; such groups of equal size are called 'fractile groups', in the terminology used in Fractile Graphical Analysis (FGA).

hypothetical case when each person (or household) has exactly the same per capita expenditure, the concentration curve would be the same as the diagonal line at an angle of 45° with both the axes. In this case, if we continue to add the successive consumption expenditure of each person (or household), the accumulated sum of such expenditure would be exactly the expenditure of each person (or household), multiplied by the number of persons (or households). Therefore, the cumulative percentage share of total expenditure would be the same as the cumulative percentage of persons; all the points would then fall on the diagonal line mentioned above. The nearer the observed concentration curve is to this diagonal line of equal distribution, the greater will be the equality of distribution. Also, the farther away the observed concentration curve is from the diagonal line, the greater will be the inequality of distribution

4. *Rise in the price of cereals* : The price of cereals was at its lowest level in rounds 8 and 9, and at the highest level in rounds 15 and 16. Relevant information on the respective price-index number, period of survey, and the number of households (with about half the number in each sub-sample) which were covered in the survey in each of NSS rounds 8, 9, 15 and 16, are given in Statement (1) below.

STATEMENT (1) : NSS ROUNDS : PERIOD OF SURVEY, PRICE INDEX OF CEREALS, AND NUMBER OF HOUSEHOLDS IN COMBINED SAMPLE

NSS round	price index of cereals 1952-53 price=100	period of survey	number of households
(1)	(2)	(3)	(4)
8	77.4	July 1954—March 1955	1,833
9	73.3	May 1955—November 1955	1,600
15	105.6	July 1959—June 1960	7,144
16	103.2	July 1960—June 1961	3,703

It would be seen that price index numbers for cereals were very low, 77.4 and 73.3 (with 1952-53 price = 100), in rounds 8 and 9; the index number increased to 105.6 in 1959-60, and was slightly lower, 103.2, in 1960-61. The general increase in prices of cereals was about 45 per cent between 1954-55 and 1959-61. The period covered by survey rounds 8 and 9 with low prices, on one hand, and the period covered by rounds 15 and 16 with high prices on the other, therefore, supply suitable data for the study of the effect of increase in prices on the consumption of cereals.

5. Index numbers of average prices of cereals in rounds 15 and 16, taken together, with corresponding prices in rounds 8 and 9, taken together, as base (=100) are shown in Statement (2). Decile groups are shown in col. (1); and corresponding index numbers are given, separately, for sub-sample 1 (s.s.1), the combined sample¹ and sub-sample 2 (s.s. 2), in cols. (2), (3) and (4), respectively, in Statement (2).

¹ The information in the National Sample Survey was collected in the form of at least two (and sometimes, more) interpenetrating net-work of samples (IPNS). These two samples are statistically independent and equivalent, and are called the two sub-samples (s.s. 1 and s.s. 2); pooling them together one gets the combined sample that is, the sample as a whole.

8(b) - INDIA

Differences between the two estimates for sub-sample 1 and sub-sample 2 supply information on the margin of uncertainty of the estimates for the combined sample inclusive of both sampling and non-sampling errors. The index-numbers are shown in a graphical form in Chart 1. It can be seen from the bottom line of Statement (2), that the index number for the whole (that is, 0 to 100 per cent) of the population was 145 showing that the general rise in prices of cereals was 45 per cent. For the lowest decile, or bottom 10 per cent of the population, index numbers, as shown in line 1, were of the order of 180 or 190, indicating that the price of cereals increased by 80 or 90 per cent in the case of the poorest section of the population. Index numbers

STATEMENT (2) : INDEX NUMBER OF AVERAGE PRICE
OF CEREALS IN NSS ROUNDS (15, 16) WITH PRICE
IN ROUNDS (8, 9) AS BASE=100, BY FRACTILE
GROUPS : ALL INDIA RURAL

fractile group (per cent)	price index		
	s.s. 1	combined	s.s. 2
(1)	(2)	(3)	(4)
0 — 10	178	191	183
10 — 20	165	162	159
20 — 30	158	154	158
30 — 40	161	151	158
40 — 50	154	142	149
50 — 60	143	141	144
60 — 70	138	149	145
70 — 80	138	139	139
80 — 90	135	141	137
90 — 100	123	137	129
0 — 100	145	145	145

for the second, third, and fourth deciles were appreciably lower; and the rise in prices was of the order of somewhat less than 60 per cent. For the fifth decile (line 5) the rise was of the order of 50 per cent. The percentage increase of prices gradually dropped in higher deciles, and was of the order of only 30 per cent in the tenth decile for the top 10 per cent of the population. The increase in the price of cereals was thus relatively much greater for the poorer sections of the people.

6. In order to maintain the consumption of cereals at the same physical level, the increase in expenditure on cereals of the poorer households would have to be relatively much larger than the increase of expenditure on cereals of the richer people. For example, the bottom 10 per cent, or the poorest section, of the population would have to incur in rounds (15, 16) something of the order of 80 per cent more expenditure, and the top 10 per cent of the population only about 30 per cent more expenditure on cereals, in comparison with their respective expenditure on cereals in rounds (8, 9), in order to keep the per capita consumption of cereals in rounds (15, 16) the same in physical terms as in rounds (8, 9).

7. *Changes in shares of cereals consumption:* The next point to be noted is that because the *actual* expenditure on cereals of the poorer households would have to be relatively greater in rounds (15, 16), in order to keep the level of consumption the same in physical terms, the *share* of expenditure of the poorer sections would also be relatively greater in rounds (15, 16) in comparison with their share of expenditure in rounds (8, 9). The cumulative proportion of the population and the accumulated share of expenditure from the bottom decile would also be higher for each decile upto the ninth decile, and would become 100 per cent in the tenth decile, that is, for the population as a whole. Relevant data for consumption of cereals in rural areas in India are given in Statement (3), in which col. (2) shows the cumulative

STATEMENT (3): AVERAGE CUMULATIVE PERCENTAGE SHARES OF QUANTITY AND VALUE OF CEREALS CONSUMED AND OF TOTAL CONSUMER EXPENDITURE IN NATIONAL SAMPLE SURVEY IN ROUNDS (8, 9) (LOW PRICES) AND IN ROUNDS (15, 16) (HIGH PRICES): ALL INDIA RURAL

sl. no.	cumulative percentage of population	cereals				total expenditure	
		quantity		value		rounds 8, 9	rounds 15, 16
		rounds 8, 9	rounds 15, 16	rounds 8, 9	rounds 15, 16		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	10	6.28	5.52	4.50	5.02	3.00	3.54
2	20	13.80	13.24	10.61	11.96	7.29	8.54
3	30	22.44	21.64	18.18	19.88	12.67	14.50
4	40	32.46	31.05	27.12	28.78	19.05	21.38
5	50	42.22	40.78	36.62	38.41	26.41	29.17
6	60	52.68	51.26	47.28	48.84	35.00	37.98
7	70	63.70	62.22	58.44	59.70	45.16	48.04
8	80	74.74	73.49	70.43	71.46	57.47	59.95
9	90	86.92	85.34	84.20	84.24	72.88	74.62
10	100	100.00	100.00	100.00	100.00	100.00	100.00
*11	top 10%	13.08	14.66	15.80	15.70	27.12	25.38

* The bottom line 11 gives the respective shares of the top 10% of the population, which are obtained by subtracting the figures for the ninth decile in line 9 from 100 in each of cols. (3) to (8).

percentage of population from the bottom. The cumulative shares of the quantity of cereal consumed up to each fractile group in rounds (8, 9) are shown in col. (3), and in rounds (15, 16) are shown in col. (4). The cumulative shares of the value of the cereals consumed, that is, the cumulative shares of expenditure on cereals, are given in col. (5) for rounds (8, 9), and in col. (6) for rounds (15, 16). The corresponding figures for total expenditure are given in cols. (7) and (8).

8. In line 1 of Statement (3), for the bottom 10 per cent of the population, it can be seen from col. (3) and col. (4), that the share of consumption of cereals, in physical quantities, was 6.28 per cent in rounds (8, 9) against 5.52 per cent in rounds (15, 16), showing a difference of 12.1 points¹ in consumption in real terms. From cols. (5) and (6) it can be seen that the share of the money value of the consumption of cereals was 4.50 per cent in rounds (8, 9) and 5.02 per cent in rounds (15, 16), showing

¹ This is a measure of the difference of the two percentages expressed as a percentage of the estimate for the period of low prices in rounds (8, 9).

8(b) - INDIA

a rise of 11.6 points in the share expenditure of cereals. From line 11 of Statement (3), for the top 10 per cent of the population, it can also be seen that the share of consumption of cereals in physical terms increased from 13.08 per cent in rounds (8,9) to 14.66 per cent in rounds (15, 16), showing a rise of 12.1 points, while the share of expenditure on cereals remained practically the same, about 15.80 per cent. Changes in the shares of the total consumption expenditure can be compared from figures given in cols. (7) and (8). It can be seen from line 1, for bottom 10 per cent of the population, the share of expenditure increased from 3.00 per cent in rounds (8, 9) to 3.54 per cent in rounds (15, 16), or by 18.0 points, which was more than the increase of 11.6 points in the share of the expenditure on cereals. From line 11, it can be seen that the share of total consumption expenditure of the top 10 per cent of the population had decreased from 27.12 per cent to 25.38 per cent, showing a reduction of 6.4 points. To sum up, for the bottom 10 per cent of the population, the share of consumption of cereals decreased by 12.1 points in physical terms but the share of expenditure on cereals increased by 11.6 points, and the share of total expenditure increased by about 18.0 points. For the top 10 per cent of the population the position was quite different. The share of consumption of cereals in physical quantities increased by 12.1 points, while the share of expenditure remained the same, and the share of total consumption expenditure decreased by 6.4 points. It is clear that disparities in the consumption of cereals had increased with the rise in prices of cereals.

9. Estimates of cumulative shares, for different decile groups, given in Statement (3), are shown in the form of concentration curves in Chart 2. Two concentration curves for consumption of cereals, in physical terms, are shown in green colour; and two corresponding concentration curves for consumption of cereals in terms of money, that is, the two concentration curves for the expenditure on cereals, are shown in red colour. For purposes of comparison, two concentration curves for the total consumption expenditure are given in black colour. In each case the solid graphs represent conditions in rounds (15, 16) when prices were high, and the broken graphs represent conditions in rounds (8, 9), when prices were low. It can be seen that the broken green graph, for the distribution of consumption of cereals in physical terms when prices were low, is nearer to the line of equal distribution than the solid green graph representing the distribution of consumption of cereals in physical terms when prices were high. On the other hand, in the case of the two graphs in red colour (for distribution of expenditure on cereals) and also the two graphs in black colour (for distribution of total consumption expenditure), the position is reversed; broken graphs, representing conditions when prices were low, are further away from the line of equal distribution than the solid graphs, representing conditions when prices were high. In other words, when prices increase, the distribution of consumption, in terms of money, tends to become more equal, but the distribution of consumption, in physical terms, tends to become more unequal. It must be noted that it is only when the increase of prices of cereals (or of other goods and services) are relatively greater for the poorer sections of the population that one would expect the distribution of consumption in terms of money values (that is, of consumption

expenditure) to become more equal, and the distribution of consumption in physical terms to become more unequal.

10. *Test of significance of change in concentration* : The greater equality of the distribution of total consumption expenditure when prices are high can be examined critically with the help of the two sub-sample estimates for s.s. 1 and s.s. 2 in the NSS data for round 8 (July 1954-March 1955) and for round 16 (July 1960-June 1961) which are given in Statement (4), and are shown graphically in Chart 3.

STATEMENT (4) : CUMULATIVE PERCENT SHARES OF TOTAL CONSUMER EXPENDITURE BY FRACTILE GROUPS FOR ROUNDS 8 AND 16 OF THE NATIONAL SAMPLE SURVEY : ALL INDIA RURAL.

cumulative percentage of population	round 8 (July 1954-March 1955)			round 16 (July 1960-June 1961)		
	s.s. 1	s.s. 2	combined*	s.s. 1	s.s. 2	combined*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
10	2.90	3.11	3.00	3.47	3.36	3.41
20	7.03	7.50	7.26	8.39	8.04	8.21
30	12.09	13.01	12.54	14.29	13.76	14.01
40	18.20	19.35	18.78	21.07	20.46	20.75
50	25.34	26.61	25.98	28.79	28.04	28.40
60	33.58	35.07	34.33	37.59	36.78	37.17
70	43.25	45.20	44.23	47.65	46.71	47.16
80	54.84	57.62	56.22	59.34	58.45	58.87
90	69.56	73.53	71.55	73.90	73.48	73.65
100	100.00	100.00	100.00	100.00	100.00	100.00

*No. of sample households : round 8, 1833; round 16, 3703 with roughly half the number in each sub-sample.

In Statement (4) the cumulative percentages of the population are given in col. (1); cumulative percentages of total consumption expenditure are given for sub-sample 1, sub-sample 2 and the combined sample for round 8 in col. (2), col. (3) and col. (4) respectively; and corresponding cumulative percentages of consumption expenditure for round 16 are given in col. (5), col. (6) and col. (7), respectively. It would be seen in Statement (4) that, for each fractile (decile) group, all three figures for round 16 when prices were high, are greater than the corresponding figures for round 8 when prices were low. The three concentration curves, for the figures given in cols. (2), (3) and (4), for round 8, when prices were low, are shown in black colour in Chart 3; and the three concentration curves for figures given in cols. (5), (6) and (7) for round 16, when prices were high, are shown in red colour in Chart 3. As already pointed out, in fractile graphical analysis, the geographical area bounded by the two sub-sample graphs gives the error-area of the combined graphs. Also the 'separation' is given by the area lying between the two graphs for the two combined samples, that is, in this case, the area lying between the graph in solid red line and the graph in solid black line. It would be seen that the three graphs in red colour are completely separated from the three black graphs¹, and are nearer the line of equal distribution.

¹ With one exception at 9th decile for round 8, namely, the cumulative per cent share 73.53 for s.s. 2 in col. (3), is slightly greater than the corresponding value 73.48 in col. (6) for round 16.

8(b) - INDIA

On the basis of studies on error-areas, it may be noted that such complete separation¹ indicates that the difference between the two distributions of total consumption expenditure during periods of high and low prices respectively, was statistically significant. It is, therefore, reasonable to think that, with an increase in prices, the distribution of expenditure on consumption did become more equal.

11. *Ranking of concentration curves by price levels*: The concentration curves of consumption expenditure for seven NSS rounds have been placed in order of increasing concentration of distribution (that is, in order of increasing distance from the diagonal line) in col. (1) of Statement (5) together with several series of index numbers of prices for the rural area. Each series of price

STATEMENT (5): RANKING OF CONCENTRATION CURVES FOR PER CAPITA TOTAL CONSUMER EXPENDITURE OF DIFFERENT NSS ROUNDS IN ORDER OF INCREASING CONCENTRATION (INCREASING DISTANCE FROM THE DIAGONAL OR EQUAL DISTRIBUTION LINE) AGAINST INDEX NUMBERS OF PRICES: ALL-INDIA RURAL

srl. no.	NSS round in order of increasing concentration*	national income deflator	NSS price index of cereals			wholesale price index (1952-53=100)		
			bottom 10 per cent	bottom 20 per cent	general population	cereals	other commodities	all commodities
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1.	15	105.6	138	131	116	105.6	122.4	119.1
2.	16	107.2	141	133	118	103.2	131.3	125.9
3.	12	100.4	131	126	111	103.0	110.8	109.3
4.	13	102.1	117	116	108	98.7	109.5	107.4
5.	7	98.0	100	100	100	92.8	104.1	101.9
6.	9	91.9	76	75	79	73.3	95.5	91.2
7.	8	90.4	76	79	82	77.4	100.6	96.1

* That is, in order of increasing distance from the diagonal line of absolute equality of distribution.

index numbers has reference to a specific field of economic activity; the different series of price-index numbers are, therefore, not strictly comparable among themselves. Nevertheless, they show general agreement in the up or down movements of prices. For the rural population, it would be seen that two top curves for the 15th and 16th rounds (lines 1, 2) are associated with the highest levels of all the series of price index numbers; the intermediate curves for the 12th and 13th rounds (lines 3, 4) are associated with intermediate levels of prices; and finally, curves for the 7th, 8th and 9th rounds at the bottom (lines 5, 6, 7) are associated with the lowest levels of prices.

12. *Effect of rise in prices on disparities*: On the whole, the higher the price level in any round of the survey, the nearer is the position of the corresponding concentration curve of total consumption expenditure to the diagonal line, that is, the greater is the equality of distribution, by size, of total consumption expenditure. Conversely, the lower the level of prices during a round of survey, the greater is

¹ The size of the sample, i. e., the number of households (1833 in round 8, and 3703 in round 16) was evidently adequate for establishing the statistical significance of the separation.

the inequality of distribution, by size, of total consumption expenditure. An increase in prices is thus associated with greater equality, and a decrease in prices with greater inequality, of distribution by size of expenditure.

13. *Association between higher prices and greater equality of distribution of expenditure*: This observed association between higher price levels and greater equality of distribution of consumption expenditure deserves further consideration. From Statement (5), cols. (6) and (7), it is seen that the increase in the wholesale price of commodities, other than cereals, was much greater than the rise in the wholesale price of cereals, and, almost certainly, was also greater than the rise in the retail price of cereals. Secondly, the rise in the price of cereals was relatively much greater for the poorer sections of the population; and it is very likely that the rise in the price of commodities, other than cereals, was also relatively greater for the poorer people. The poorer sections of the population would have to pay relatively higher prices to maintain the same level of real consumption, and their share of total consumption expenditure would increase without, however, an increase in their share of real consumption, in comparison with sections which were better off. That is, an appreciable increase in prices would affect poorer sections more adversely, and would generally lead to an increase of disparities in consumption in real terms. It may be noted, however, that disparities may increase, or may remain the same, or may decrease, while the absolute level of consumption changes in one direction or in another. For social justice, it is necessary that the absolute level of real consumption should increase and inequalities should also decrease. Unfortunately, a rise of prices in India was found, very often, not only to have increased disparities but also to have lowered the absolute level of consumption in real terms of the poorest sections of the population¹.

14. *General conclusion*: From the information and analysis given in the paper in the press², it seemed likely that there was some improvement in the level of living and an increase of opportunities in many directions for most of the people of India during the first decade of planning (1951-1961). The rate of growth of consumption was still low. Benefits arising from growth also seemed to have accrued, in a larger proportion, to the richer sections of the people. In consequence, disparities between the rich and the poor did not generally decrease, but increased in many respects. The level of living is so low for a very large part of the population that it would be difficult to bring about an appreciable improvement in the absolute level of real consumption or a reduction of disparities without a large and rapid increase of income and availability of goods and services in the near future.

¹ International comparisons could not be made because it was not possible, up to 1966 or 1967, to secure, or to trace, published data, (or unpublished information to which access could be obtained) on changes in prices of cereals over time, separately, for different fractile groups, or for different income-brackets or different expenditure-brackets, that is, separately, for households of different levels of living for any other country (not excepting U.K., U.S.A., U.S.S.R., Japan etc.)

² The present note is a slightly revised version of a note which was placed before the Seventh Conference of Commonwealth Statisticians in New Delhi in November 1970.

The opposite effects of rise in prices on concentration of consumption in money values and in real terms were placed by me as Chairman with other materials before the Income Distribution Committee of India in February 1964 as a draft of part 2 of its report. Because of differences of opinion, further studies were undertaken during the next four years. As no unanimous decision could be reached, part 2 of the report was published in July 1969 excluding the detailed studies which the Committee decided would be published as a scientific paper by me in my individual capacity; this paper is in the press.

8(b) - INDIA

CHART 1: NATIONAL SAMPLE SURVEY : ALL INDIA RURAL : INDEX NUMBERS, BY FRACTILE GROUPS, OF AVERAGE PRICE OF CEREALS IN ROUNDS (15, 16) WITH PRICE IN ROUNDS (8, 9) AS BASE = 100.

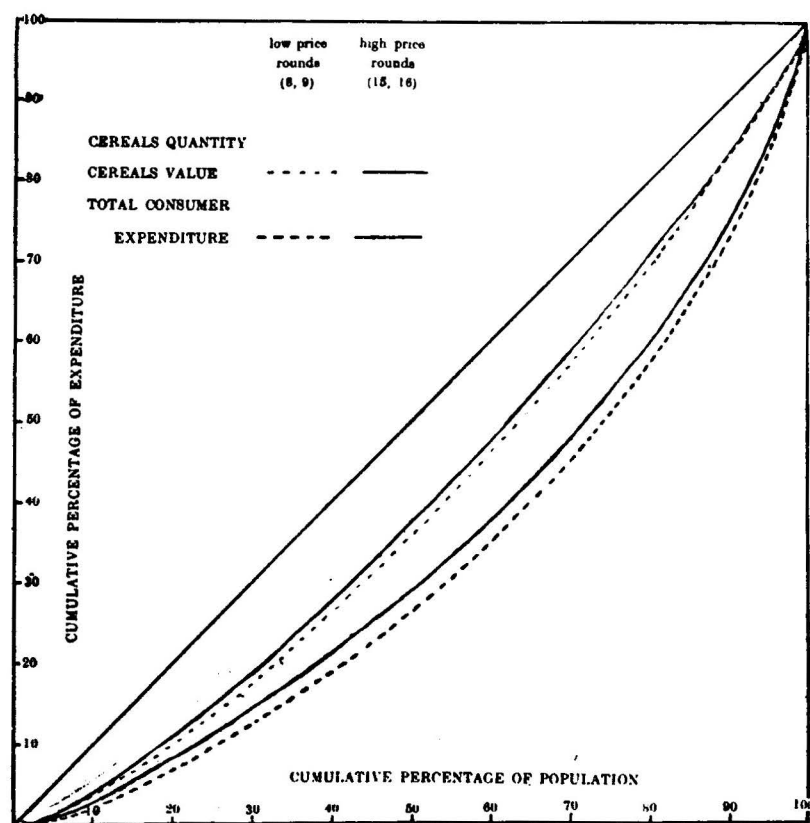
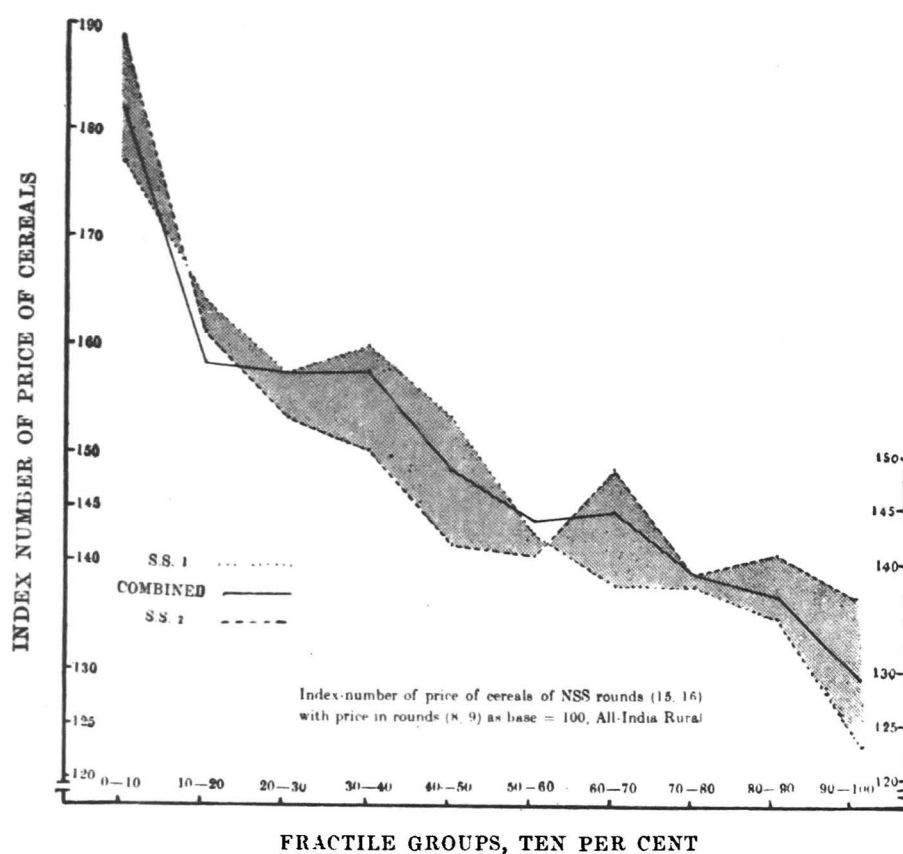


CHART 2 : NATIONAL SAMPLE SURVEY : ALL INDIA RURAL: CONCENTRATION CURVES FOR PER CAPITA CONSUMPTION OF CEREALS PER 30 DAYS, IN PHYSICAL TERMS (QUANTITY) AND IN VALUE TERMS AND ALSO FOR TOTAL CONSUMPTION EXPENDITURE IN ROUNDS (8, 9) AND (15, 16).

8(b) - INDIA

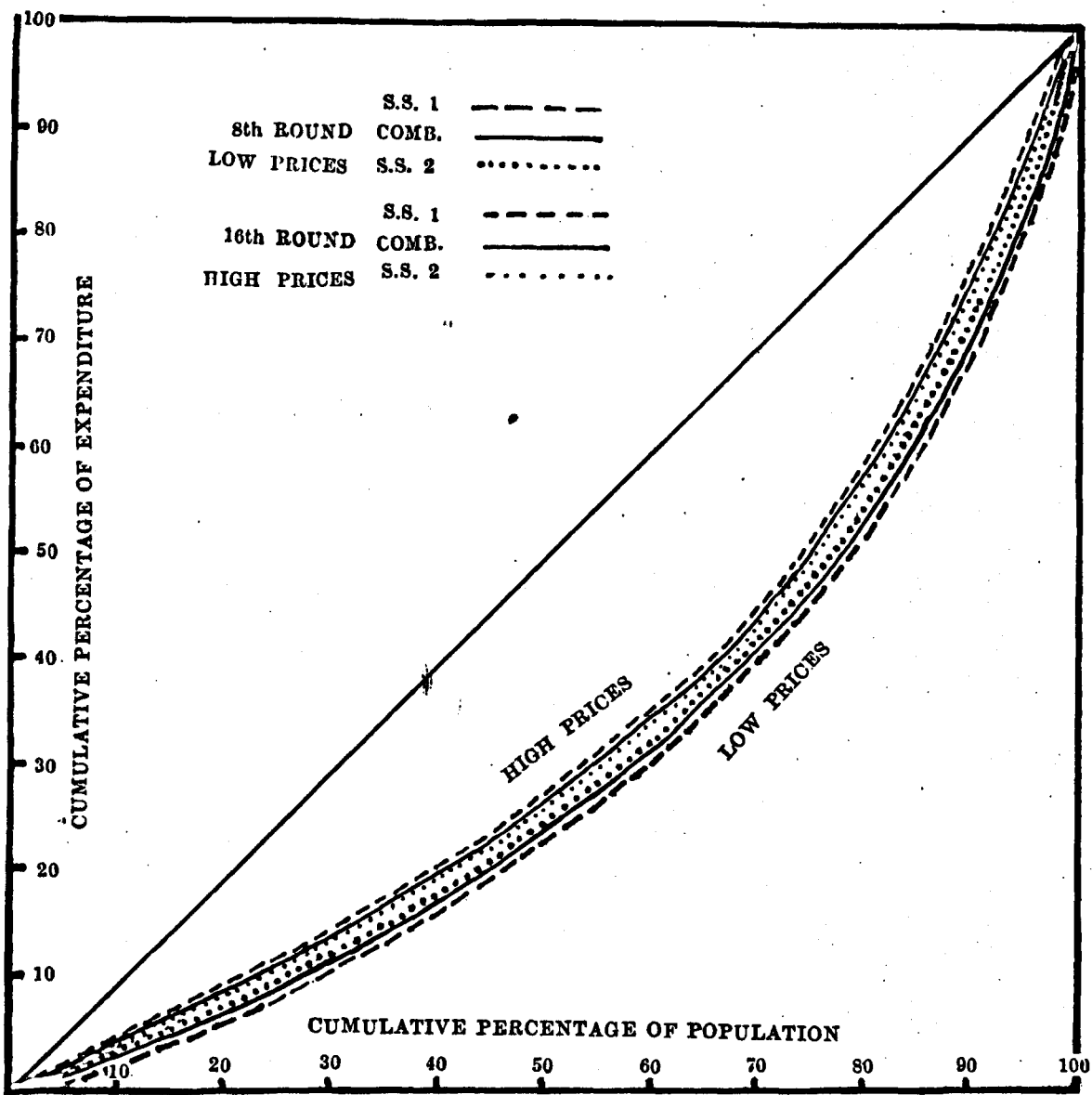


CHART 3 : NATIONAL SAMPLE SURVEY : ALL INDIA RURAL : ROUNDS 8 AND 16 :
 CONCENTRATION CURVES FOR THE DISTRIBUTION OF TOTAL
 PER CAPITA CONSUMER EXPENDITURE PER 30 DAYS.