

Proceedings of the Symposium on Problems and Prospects of Large Scale Computerisation of the 1981 Census

THE Symposium was held at the Central Statistical Organisation in New Delhi on November 4-5, 1977. It was organised by the Indian Association for the Study of Population jointly with Central Statistical Organisation, the Office of the Registrar General, India and the Indian Social Institute, New Delhi. The specific purpose of the Symposium was to explore the possibility of more effective large scale utilisation of computer technology in processing the census data, in order to better serve the varied purposes for which the census data are used, including the needs of planning, in particular and social research, in general. The Symposium brought together the producers of the census data, the specialists in computer technology and the users of the census data, particularly those dealing with planning exercises and those engaged in social and economic research based on data regarding population. A total of 39 participants drawn from 19 agencies and institutions attended the Symposium. The Symposium discussed the subject in four sessions devoted to the following topics : (i) General issues relating to large scale computerisation of census data; (ii) issues relating to sampling; (iii) problems relating to computer hardware and software; (iv) financial and administrative aspects of computerisation. A final brief session was devoted to the adoption of conclusions and recommendations emerging out of the four sessions devoted to substantive discussions. In this report we first present the conclusions and recommendations of the Symposium followed by a brief resume of the discussions. Then we give the background papers for the

different sessions, together with points made by the discussants selected for different sessions. The Association wishes to place on record the generous hospitality of the Central Statistical Organisation in making all the necessary arrangements for the holding of the Symposium and their deep sense of appreciation to the Office of the Registrar General and the Indian Social Institute for their active collaboration in making the Symposium a worthwhile venture.

Conclusions and Recommendations

I. General Issues Relating to Large Scale Computerisation of Census Data

1. Electronic data processing has a crucial role to play in the 1981 population census for timely and comprehensive tabulation of the census data and for storage and retrieval of information. Efforts should be made for better exploitation of computer hardware and software than was possible in the previous census.

2. Extensive manual tabulation will still have to be resorted to for obtaining quick results of the total population count and for the preparation of the primary census abstract for each village/enumeration block.

3. The list of items to be covered by the Census may broadly be the same as the one followed in the 1971 Census with minor revisions and modifications as are deemed necessary in the present context. The items to be canvassed should be reviewed in the 1981 Census also with a view to improving the quality of data.

4. The more complex census items may be tabulated with the help of the electronic computer as in the 1971 Census on a sample basis. The sample should be of adequate size to provide all the census tables with sufficient reliability at the level of a district at least and preferably at the level of a Community Development Block. Block level estimates would be extremely desirable to meet the current requirements of integrated rural development, for which retrieval facilities are essential.

5. Data processing should be so designed as to obtain all the final results (up to the district level) latest by April 1984 in a printed form. It is necessary for this purpose to complete the operations of data transfer and computer editing within a period of 24 months of the census and also organise rapid processing and printing of the results. Release of tables could follow a pre-assigned pro-

gramme of priorities in which migration, age distribution, education and economic activity may be considered.

II. Issues Relating to Sampling

6. The main issue relating to sampling is whether it should be adopted at the stage of census enumeration or whether the past practice of its use at the stage of tabulation should continue. As the introduction of sampling at the enumeration stage is known to pose operational and technical problems, it is necessary to undertake field investigations to study these problems and to examine the feasibility of this approach which has decided advantages in terms of economy and control of field work. The Registrar General may initiate these field studies.

7. Special studies are also required to determine the nature and extent of sampling. A household is considered as an efficient primary sampling unit and should be preferred to the individual, or a geographical entity such as a village/enumeration block.

III. Problems Relating to Computer Hardware, New Computer Needs and Problems of Software

8. Data preparation should be on a decentralised basis using direct key-to-tape or key-to-disk systems. A mix of direct data entry systems and other systems could be considered, if necessary. In so far as data entry systems are concerned, efforts should be made to make use of spare computer capacity of other, agencies.

9. Computer facilities available at various State and regional centres should be utilised.

10. Priority should be given to the census jobs at State and regional computer centres for both data preparation and computer processing.

11. In the event of indigenous key-to-tape/disk facilities being not available in time, import of these items may be considered. Alternatively, punching/verifying machines available at various centres may be requisitioned for census jobs.

12. The Registrar General's Office may initiate a pilot project to assess the use of mark-sensing for census data.

13. Irrespective of whether National Informatics Centre (NIC) is fully functional or not, a computer system with provision for direct data entry, mark-sense card recording, optical reading facilities, tape peripherals of high density, disk, printer, graphic facilities should be provided to the Registrar General's Office in 1979. This system should be compatible with the NIC system and capable of being linked with the NIC main system.

14. The Registrar General may set up a software group to look into both the system software as well as system specifications and standardisation for cleaning of data, criteria for cleaning, maintenance of data bank, storage and retrieval of data etc. Demographers and other users of census data may also be associated in the preparation of validity checks.

15. Computer aided printing should be investigated for adoption.

16. It is important that a high-powered committee of direction is set up, with authority to give precedence to census operations by the Government computers and other data processing agencies; the Registrar General should be a member of this committee.

IV, Financial and Organisational Aspects of Computerisation

17. It was felt that the estimate of Rs. 17.10 crores as the cost of computerisation for the 1981 Census was realistic. Minor adjustments will have to be made within this figure of Rs. 17.10 crores for the staff to be appointed. To this total figure, a small additional sum may have to be added as training cost.

18. On the basis of this estimate, it was suggested that a committee comprising representatives of the Registrar General's Office, the Central Statistical Organisation, the Department of Electronics, Electronics Corporation of India, and the Planning Commission should begin work on closer estimates of the total cost and to spell out more details in the light of the above recommendations.

19. Adequate account should be taken of the subsequent facilities that will have to be provided to a variety of users demanding data at various levels of disaggregation. This question can only be taken on hand around 1982 or 1983 when the Registrar General divests himself of the main responsibilities of producing the stipulated census and ancillary tables.

20. It was felt that so far as Government users are concerned they might be required to pay the full costs; so would the commercial users. As for academic,

Specialised research, institutions and individual research scholars, a formula would have to be worked out for subsidised rates.

21. The committee constituted as per recommendation 18 above may be requested to go into the question of comparative costs of conventional printing followed so far by the Registrar General's Office and of printing straight from the computer output. In making these estimates of printing costs, the time gained or lost by either process leading to denial of valuable information to essential users will also have to be taken into account. This matter should be viewed as next in importance only to the desirability of computerisation of the 1981 Census data.

Resume of the Symposium

The first session of the Symposium under the chairmanship of Sri V. R. Rao, Director, Central Statistical Organisation, considered the general issues relating to large scale computerisation of census data. The Chairman in his introductory remarks drew attention to the fact that the preparations for 1981 Census were entering in an advanced stage where the consideration of computer processing of the data is quite opportune and worthwhile. Recalling that the matter had received some attention in the Data Base Seminar of 1974, the Annual Conference of the IASP in 1976, C.S.O. sponsored seminars of 1977, and the meetings of the U.N. Statistical Commission and the Statistical Advisory Committee, he underscored the growing interest in large scale computer processing of census data in order to provide timely and comprehensive tabulations and to facilitate storage and retrieval of information at very modest cost. Then followed the presentation of papers and notes by P. Padmanabha, the Registrar General of India, Asok Mitra and Stan D'Souza.

In his presentation, the Registrar General spelt out the present thinking about the different aspects of the 1981 Census including the scope of information to be collected and the work-load involved. He also referred to estimates of time and costs for the use of computer on the assumption that the 1981 Census would broadly follow the pattern set by the 1971 Census. Asok Mitra drew attention to the reservations with which the quality and reliability of the census data must be accepted. He dwelt upon the system of data collection, the difficulties of planning computer capacity and time, paucity of qualified computer personnel,

inadequacy of financial provision and lack of printing facilities. He desired careful consideration of the extent to which computerisation is worthwhile in the light of the obtaining level of quality and reliability. He emphasised that quite a sizeable part of the tabulation would have to be done manually in order to provide a quick overview of the census findings.

Stan D'Souza narrated his experience with electronic data processing in connection with the 1974 Census of Bangladesh. He drew attention to several aspects of the problem including the assessment of data needs, technical advice, choice of the stage at which sampling can be introduced, and training and supervision of the field and processing staff. He considered sampling at the enumeration stage to be more efficient and less expensive and pleaded for generating information on socio-economic characteristics separately for rural and urban areas through carefully designed sampling procedures.

The presentation of these papers evoked considerable discussion which covered a variety of topics including the scope and the methodology of data collection, relative proportions in which manual and computer processing could be combined, the organisation and maintenance of a data bank facilitating general access to all types of data including religious and linguistic information, time schedule of tabulation, decentralisation of data processing, employment of women and students as enumerators, training and supervision procedures, increase in the financial outlays on the census and the question of acquiring new computer system sufficiently in advance.

Ashish Bose made a strong plea for co-existence of hand tabulation and electronic processing of data in the 1981 Census. He was of the view that the financial costs involved in maximising the use of census data should be included in budget for planning development. He further suggested that the use of computer technology warrants the provision for a matching printing technology and so there should be a printing press at the entire disposal of the Registrar General. From this discussion there emerged a set of five conclusions and recommendations, included in the first part of this report.

II

The second session, held under the chairmanship of Prof. A. L. Nagar of the Delhi School of Economics, Delhi University, took up the different aspects of the use of sampling procedures in the census. The Chairman said that while he

Was not quite familiar with the approaches developed by the Indian Census, it was obvious that sampling could be used with profit in the census operations as it covers a very vast population about whose social and economic conditions, it is necessary to have reliable information for purposes of planning for social and economic development. The session had before it a very exhaustive paper which was prepared by the Director, C.S.O.

In presenting the paper he laid great emphasis on the importance of sampling for the general improvement of the quality of data, for reducing *work-toad* of enumeration and data processing, for shortening the time-lag between the collection of data and the release of results, and for making the census more economical. He drew attention to the detailed discussion in the paper of the different aspects of the problem of introducing sampling in the census. In doing so, he identified items which required universal coverage and! those that could be canvassed on a sampling basis. Further, he identified the possible sampling units, described sampling procedures including the choice of sample fractions, and developed a broad strategy for the use of sampling procedures in the census. He suggested that in the present context it was necessary to undertake a series of studies and field trials in order to develop an appropriate *sampling* plan for the census.

In the detailed discussion that followed it was generally recognised that sampling should be used in the 1981 Census. The discussion related to the following issues: (i) whether sampling should be operated at the enumeration or the tabulation stage; (ii) whether the unit of sample should be individual, household or village/urban block; (iii) the optimum size and *methodology of sampling*; (iv) practical difficulties in the identification of the sample households in the field; (v) substitution of samples in the case of non-availability; (vi) the difficulties in simultaneous canvass of universal slip and a sample schedule; (vii) the question of separating the sample canvass from the universal canvass of individual slip through field staff of organisations like the N.S.S. and the State Statistical Bureaus; (viii) tailoring of the training programmes to the requirements of universal count and the sample canvass; and (ix) adequate remuneration to the different categories of field staff. The discussion concluded by making two recommendations emphasising the need for exploring the subject especially by conducting field trials (see para 6 and 7 of the first part).

III

The third session, held under the chairmanship of Brig. V.M. Sundaram, Direc~

tor, Army Computer Centre, focussed attention on a large variety of issues connected with developing computer hardwares and softwares to meet the needs of processing the voluminous census data. The Chairman reviewed the evolution of computer facilities in the country, identifying the different centres that have come up, and further noted that a national net-work of computers was emerging in the country with NIC acting as a nodal point. He envisaged a fair possibility of the Census Organisation acquiring one of the terminals in the net-work.

The discussion in this session began with two presentations, one by N. Seshagiri of the Electronics Commission and the other by Brig. A. Balasubramanian of the Department of -Electronics, In the former presentation Seshagiri spelt out the objectives of the National Informatics Centre as related to the development of hard and soft wares in the country. He referred to the installation of a high powered computer at the centre with VHF/UHF communication systems and terminals at several different places, of which the R.G.'s Office could be one. He, however, pointed out that the Census requirement warranted 'peak-load planning' implying utilisation of spare computer capacity throughout the country. He suggested use of multiple key-to-tape or key-to-diskette devices for the census and suggested that the developing validation and preparation software could be made available through the NIC to the Census free of cost. He expressed the readiness of the NIC to assist and collaborate with the Census in all aspects of the use of computer technology.

Brig. Balasubramanian gave details of the tabulation done on IBM 1401 computer for the 1971 Census with punch cards as input media. He suggested that raw data after requisite editing could be made available to various data users on request. For this purpose, it was desirable to hold data in random access files containing multiple indices. He suggested the use of key-to-tape data entry devices for improving the quality of data and reducing costs. He was in favour of decentralisation of coding, checking and data entry operations, increased use of data entry devices, high density (1600/3200 BPI) 9 track tapes for storage of data, direct access media for data analysis and continuous utilisation of computer time.

In the discussion that followed it was recognised that it was also necessary to provide for the supply of low density tapes to users unable to use the high density tapes. Reservations were also expressed about the introduction of mark-sensing techniques. The general opinion was that the whole area of the development of hard and soft wares should be thoroughly explored by a committee

of experts. Such a committee could study the problem of establishing data entry devices, computer systems, training procedures, validation checks etc. The recommendations emerging from this discussion are given in the first part (see para 8 to 16).

IV

The final session of the Symposium discussed, under the chairmanship of Prof. Asok Mitra of the Jawaharlal Nehru University, the financial and administrative aspects of computerisation. The Chairman emphasised the overriding need of improving the quality of census data through a thorough-going reform of the data collection procedures involving census enumerators. He advocated increase in the honorarium to the enumerator to Rs. 100/- and thought that it was vitally necessary to ensure that the enumerator takes up the work voluntarily and carries it out with responsibility and keen interest in the job.

The Registrar General presented in this session details of financial and organisational implications of the programme of computerising the census operations on the pattern of the 1971 Census. He envisaged that 24 coding and punching cells will be set up for input preparation, transfer of data to tapes and correction and editing of the data. He said that while all these operations could be carried out at the zonal computer centres, the final processing will be done at the Registrar General's Office. He gave details of the staff requirements and estimated the total expenditure involved to be of the order of Rs. 16.20 to 17.09 crores depending on the type of input media used.

There was a general agreement about the reasonableness of the cost estimates presented by the Registrar General. The discussion centered around a great variety of administrative problems connected with collection, processing, storage and retrieval of census data. The general emphasis was on the need to take all feasible steps for improving the quality of data. It was felt that an expert committee comprising, for example, Brig. , Seshagiri and the Registrar General may go into the question of replacing the present IBM 1401 by a more sophisticated system with the facility of off-set printing, so that the delay in the release of census publications is significantly reduced. It was also suggested that the incidence of power failure which caused delay in computer processing could be avoided by installing their own generator at the census office.

In the brief final meeting of the Symposium, the conclusions and recommendations set forth in the beginning of this report were adopted and it was resolved that the President of the Indian Association for the Study of Population be authorised to forward the proceedings of the Symposium to the concerned official agencies including the Deputy Chairman of the Planning Commission, the Home Ministry, Finance Ministry and the Office of the Prime Minister.

Session I: **General Issues Relating to Large Scale Computerisation of Census Data**

PAPERS

P. Padmanabha

INDIAN census has over the years relied on manual processing and has, by and *-large, served the essential data needs of the country. In the post-independent India, the need for more data was felt in connection with planning for development. The Census organisation which processed the data manually as a one-time operation found it difficult to re-process the schedules to meet various data needs. Retrieval, complex cross-tabulation and accuracy posed problems. By then, Electronic Data Processing had come to be recognised as the most convenient and efficient mode of handling and processing of such data. The Indian Census Organisation, therefore, decided to go in for computer processing of the 1971 Census data and in preparation for it, hired a small computer system (IBM 1401 with 4 K memory) in October 1966. This system was augmented by increasing the memory from 4 K to 8 K with four tape drives in 1968 and further stepped up to 16 K with six tape drives in April 1971. This provided for input preparation through the media of punched cards and subsequent transfer of data to magnetic tape.

The two 'supplier—firms' in the field, IBM and ICT (now ICL), between themselves, were able to supply 650 punching and verifying machines against the requirement of 2000 machines to prepare the entire input in about two years' time. Then, there was the question of providing basic population statistics at the village/urban block level for various administrative purposes, including delimitation of constituencies. Having regard to all these limitations, it was decided that the 1971 Census data should be processed partly by manual process and partly on the computer. Accordingly, the following programme was drawn up.

- I. (1) Manual processing of 100 per cent of the main census schedule for basic population statistics.
- (2) Manual processing of a 10 per cent rural sample for further tabulation.
- II. (i) Processing of a 20 per cent urban sample of slips on EDP,
- (2) Advance tabulation with a 1 per cent sample of slips on EDP (rural and urban).
- (3) Processing of a 20 per cent sample of houselist data on EDP.
- (4) Processing of 100 per cent establishment data on EDP.
- (5) Processing of 100 per cent data pertaining to degree holders and technical personnel on EDP.
- (6) Processing of 4 per cent sample schedules pertaining to population record on EDP.

Thus a combination of manual and electronic data processing was resorted to for handling 1971 Census data. The following table gives an account of the volume of 1971 Census processing done on the aforesaid computer.

<i>Job</i>	<i>Vol. of records (in million)</i>	<i>Whether sample processed</i>	<i>Vol. of records processed (million)</i>	<i>Target date of completion</i>	<i>When completed</i>	<i>Remarks</i>
Houselist	144.0	20%	28.8	Feb. 72	Feb. 73	Priority Change
Establishment Schedule	10.2	100%	10.2	Feb. 72	Feb. 73	Priority Change
Degree holders and technical personnel cards (DHTP)	2.2	100%	2.2	Dec. 73	April 72	Higher Priority assigned
1 per cent advance tabulation	550.0	1%	5.5	Dec. 72	Sept. 72	Change of priority
Urban individual slips	109.0	20% (100% for small areas)	26.5	Dec. 73	July 74	
Census Population record	550.0	4% (100% for small areas)	39.3.	Oct. 75	Oct. 75	

After the main tabulation, the secondary processing of census data is being done on our own, as well as at the instance of other Ministries/Departments.

Preparation for the 1981 Census has begun. It will gain further momentum in the coming years. Suggestions for the coming census are being received. While formal consultations with data users will begin from early next year, this process had commenced much earlier. For instance, the topic was discussed at the Data Base Seminar of the Indian Association for the Study of Population (IASP) in collaboration with the Indian Econometric Society and the Indian Statistical Institute (ISI). A technical advisory group under the chairmanship of the Director, CSO also went into the topics and concepts for the ensuing census. The IASP organised another seminar at Bangalore in December 1976 where the approach to 1981 Census was discussed and has followed it up by the present symposium where the main focus is on computerisation. At this point, a brief discussion on the present thinking about topics to be canvassed in the forthcoming census may be relevant.

Houselist Form

As in the previous censuses, a houselist form will be canvassed at the coming census also. The enquiry will precede the population count and the possibility of a synchronous houselisting operation throughout the country will be explored. In this form the scope of enquiry has been enlarged by the inclusion of new topics, viz., (i) toilet facilities, (ii) bathing facilities, (iii) drinking water supply, (iv) domestic lighting, (v) cooking facilities, and (vi) number of married couples in the household. The Census Commission is also considering collection of some basic demographic particulars of residents enabling linkage with housing characteristics.

Establishment Schedule

The tentative schedule presented at the Bangalore seminar remains basically the same as its predecessor except that the employable capacity of the establishment is also proposed to be sought in 1981. There is, however, a thinking that in the wake of the Economic Census and follow-up survey by the Central Statistical Organisation (CSO), a census of establishments so soon after the Economic Census may not be necessary. The census of establishments covers all types of establishments and provides a frame for surveys whilst the Economic Census confines itself to establishments having at least one hired worker. Thus

the scope of the latter is somewhat restricted and may not serve the purpose of a frame for users like the IAMR, for example, or for formulation of programmes relating to household industry.

Individual Slip

Suggestions have been received from different quarters to collect detailed information on (i) physically handicapped persons; (ii) caste/tribe (as distinct from scheduled caste/tribe); (iii) distance of place of work from place of residence; (iv) more detailed data on women; and (v) knowledge and practice of family planning. The Census Commission on its own proposed that topics like 'nationality' and 'unemployment' should be revived at the next census. These suggestions and those on concepts had been discussed by the Technical Advisory Committee. On the question of concepts, the Committee arrived at the conclusion that the 1971 concepts should be adopted for the 1981 Census for the sake of comparability, if not for anything else. The suggestions are still under examination.

Household Schedule and Census Population Record

Here again the proposal is to revive the 1961 household schedule which besides being a population record elicited information on household cultivation and household industry. It has further been suggested that the scope of the population record should be enlarged to cover industry and occupation of each individual.

Degree Holder and Technical Personnel Card

This schedule is being canvassed since 1961 at the instance of the CSIR. No specific suggestions have yet been received from its authors as yet. The Census Commission, however, feels that to ensure better coverage it would be necessary to canvass the special schedule among all matriculates and above. The screening can always be done at the tabulation stage. With this in view the question on educational level in the individual slip is proposed to be split into 'matriculation and above' and 'others', the former serving as control for canvassing of the special schedule.

We may now proceed to examine the work-load and implications of computerisation of the 1981 Census. The following assumptions have been made :

- (1) India's population will be around 700 million in 1981.

- (2) The five schedules discussed above will be canvassed in 1981.,
- (3) Sampling will be resorted to at the tabulation stage for the following with the intensity noted against each :
- | | |
|-----------------------------|---|
| (i) houselist | 20 per cent |
| (ii) establishment schedule | 20 per cent |
| (iii) individual slip | 2 samples will be drawn: (a) 1 per cent national sample for advance tabulation, and (b) 20 per cent sample for detailed tabulation. |
| (iv) household record | 10 per cent sample. |
- (4) The degree holder and technical personnel card will have to be renamed consistent with its extended scope and will be processed on 100 per cent basis.
- (5) The individual slips will be processed manually (100 per cent) for yielding basic population statistics at the village and ward/block level.
- (6) Sophisticated input preparation devices like, mark-sensing, optical reader, microfilming by machines such as FOSDIC of the US Bureau of Census, etc. will not be available.
- (7) The tabulation programme will broadly follow the 1971 pattern.
- (8) The present computer system i.e., IBM 1401 with 16 K memory and 6 tape drives will be available.

Table 1 presents in a tabular form the work-load and time estimate in respect of each schedule and for each of the operations involved in computer processing of the 1981 Census data. There has been a suggestion to operate a 100percent computerisation of the data in 1981. Its implications have also been shown in this table. The norms used for computing the time estimate are given in Table 2.

TABLE 1—WORK-LOAD IN TERMS OF TIME AND QUANTUM FOR DIFFERENT CENSUS OPERATIONS

	Records in lakhs	Cards in lakhs	Coding and Checking		Punching and Verification						Operation					
			Cod- ing months	Check- ing months	Hand pun- ches and verifiers		Automatic punches and verifiers		Key to tape		Card to tape, correc- tion and merging hours	Sort- ing hours	Edit- ing hours	Summ- ary hours	Sort- ing hours	Printing hours
					P	V	P	V	P	V						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1. Houselist 20%	340 ^a	170	4250	3400	2125	1889	1700	1416	1416	1416	1700	—	680	680	340	340
2. Establishment schedule 20%	24 ^a	8	480	400	100	89	80	67	67	67	80	96	48	48	24	24
3. Advance sam- ple 1%	70	70	875	700	875	778	700	583	583	583	700	—	210	840	420	210
4. Individual slip 20%	1500 ^a	1500	18750	15000	18750	16667	15000	12495	12495	12495	15000	—	4500	18000	9000	4500
5. Degree holder and technical personnel card 100%	120	120	3000	2400	1500	1333	1200	1000	1000	1000	1200	960	360	360	120	120

6. Household schedule and Population Record 10%	1100	400	13750	11000	5000	4444	4000	3332	3332	3332	4000	—	2200	2200	550	550
Total	3154	2268	41 105	32900	28350	25200	22680	18893	18893	18893	22680	1056	7998	22128	10454	5744
Individual slip80%	5500	5500	68750	55000	68750	61111	55000	45815	45815	45815	55000	—	16500	66000	33000	16500
Grand Total	8654	7768	109855	87900	97100	86311	77680	64708	64708	64708	77680	1056	24498	88128	43454	22244

NOTE : a. The figures in col. 1 take into account 100 per cent records ID the case of smaller territories.

b. Excludes areas for which 100 per cent processing is envisaged in item 4. The figures indicate additional work-load in the event of 100 per cent computerisation.

TABLE 2—NORMS ADOPTED FOR VARIOUS OPERATIONS

Coding	400 slips/lines per day/operator (200 for DHTP cards)
Checking	500 slips/lines per day/operator (250 for DHTP cards)
Punching	400 cards per day/operator (hand punch) 500 cards per day/operator (automatic punch)
Verification	450 cards per day/operator (hand verifier) 600 cards per day/operator (automatic verifier)
Key-to-tape operation	600 slips per day/operator for transfer and verification each operation

Following the estimates shown in Table 1, the following picture emerges ; (Does not envisage 100 per cent computerisation)

I Coding	41,105 man months
Checking	32,900
IIA Punching	28,350 man months (hand punch) 22,680 man months (automatic punch)
Verification	25,200 man months (hand verifier) 18,893 man months (automatic verifier)
IIB Key-to-tape	
Transfer	18,893 man months
Verification	18,893 man months
III Card-to-tape, correction and merging	50 months @ 450 hrs. per month in 3 shifts
IV Other operations	105 months

The requirement of personnel for different time periods by which the input should be prepared would be as follows:

TABLE 3—NUMBER OF PERSONS REQUIRED ON ONE SHIFT BASIS

<i>Operation</i>	<i>24 months</i>	<i>36 months</i>	<i>48 months</i>	<i>60 months</i>
Coding and Checking	3084	2056	1542	1233
Punching and Verification:				
Hand operated	2231	1487	1115	893
Automatic *	1732	1155	366	693
Key-to-tape	1574	1050	787	630

In 1971, Registrar General's Office computer installation was set up piece meal. For example, it started with a small system — IBM 1401 computer with 4-Kmemory. Incapacity was raised to 8-K with 4 tape drives. Finally, in April 1971, the capacity was stepped up to 16-K with 6 tape drives. The computer personnel were also appointed more or less on ad hoc basis which fell short of an integrated organisational pattern. The 1981 Census proposes to reorganise the computer division in order to ensure that it is able to handle the gigantic tasks which it will be called upon to manage.

The cost of processing of the 1981 Census data is about Rs. 16.20 crores over a period of seven years at the headquarters and three years each for the zonal centres and state coding and punching cells. The break-up is given below (on current rates).

TABLE 4- ESTIMATED COST OF PROCESSING OF THE 1981 CENSUS DATA

A.	1. Rentals for Central Computer systems and regional computers	210.20 lakhs
	2. Computer Stationery	5-10 „
	3. Printer ribbons	0.25 it
	4. Salaries of Staff	921.45 i>
	5. Cost of cards for 1981 census at 1971 rates	109.00 >>
		1246.00" ii
B.	1. Cost of tapes	22.82 »
	2. Tape cabinets, trolleys	1.90 »
	3. Cost of hand punches and verifiers and their maintenance	63.17 „
	4. Supervisory staff	286.59 »
		374.48 >
Total		Rs. 1620.48 lakhs

The estimate of Rs. 16.20 crores is based on the assumption that hand punches and mechanical verifiers are used. If automatic punches and verifiers are used (outright purchase), the cost works out to **Rs.** 16.86 crores, as against Rs. 16.20 crores by hand punches and mechanical verifiers. On the other hand, if automatic punches and verifiers were to be taken on rental basis, the cost would come to Rs. 16.37 crores and with key-to-tape device (outright purchase), the cost works out to Rs. 17.09 crores.

In 1971, a sum of Rs. 4.78 crores was spent for the processing of 64,2 million cards. This amount did not include the expenditure incurred on supervisory personnel, cost of tapes, tape cabinets, trolleys etc., cost of hand punches/verifiers and their maintenance. The estimated total expenditure of Rs. 16.20 crores compares well with the 1971 expenditure taking into account the fact that the volume of cards to be handled in 1981 would be a little more than 3,5 times 1971 work-load, the steep rise in cost of cards, all-round price hike for ancillary equipments and substantial rise in the emoluments of the personnel. The increase is also due to the anticipated additional processing time on the computer for tabulations to be generated from the household schedule.

It is too early to give an idea of the saving in cost and time that would accrue due to computerisation in the 1981 Census. This would depend on the questions to be included, the sampling fraction, the type of machines used and other factors. An idea of the broad magnitude of saving can, however, be inferred from the 1971 Census experience. It has been estimated that if all the projects which were done on the computer by making use of about 1,100 persons for 60 months, were to be done manually, the processing time required would be 135 months. On the other hand, if it had been stipulated that the entire work was to be finished in 60 months, we would have had to employ 2,480 persons. The comparable costs would have amounted to 47.8 million rupees for computer processing as against 59.4 million rupees for manual processing. The latter cost does not include costs of personnel management, and other over head costs like office accommodation, furniture etc.

The greatest advantage of computer processing would, however, be the efficient storage, reliable retrieval, uniform cross-classification, economy and speed of operation. Once the data are put on tapes, secondary processing would be cheaper; the data would be on tape in a compact form whereas manual processing would involve maintaining and handling millions of data repeatedly.

EARLY in 1971 the Registrar General published *the 1971 Census Tabulation *[^]Plan*. This booklet gave an account of the number of tables projected on the basis of (a) full count, and (b) sample count. The full count tables were supposed to be manually tabulated. The sample count tables were expected to come out of the electronic computer. The record of publication of manually tabulated and computerised sample tables of the 1971 Census gives us food for thought on the prospects of computerisation of the processing of the 1981 Census data. On any showing the computer has taken much more time than manual tabulation, even when it has worked on small samples to produce not-too-complicated tables. Here is a brief summary :

PUBLICATION STATUS OF THE 1971 TABLES

Tables Series	Coverage Full (F) or Sample (S)	Operation Manual (M) OF Computer (C)	No. of Tables	Printed and Pub- lished	In Progress	To be sent to press
1. General Population : A 1-V	F	M	5	4	1	—
2. Economic Tables : B I Part A; B II BIB; B11A, B; BTV A to D; BV A, B; BVIA (i, ii); BVIB (i, ii); BVII; BVIII; B IX A, B	F	M	2	2	—	—
3. Social and Cultural : C V; C VII; C VIII C I A; C I B (i, ii); C II; C m A, B; C IV; C V	S	C	17	—	17	—
4. Migration : D I to VI	F	M	3	3	—	—
5. Establishment : E I, E II A, B, C; E IH, E IV	S	C	8	4	—	4
6. Fertility : F I to V	S	C	6	—	6	—
7. Special tables for degree holders and technical personnel : G I to XII	F	C	6	5	1	—
8. Housing : H I to IV	S	C	5	—	—	5*
9. Special tables for degree holders and technical personnel : G I to XII	F	C	12	12	—	—
10. Housing : H I to IV	S	C	4	4	—	—
Total (All India)			68	34	25	9

Special All India Tables on 1 per cent Sample basis : In mimeograph only : limited availability.

*Tables based on 1 per cent sample are likely to come out of press in December 1977 and those based on 10 per cent rural and 20 per cent urban are going to press.

The figures tell their story. Of the 10 tables designed for hand tabulation as many as 9 tables have been printed and published at the all-India level, while 27 have been printed and published at the state level. At the all-India level the remaining one table is in the press awaiting publication. Nothing is pending tabulation. Apart from these 10 manually processed tables, the biggest single table, the Primary Census Abstract, in a minimum format of 35 columns, for each village, town or city block, taluk, sub-division, district, division and state, were all printed and published at the all-India as well as state and district level by 1974.

Not so with the computerised tables. Of the 58 tables designed for computer tabulation, of which 18 are on full count and 40, on a sample basis, only 23 have so far been published at the all-India level. As many as 24 tables at the all-India, and some of them at the state, level are still in the press, while some 4 tables are not even ready for the press and 5 based on 10 and 20 per cent sample are still to be sent to the press.

Computerisation, it will be readily conceded, cannot avoid teething troubles[^] Computerisation was undertaken for the first time in 1950-51 in the U.K. and U.S.A. Some of the most important tables were not published in either case until 1957-58. Much effort was spent on both hard and soft wares from 1954 onwards until tables began to appear according to schedule only in the 1970s. The bugs were not fully removed in the sixties, because some of the important 1960 series of the U.S. and U.K. tables were as late in appearing as 1965-66.

In India, electronic computerisation was tried out on a limited scale in 1961 for constructing eleven complex household economic tables, based on a 20 per cent sample of the total count. These computerised tables were all published by September 1964. But the bulk of the manually processed tables were published almost simultaneously.

Computerisation takes time in any country. It does not necessarily shorten time as is often naively imagined. It shortens time only when continuous research and effort are expended on improving hard and soft wares. This depends a great deal upon continuous investment in men and equipment involving large sums of money.

At this point I must hasten to add that I am talking under false pretences. I do not know anything about computers nor about the problems that are apt to

arise over large scale data-processing on computers. I am not aware either of the immense feats of improvement that have recently been accomplished in other parts of the world but are not readily accessible in India.

In the circumstances I shall confine myself to a few imperatives of the Indian situation, some of which have nothing to do with the question in hand but will, at the same time, determine the pace and quality of computerised tabulation in 1981 and even in 1991. I hope I shall be able to pose my questions concretely enough to invite informed discussion both at the policy and technical level.

The first question to ask ourselves is how much of the census information will deserve high level computerisation in our country. Computerisation justifies the expense and time when the quality of information collected is good. In the Indian Census we are plagued with data of doubtful quality and validity for several reasons.

The first, is the very low level of general literacy which reduces the level of response of the population censuses as well as that of the census taker. This is the basic malady which precludes the Indian Census from going in for major improvements in the questionnaire from one census to another. The same reason prevents the Registrar General from adopting more meaningful concepts, definitions, methods and classificatory systems in tabulation. Every type of data, age, marital status migration, language, economic activity, is only broadly, even vaguely, and at best approximately, described, and will not stand scrutiny except at certain levels of aggregation, where large numbers confer their own quality on the defective data. In the circumstances, computerisation cannot be expected to improve the quality of census data. If anything complex cross-tabulation will possibly bring out more garbage than one suspects.

The second, is the ad hoc and rather disheartening quality of the census taker. He is not paid for his work, insufficiently trained and virtually conscripted. He is often no more than barely literate, drawn from a bewildering variety of assignments few of which demand intellectual rigour. Many of the concepts and definitions so carefully formulated in the printed instructions escape his comprehension. In the result the census authorities have to be content only with the broadest and crudest responses to the census schedule. The majority of answers are given in very general terms, not lending themselves readily to detailed sub-classification. The computer scarcely gets a chance of showing its paces. It

has to be content with largely undifferentiated aggregates, not through any fault of its own, but for lack of detailed and intelligent response.

The third is the lack of planning of computer capacity and time and of the right quantity of computer personnel. As a result a large number of data processing problems are solved in an ad hoc manner as and when they arise at particular time and place. It is not always possible to work out a common programme even in the case of complex cross-tabulations and the programmer is often left uninitiated in how certain parts of a programme should be developed differently from certain other parts.

The fourth is the unwillingness of finance to commit money and machinery well in advance to be commensurate with the task envisaged. As a result computerisation is often compelled to proceed on as-and-when basis on a protracted, Zig-zag course.

The fifth is the lack of printing facilities straight from the computer print-out. The U.S. and U.K. Censuses made a success of the computer and also of publication time only when printing facilities of this kind became available. Census computerisation in India will never reap the credit of timely publication until printing facilities straight from the computer print-out, are commonly available.

Let us now take up several other considerations which will have a large say on the extent to which the computer should be used in the Indian Census. The Indian Census publishes its provisional population figures, sub-classified by rural-urban, male-female, literate-illiterate and one or two other characteristics for each taluk, sub-division, district and state within a fortnight of the conclusion of the enumeration. These provisional totals are fairly accurate and have seldom deviated by more than a small fraction of one per cent of the final population totals. This work is done entirely by hand. This is not only a matter of legitimate pride for the Indian Census but insulates it at the start against any suspicion of tampering which is a very important consideration in this multi-lingual, multi-ethnic, multi-cultural country. Speed in processing and publication of census results in India is as important for credibility as for ensuring their wide use and acceptance. The computer could not hope to compete with the speed with which the first results are published at every census.

Secondly, since 1951, the Primary Census Abstract has come to stay as the single most important data source for a variety of planning and research con-

cerns. This Abstract swelled to 35 columns in 1971 from about 27 in 1951. This information is finalised, printed and published from the level of the block and village upwards within two years and a half of the conclusion of the Indian Census. This particular piece of work and the speed with which it is accomplished to the last detail is possible only through manual processing. The computer will never be able to compete with manual tabulation in this area. And the Primary Census Abstract has become indispensable at all levels.

Thirdly, the concept of area planning demands that in addition to the Primary Census Abstract, the A and C series of Tables be available preferably on a 100 per cent count basis and that as soon as possible. Since the number of units and the classification cells in the A and C series are large but reasonably uncomplicated for sorting, manual tabulation for them may be more rewarding and time saving than computer tabulation.

These leave us mainly with the B and D—the Economic and the Migration series. In addition, should the census revive some kind of household economic tables following the 1961 Census precedent, there will be the household economic series. These along with the Establishment and Housing Tables are the kind of tables which will very naturally demand computer handling and processing, for manual tabulation has so far made limited success of these areas. The quality of census tables in these areas has so far been much too inadequate for purposes of national accounting. In these areas again will arise complicated problems of sample-taking, the designing of cross-tabulations, the tabulations of complex details, the matching of samples, the reconciling and curing of what at first sight will seem to be garbage. Computer programming in these areas will need very careful adjustment with manual tabulation programming, this demands that a whole array of steps be worked out, well in advance, about how the arrangements for manual tabulation will mesh with computer tabulation both in time and space and in respect of different personnel and offices handling them. Separate but concurrent printing and publication programmes will have to be worked out.

The basic question to be kept always in mind is how much of the data will bear computer tabulation in respect of their quality and detail. For computer will help improve the quality of data to the extent that they will bear disaggregation.

Stall D'Souza

I. Introduction

THE 1974 Census of Bangladesh is presented here as a case study that may help the deliberations of this Symposium on the subject of large scale computerisation of the 1981 Census. Though the questionnaire to be utilised is not the focus of this workshop, it seems to me that some consideration should be given to the objectives of The census to be taken in 1981. The 1872 Census Report of Bengal justified the need for a census as follows :

Without precise information regarding the number of people, there was felt to be a sense of inconvenience in the administration of Bengal. Without information on this head, the basis is wanting on which to found accurate opinions on such important matters as growth and rate of increase of population, sufficiency of food supplies, the incidence of local and imperial taxes, the organisation of adequate judicial and police arrangements, the spread of education and public health measures.

The earlier censuses constituted an integral part of a process that helped the colonial power to maintain its dominance and to obtain 'local and imperial taxes'. Within the context of India today, especially with the availability of computer facilities, the census operation is a central area of the statistics system of the country, that must be geared towards the goal of development. In terms of population, India is the largest country, where censuses are taken every 10 years. China, the world's largest country, does not take regular censuses. Its developmental processes are made possible by the utilization of local data collection at the 'commune' level.

The United Nations Research Institute for Social Development (UNRISD) has set up a set of 100 development indicators. A table on the availability of data on these indicators, shows that the availability relates to 85 of these indicators in the case of India but this is true in respect of only 54 indicators for

Pakistan, of which Bangladesh formed a part till 1971 [11]. This relatively favourable position of India does not mean that the available data are necessarily of good quality or of adequate usefulness. This paper discusses the case of the 1974 Census and makes suggestions for the 1981 Census of Bangladesh in the first two parts. In the discussion that follows the status of the Census data is considered in the context of development statistics.

According to the 1974 Census of Bangladesh, its population on March 1st was 71.48 millions. A post enumeration check revealed under-enumeration, in the light of which the population was adjusted to 76.40 millions. The estimated population of January 1st, 1977 is 82 millions. Bangladesh thus ranks 8th amongst the world's largest populations.

Apart from its size, there are a number of special features of the population of Bangladesh, which differentiate it from most other large countries. Bangladesh is one of the world's most crowded areas with an estimated density of 531 persons per square kilometer (March 1974). Population pressure such as this exists, possibly, in no more than a few small island nations. None, perhaps amongst even these, are as dependent on the land for their livelihood as Bangladesh is. Countries with population densities in excess of 200 persons per square kilometer usually have a high proportion of the population living in metropolitan or urban areas. In contrast, 91.2 percent of the population of Bangladesh live in rural areas.

As compared with the major regions of the world, Bangladesh has an unusually high dependancy ratio of 105, resulting almost entirely from an exceptionally high proportion of persons below 15 years of age (48.1 percent).

Marriage in Bangladesh is near universal, particularly for women. According to the 1974 Census, somewhat less than 1 percent of females had never married by age 30. Though for males the never married rate is higher, still little more than 1 percent fail to marry before age 50.

Age at first marriage is younger in Bangladesh than in most societies. The singulate mean age at marriage derived from the 1974 Census is 15.9 years for females and 24.0 years for males. A sample, restricted to ever married women, interviewed in the Bangladesh Fertility Survey yielded a mean age at first marriage of 13.6 years for urban women and 12.4 years for rural women.

The Bangladesh literacy rates for the population aged 5 years and above are 24.3 percent for both sexes, 32.9 for males and 14.8 percent for females. For the urban areas these rates are 44 percent for both sexes, 52 percent for males and 33.3 percent for females.

II. An Overview of the 1974 Population Census

Administrative Background

The first census of population in the area now comprising Bangladesh was taken under the British rule in the year 1872. The next census was held in 1881. Since then, the census of population in this area has been as a rule taken every ten years. The last decennial census of population was, therefore, due in 1971. While most of the plans for conducting the Census were ready, the attempt to take the census in 1971 had failed owing to the political instability and the subsequent war of liberation.

The new Government appointed a Census Commissioner on 1st July 1972, and entrusted him with the task of conducting the first census of population in the newly independent country. Having considered the unsettled situations arising from the war of liberation, the urgent problem of rehabilitating the estimated 10 millions homeless people affected by the war, the job of reconstructing the transport and communication system and the overall need of improving the law and order situation, the Government decided to conduct the First Population Census of Bangladesh in February 1974.

Under the overall census plan, a Census Advisory Committee (consisting of representatives from all the Ministries, Divisions, Universities and other Government and Non-Government Research Organisations) was formed to solicit their views as users on 20th July, 1972. The Committee was to advise the Census Commission on the methodology of taking the Census, drafting forms/questionnaires, the tabulation plan, processing of census data, etc. The newly constituted Census Advisory Committee considered the 1971 plan of the Census, according to which information was to be collected on only eight population characteristics in the "Big Count" of population, leaving other items, including housing, social, economic and demographic characteristics of the population, to be collected by a nation-wide sample survey immediately after the population count. The Census Advisory Committee, having considered the various developmental needs of the new nation, advised the Census Commission to conduct the following censuses :

(a) the census of housing; (b) the census of establishment; and (c) the census of population. The Census Commission, however, had to undertake one more census, namely, the Census of Handloom Industries in July 1973 under the directives of the Prime Minister.

The methodology of taking the above mentioned censuses, the items to be enumerated, the tabulation plan and the data processing methods were all worked out in detail by the Census Advisory Committee.

Field Operation

CENSUS PRETEST. A pilot census was conducted between 27 November and 6 December, 1972 in 10 urban and 30 rural enumeration blocks, which were purposefully selected from 5 administrative districts to evaluate (1) Questionnaire design, (2) Enumeration block size and work loads of enumerators and supervisors, and (3) Methodological questions regarding training of enumerators and supervisors.

The officers of the census headquarters and field officers drawn from the district administration were involved in this census project. The census material and data obtained were analysed and provided valuable information for the final questionnaire design as well as for the planning and operational aspects of the census.

TRAINING OF CENSUS FIELD OFFICERS. The training of field officers was undertaken in three phases. In the first phase all the District Superintendents of Census (DSCs) and Census District Officers (CDOs) were trained by senior officers of the census headquarters on the 14-15 January, 1974. In the second phase Census Charge Superintendents were trained in Groups at 10 district headquarters in two-day training sessions held during 18-28 January, 1974. The training was conducted by the Deputy Census Commissioners with local assistance from the respective DSCs and CDOs. In the third and final phase, the Circle Supervisors and Enumerators were trained in two-day sessions at the "Thana Headquarters by the respective Charge Superintendents. The CDOs supervised the training which was conducted during the first week of February, 1974.

Appropriate training manuals for enumerators and supervisors were prepared and distributed after adequate explanation regarding their use. Rural area maps

at "mouza level" were obtained from the Malaria Eradication Program. Urban area maps were prepared at the Census Commission using sketch maps from respective municipalities.

ENUMERATION. The enumeration took place from 10 to 28 February, 1974. Household lists of the 1973 housing census were used. A final round was made throughout the block by the enumerator on the night of February 28 until the dawn of March 1, 1974; this was termed the Census Night. The reference date of the Census was thus March 1, 1974.

Bangladesh nationals living in foreign countries, including those working in Bangladesh diplomatic missions abroad, were excluded from the count. Members of foreign diplomatic missions located in Bangladesh were also excluded. Foreign nationals resident in Bangladesh throughout the enumeration period were however, included.

The decennial censuses of population in the area now comprising Bangladesh, had, since the Census of 1941, been taken on a *DE JURE* basis; every person was counted on the basis of normal residence regardless of where he was during the enumeration period. In the 1974 Census a system nearer to a *DE FACTO* basis was introduced. 'Usual residents present' and 'usual residents absent' who would return by 28 February, 1974 were enumerated at the usual place of residence but 'visitors' who would not return to their usual place of residence were enumerated at the place of visit on the Census Night.

POST-ENUMERATION CHECK (PEC). A post-enumeration check was conducted during the month of April 1974. A sample of 0.50% of the census blocks was selected systematically from each of the 66 census districts. A total of 482 census blocks were thus selected. The re-enumeration operation was conducted in all these blocks. The other two operations ran into difficulties. Census records could be found for only 119 out of the 482 sample blocks. These 119 blocks, however, did not represent all the 66 census districts; 14 census districts were left out. It was decided to match data for one rural sample block from each of the 48 census districts for which enumeration records were available. For the urban areas, the matching was conducted for 11 sample blocks from the 4 major cities of Dacca, Chittagong, Khulna and Narayanganj. The under-estimation involved in the total count for the 4 major cities was 19.3% and 6.5% for the rest of the country. On this basis, the population of Bangladesh on March 1, 1974 was readjusted to a figure of 76.398 millions. Due to the small number of blocks matched, the results of the PEC remain of limited value.

TALLY SHEETS FOR PRELIMINARY CENSUS FIGURES. Each enumerator was asked to prepare a summary for his block, after completing the enumeration, by compiling the data on a 'tally sheet' for the following items : (i) total population of males and females; (ii) population by age groups 0-4 and 10 and over for each sex; (iii) number of literates for each sex; (iv) labour force classified by "working" and "Looking for Work" for each sex; and (v) number of cultivators by sex.

He was also asked to give in the summary of the tally sheet the total number of houses and households he had counted in his census block. On these summary returns, supervisory officers upto the level of the Census District Officer prepared summary figures for their respective areas of jurisdiction.

On the basis of the results received telegraphically by the Census Commission at Dacca from the District Officers, preliminary census figures of population by sex for each administrative sub-division were released in June 1974 as a press handout. These preliminary population figures, along with a few details regarding area and density of population till the sub-divisional level, were subsequently published in the Census Bulletin No. 1. The second bulletin *provided final population* figures by sex for all administrative areas down to thana level on the basis of tally sheets. These tally sheets were checked at random for compatibility of the summary figures with the data on enumeration slips for census blocks. This procedure was adopted to meet the urgent demands of development planners, workers and other users. The third bulletin presented population by sex, literacy and household for Unions which are the smallest administrative units of the country. It also presented the same information for urban localities separately. This bulletin was prepared from the same tally sheets which were used to compile the tables of the first two bulletins. A list of villages for each district was also compiled from the data of tally sheets on sex, literacy and number of households in each village.

Tabulation

DATA PROCESSING. Computerised data processing has been resorted for the 1974 Census. A 100 per cent tabulation of both rural and urban had been envisaged in the master-plan but on account of delays in the arrival of the "encoding machines", it was decided to provide sample tabulations of the census data. Due to administrative and logistic difficulties, several census slips were,

however, lost or misplaced and sampling had to be done from the "available slips" and consequently an unknown bias is retained in the results.

Delays in processing the data have been caused by changes in the administrative, technical and advisory setup of the computer centre, and subsequent technical changes in editing and processing the data. Individual programs that had been prepared previously were replaced by the COCENTS package program. Preliminary tabulations for the whole country were thus available for all the tables only at the end of May 1976.

TABULATION PLAN. A set of 24 tables has been sent for publication. An effort was made to ensure internal consistency of the tables. A study of the tally sheets has shown that while the totals were reliable, characteristic data presented there required serious editing. Characteristic data tables, to be published, replace results published in earlier census bulletins. The tables are presented upto the district level.

The sample design followed for tabulation was a systematic selection of 5% of the available rural blocks and 20% of the urban blocks from each of the census districts with the exception of 3 sub-divisions of the Chittagong Hill Tracts where 100% sampling has been done. The tabulations in this volume have been obtained by "blowing up" the sample totals using appropriate weights for the rural and urban sample of males and females separately,

The multiplying factors for urban samples other than those from the Chittagong Hill Tracts should have been around 5, since a 20% sample of blocks was selected. In fact, the median value of the weights for the urban male sample is 5.35 and for the urban female sample is 5.19. On the other hand the range of values varies from 1.41 to 42.81 for males and from 1.41 to 42.53 for females, indicating that in some areas many urban slips were not available. These areas include the census districts of Chittagong Sadar North and South, Narayanganj, Dacca Sadar and Tangail. For the rural areas since a 5% sample of blocks was drawn, the weights should have been approximately 20. The median values of the weights of samples for males is 20.88 and females is 20.86. Here too, many slips were missing and the range of weight values varies from 15.93 to 55.30 for males and 10.00 to 55.65 for females. The areas with the largest number of missing slips are Jaipurhut, Goalunda, Rangpur, Kurigram and Noakhali Sadar Census districts.

As mentioned earlier an unknown bias has been introduced into the results

due to the "non availability" of slips. An important factor affecting the results is the sample design itself. The selection of sample blocks instead of households or individuals from *each* enumeration block is probably adequate for estimates of "characteristics" in rural areas generally. Small villages with special characteristics could, however, be omitted by this sampling procedure. For urban areas the design has serious defects. Urban blocks are often internally homogenous but with great heterogeneity existing between blocks. Care must be taken in the use of data from tabulation cells with few sample points. Thus while national level estimates will not be greatly affected by the design, estimates for small administrative areas are less stable. The tabulation plan is a considerably reduced version of that envisaged in the master plan since the results for detailed tabulations would in fact have little meaning.

The use of tabulations has meant that a slight discrepancy exists in the totals of various tables due to rounding-off-error, depending on the number of times such a procedure is used within a particular table. The COCENTS program ensures that figures add up within any table. Further, the weights used were corrected to only two places of decimal and hence the "blown up" sample totals inevitably differed slightly from those published in the Census Bulletin No. 2. The total population of Bangladesh in Tables 1 and 2 obtained from Bulletin No. 2 is 71,479,913. In Tables 3-10 the population total varies between 71,477,913 and 71,477,635, thus remaining within a variation upper limit of 1500. Given the various limitations of the data, the results should be deemed consistent at the thousands level.

III. Some Proposals for the 1981 Census

In this section some proposals for the 1981 Census are made on the basis of the experience of the 1974 Census.

Assessment of Data Needs

Of prime importance is the necessity of having a clear idea of the data requirements. Population censuses are "primary sources of the basic national population data required for planning and administration, as well as for many aspects of economic and social research" [12]. A distinction has to be made between the data requirements for planning at the national level and large areas such as districts and sub-divisions, and those for small areas like thanas and unions. Carefully designed sample surveys would provide national level "characteristics data", though a frame would have to come from a census. For small areas,

sample surveys cannot provide accurate information. At the planning stage, then^ several Important decisions have to be taken, and the setting up of a *Technical Advisory Committee* (TAC) for the census is of prime importance.

The TAC, however, should meet at every stage of the conduct of the census, until the final publication of the census documents. An efficient TAC helps maintain the standard of the census and provides the administrator with the scientific input he requires. The 1974 Census TAC was constituted in July 1973 for a period of one year. The role of the TAC virtually terminated after the overall census plan was adopted although a sub-committee was consulted regarding revisions in data processing thereafter. A new TAC should be brought into existence to examine the various aspects of the 1974 Census data. An immediate decision would be about date of the next census. In keeping with the census history of the area, an appropriate year will be 1981,

Sampling

In the 1974 Census, sampling has been used at all the stages. No further analysis for the data collected is envisaged. Clearly sampling at the enumeration stage would have been a more efficient and a less expensive process. The detailed socio-economic characteristics necessary for planning can be obtained by carefully designed sample frames adapted to urban and rural areas. Particular specifications for the 4 largest cities of Dacca, Chittagong, Khulna and Narayanganj could be introduced.

The 1974 Census Organisation was also responsible for a housing, establishment and handloom census, besides the population census. Further a retrospective survey and a post-enumeration check were conducted. A large field effort was thus necessary, but the data realised from the censuses have not been adequately tabulated. The TAC should deliberate on the advisability of undertaking so many field operations. A count of the actual number of households associated with the housing census, *would* be prepared immediately before the next population census.

A PEC on a few questions, like coverage by age and sex, could be undertaken simultaneously as the Census. The socio-economic sample survey could incorporate a few items on mortality and fertility as well as some details on housing. The work of the Census Organisation will thus be trimmed.

Post Enumeration Check

As a result of the PEC conducted for the 1974 Census nearly 5 million persons were added to the total census count of March 1, 1974. The earlier 1961 PEC was also poorly executed. A check in 143 urban and 56 rural census blocks gave an over-enumeration of 1.6%. Other checks based on demographic analysis of age data, on the other hand, showed under-enumeration ranging from 5.1% to 8.9% [1], [-5]. In view of the importance, especially of the total figure, to national planning, it seems vital to execute a carefully planned post-enumeration check soon after the census count. The PEC should lay great stress on all aspects of undercoverage—numbering of houses, households, coverage inside households, mapping errors, and floating population [4]. In this connection, the use of women enumerators for a PEC should be seriously considered.

Questionnaire Design

The questionnaire used in the 1974 Census contained several items which will not be tabulated. Clearly, a great loss of time and finance has resulted, apart from the usual diminishing returns obtained from a cumbersome questionnaire. Several consultative meetings have been held with the demographers and experts in the country as well as with visiting experts. A fair degree of consensus, regarding the content of the next population census questionnaire, has been reached. The schedule should contain only about 7 or 8 simple characteristics: name, relationship to head of household, sex, age, marital status, religion and literacy.

The use of a limited number of questions will facilitate a major design change in the format of the census slips; a multiple enumeration type of schedule can be adopted. As a result, data entries for approximately 1000 persons in an enumeration block can all be made in a single booklet of about 40 pages. About 90,000 such booklets would cover the country assuming the population to be counted in 1981 to be around 90 millions. In the 1974 Census, storage space for the 71 million slips provided a serious problem, besides the tabulation and other logistic problems involved in moving the data from the field to the Data Preparation Centre in Dacca. The use of the booklets will thus circumvent a serious management problem. Besides, if copies of the booklets are maintained at the union level, the beginning of a nation-wide registration system can be set up.

The questionnaire for items to be obtained by a sample survey will, of course, be more detailed. Care should be taken that priority is given to items that make international comparisons possible as well as to items that have been consistent features in the census history of the country.

Training and Supervision

A key element in a census is the training imparted to enumerators. A simplified questionnaire would reduce enormously the task of training over 80,000 enumerators. The 1974 Census experience has shown that the training at the thana level left much to be desired. The Census organisation contained a small nucleus of officers trained in census procedures but a great majority of personnel were recruited from the existing administrative machinery. These officers, with little previous census experience, normally had other important tasks to perform and handling the census was just an additional chore. The enumerators themselves who were teachers or other village officials were given a two-day "theoretical training" with no field training. Hopefully this situation can be remedied in the next Census. The merger of the Census Commission with the Bangladesh Bureau of Statistics (BBS) could result in a nation-wide field structure of about 300 trained officers and staff becoming available for the census operations. It will be vital, however, that during the conduct of all census operations, a fair degree of autonomy is enjoyed by the Census Wing of the BBS. This must include logistic support if the census operation is to be successful.

The Bangladesh Fertility Survey (1975) has shown that women interviewers can be recruited and perform well in the field. The problem of under-enumeration for infants and women in certain age groups, would be best tackled by trained female interviewers.

Greater 'people participation' should also be encouraged. It will be important to mobilise all the village level organisations for census operations. The perception of such organisations of the utility of an accurate population register at the village level will go a long way in ensuring that good data is collected in the field.

Though maps were prepared at great cost for the 1974 Census, they were hardly used and few of them have been retained. Appropriate use of maps increases enormously the quality of data and provide an independent check of houses covered. In urban areas maps were utilised to a greater extent than in rural areas. Maps are required for delimitation purposes: and their updating

should be taken up immediately. The use of available serial photographs in the preparation of maps should be investigated as this can mean great savings in time and money, and improvement of the quality of the data. Appropriate coding of geographical areas should be undertaken since computerised tabulation is used. In this connection, it should be stated that census delimitation should coincide with existing administrative demarcations. The creation of artificial census units makes field identification difficult and creates problems when separate tabulations are required for administrative jurisdictions.

Appropriate *supervision* of enumerators is vital to the collection of good, quality data. For the 1974 Census, the proportion of supervisors to enumerators was 1 : 16. For the next census this ratio should be increased to an optimal 1 : 5. The use of union level officials should be considered in this regard. It is illusory to assume that the mere use of sophisticated equipment, such as *mark-sensing*, will reduce error. Besides the dependence on outside sources that it creates, and the level of technical correctness it requires in the marking of forms, the substitution of manual operations by machines should be carefully considered. Manpower is one of the resources of the country and should not be lightly brushed aside; rather, appropriate training should be given so that availability of such a resource is best used.

The concept of supervision is not confined to the immediate "enumerator-supervisor" relation only, but should extend throughout the framework of the census organisation. This can be done by adequate means of communication. The reliance on the mailing system in 1974 bore little results and more effective measures like the use of radio should be considered to link the central office within the various parts in the field.

Yet another level of training should be considered viz., that of the officers of the Census Wing. The 1974 Census included serious advisory inputs, which were mainly available after the Census had been taken. The timeliness of advice is vital for the efficient conduct of the next census. An important area of consideration for the TAC at the planning stages of the next census is in the training of Bangladesh personnel, and the amount of advisory services necessary. Advisory inputs at all phases of the census should normally taper down for the next census. This can be achieved by training the present staff, abroad if necessary, and by better utilisation of the specialists in the country.

Tabulation

The tabulation plans for the 1974 Census were greatly curtailed for reasons mentioned earlier. About 3.5 million census slips were processed. It is important to realise, however, that all the raising factors utilised for the census tabulations for rural and urban areas, have been obtained from "tally sheets" which were manually totalled. A careful comparison for blocks of "enumerators tally" versus a "tally at the Census Commission" is shown in Table 1.

TABLE 1—COMPARISON OF ENUMERATORS TALLY AND OFFICE TALLY
(for 84 Blocks)

	<i>Enumerator^us Tally</i>	<i>Office Tally</i>	<i>Percentage of Discrepancy</i>
Number of Persons			
Both sexes (BS)	65,368	65,310	+ 0.09
Male (M)	34,074	34,096	- 0.06
Female (F)	31,294	31,214	+ 0.26
Number of age groups			
Both sexes			
0-4	11,086	10,827	+ 2.39
5-9	12,207	12,131	+ 0.63
10 and over	42,075	42,352	- 0.66
Males			
0-4	5,601	5,476	+ 2.28
5-9	6,318	6,209	+ 1.76
10 and over	22,155	22,411	- 1.14
Females			
0-4	5,485	5,351	+ 2.50
5-9	5,889	5,922	- 0.56
10 and over	19,920	19,941	- 0.11
Number of literates			
Both sexes	13,208	14,460	- 8.66
Male	9,095	10,051	- 9.51
Female	4,113	4,409	- 6.71
Size of Labour Force			
Both sexes	20,961	18,156	+ 15.45
Male	16,431	17,476	- 5.98
Female	4,530	680	+566.18
Number of Agriculturists			
Both sexes	11,551	12,651	- 8.69
Male	10,690	12,212	- 12.46
Female	861	439	+96.13
Number of Households	11,238	11,117	+ 1.09

It is clear that while totals for "number of persons" showed small differences, "characteristics totals" differed a great deal—by as much as 566.18% for "women in the labour force". This discrepancy shows that enumerators had difficulties in comprehending for more complicated questions. Table 2 compares differences between the totals given in the second bulletin and computerised output based on sample data. Again we see large discrepancies for characteristics data. A small portion of the discrepancy could be due to the sampling procedures used,

TABLE 2—COMPARISON OF BULLETIN 2 AND COMPUTER TABULATIONS FOR SOME "CHARACTERISTICS DATA" OF THE 1974 CENSUS

Sources	Literate a? percent of population 5 yrs. and over			C.L.F. as percent of total population			A.L.F. as percent of C.L.F.		
	B.S.	M	F~	~B^S.M~F	~B^S.M~F	~B^S.M~F	B.S.	M	F
Bulletin 2	22.2	29.9	13.7	35.2	50.6	18.7	55.4	67.8	19.4
1974 Computer Tabulations	24.3	32.0	14.8	28.7	53.0	2.5	77.2	77.5	69.8

CLF: Civilian Labour Force

ALF: Agriculture Labour Force

The utilisation of manual facilities for preliminary bulletins should be encouraged. Some decentralisation of tabulation facilities may be considered in this connection. The results shown above would indicate that population totals prepared manually could be made available quickly and accurate sampling of characteristics data will mean that overall data processing tasks will not be substantially larger than for the 1974 Census and the additional censuses and surveys that were conducted. Equipment bought for the 1974 Census was expensive and, in many cases, defective. Of the 90 encoders purchased only 20 are in operation. The role of agencies like NCR in dumping defective or out-of-date equipment should be examined by the TAG and demands for the purchase of new equipment should be carefully examined.

IV. Discussion

The proposals for the 1981 Census that I have suggested include the limitation of the 100% Census count to only a few items. The importance of large scale computerisation inserts itself at this stage *through the use of linkages*. If appropriate identification is provided for individuals and households, population census results can be linked with other statistics such as the housing statistics

and economic statistics. National Sample Surveys (NSS) and other sample surveys related with census data would thus become more useful provided, of course, that they use the same or similar identification, concepts and definitions. A report of a group of experts meeting in Geneva, December 1-3 1975, has the following :

Most development plans today announce policies which aim at reducing poverty, improving distribution, cutting down unemployment and under employment, improving rural health service, reducing regional disparities and reforming land-tenure systems. It is discouraging to note that in general the related data are among the poorest in the whole field of development statistics. Despite international efforts and important programmes at the national level, the actual measurement of these factors in conjunction with development planning has not made much progress during the past 20 years [10].

The focus then should be that population census results be inserted within the development strategy of the country. Large scale computerisation plays an important role so that planners can have at their disposal up-to-date data on a variety of fields relevant to decision they have to make. In Malaysia it is reported that census data is stored at the University Campus and data requirements by the Department of Social Welfare, for instance, are readily met*.

The place of 'people participation*' at the village level has been mentioned. The establishment of a 'population register' [9] that could be computerised should be given serious thought. Development can only come from the grass-roots perception of the country's problems which relate to land holdings, population birth rates etc. Short-sighted campaigns to limit population have failed to take into account that the high level planners* view of the problem at the macro-level may not coincide at all with the perceived nature of the problem to the Individual at the micro level. The limitations of planning at the macro level are now being recognised and it is generally agreed that local level requirements cannot be ignored. The next five year plan will concentrate on "block level" planning to ensure 'minimum human needs'. Attempts will be made to transfer resources to local 'target' groups [7]. A computerised data system with appropriate linkages and identification would assist planners in the deter-

*Adnan, B.H.A., Director General of Social Welfare, Malaysia (Communication at Asian Regional Seminal, International Association of Trained Social Workers, Oct. 11-16, 1977, New Delhi).

mination of requirements and the allocation of resources. The development of appropriate socio-economic indicators by research institutions would be greatly facilitated. In-depth village studies are now being undertaken in various parts of the world. The Institute of Development Studies, Sussex, for example, has compiled data on village studies to assess the labour situation in various developing countries [2]. Census data accurately collected at the village level often forms the basis of serious studies and selection of the villages for the purpose. Experimental designs set up to study the progress in a village require appropriate 'controls'. The use of large scale computerisation would make possible the 'matching' of a few characteristics so that the required control villages could be carefully chosen.

The usefulness of statistics is greatly reduced if they are not available when needed. Tabulated data from censuses are often characterised by large delays. Plans based on such data can hardly hope to be effective. The logistic problems in conducting a census in a vast country like India are enormous. The introduction of large scale computerisation could, of course, cause bottlenecks in the transfer of data from the questionnaire to magnetic tape and at the programme level for tabulations. An efficient decentralisation of computerised facilities should, therefore, be envisaged. The deployment of a dozen or more regional computercentres that would supply 'clean' data tapes to a central computer centre, should be investigated. The use of encoders that allow direct transfer of data from the key punch stage to discs or magnetic tapes without the use of computer cards should be envisaged.

Computerised cards often require special storage facilities and create a variety of problems. Verification of data is possible with the use of encoders whereas one step namely the transfer of information from computer cards to tapes is eliminated. The use of package programmes such as COCENTS would facilitate rapid tabulations. Such package programmes however, have some inflexibility including large core requirements. These are technical details calling for computer expertise. The passage from tabulated data to printed results is another time consuming operation. The installation of 'offset' printing machines that work directly on computer output could cut down enormously the time required for publishing a census document.

The quality of the data should be an area of serious reflection. Careful training of interviewers and supervisors and efficient editing procedures go a long way to diminish error. When sampling is used it should be remembered that

m many developing countries, non-sampling errors are often much more serious than sampling error. A recent example of an attempt to assess the "balance between sampling and non-sampling errors in Spanish Official statistics" has been made [8]. From its results, it is concluded that it would not be worthwhile "to use more resources to reduce sampling error until a drastic decrease in non-sampling errors was achieved". The use of women interviewers could reduce non-sampling errors in the data on women. Decentralisation of data processing may mean problems in maintaining uniform quality at national level. Careful and uniform rules regarding editing of the data at the field level as well as the use of 'imputation' and 'hot decks' can improve the quality of the data. Clearly large scale computerisation thus has an advantage over older and slower methods of manual editing.

The various papers and comments by the learned participants will undoubtedly throw more light on the advantages and disadvantages of large scale computerisation in the Indian context. It is my sincere hope that the reflections set out in this paper against the background of work done in Bangladesh will be a useful complement to our discussion here. I would like to record my indebtedness to my counterpart, Mr. S. Rahman, Deputy Census Commissioner, Dacca, Bangladesh for his contribution to the preparation of this paper.

Refereuces

1. **Bean, L. *eta'*, 1968, *Population Projection for Pakistan, 1960-2000, Pakistan Institute of Development Economics.***
2. Connell, J. and Lipton, M., 1977, *Assessing Village Labour Situations in Developing Countries*, Oxford University Press.
3. Khan, A. R., D'Souza, S. and Lewis, L., *Current Population Trends, Factors influencing Fertility, Projections to year 2000* (unpublished background paper prepared for the Population Control and Family Planning Division—Bangladesh, for the draft IInd Five Year Plan).
4. Krotki, K. J. and Hashmi, S. S., 1962, Report on a census enumeration, *Pakistan Development Review*, 11(3), 377-405.
5. Krotki, K. J., 1963, Population size, growth and age distribution; Fourth Release from the 1961 Census of Pakistan, *Pakistan Development Review*, III(2), 219-305.
6. Rabbani, A. K. O., D'Souza, S. and Rahman, S., 1974 *Census Estimates of Fertility Levels in Bangladesh*, (Presented at a Seminar on Fertility at Cox's Bazar, Dec. 21-23, 1976).
7. Raj Krishna, *Key Note Address at Regional Asian Conference on Family as a Unit of Welfare in National Planning*, Oct. 17-21, New Delhi.
8. SanchezCrespo, J. L., Balance between Sampling and Non-Sampling Errors in Spanish Official Statistics, *Contributed Papers*, 39lh Session of the International Statistical Inst., pp. 841-878.

9. Schubnell, H. and Herberger, L., Prospects for census programmes in view of the growing use of sample surveys and computerised registers in Public Administration, *Contributed Papers*, 39th Session of the International Statistical Institute, pp. 898-904.
10. UNRISD, Improvement of development statistics, *Report of a Group of Experts*, Dec. 1976, Geneva, Report No. 76.4.
11. ____, *Research Data Bank of Development Indicators*, Geneva 1976, Report No. 76.1.
12. **United Nations** Secretariat, The availability of demographic statistics **throughout the World**. *The Population Debate: Dimensions and Perspectives*, **Papers of the World Population Conference, Bucharest 1974**, 1, pp. 371-384.

Ashish Base

THE experience of the 1971 Census was not happy in respect of speedy publication of results. The one per cent sample tables came up quite early but very little followed. Part of the fault lies not in computer technology but in printing technology. The Government presses are under the Ministry of Works and Housing and not under the Home Ministry under which the Registrar General's Office functions. I had warned well before the 1971 Census that computer technology without a matching printing technology will be fruitless and this is what exactly happened in the 1971 Census.

Another point made around 1971 was the need for putting the census data on tapes and making available these tapes to recognised institutions and scholars through the ICSSR for detailed cross-tabulation according to their requirements. But nothing was done in this regard. The Office of the Registrar-General with its limited staff in the Computer Unit could not handle the ever increasing demands made by various organisations and individual scholars. Further, the British tradition of maintaining census data confidential was followed. As a result, an odd situation has arisen where research can be done only on the basis of the obsolete 1971 Census data which may hopefully be available by 1981. The American system of selling data through tapes was not adopted nor any arrangement made to make data available except, in a limited manner, on a personal basis.

In several quarters, there is a feeling that computerisation has in fact delayed tabulation of census data instead of expediting this work and it is quite possible that the demand for large-scale computerisation is opposed by some higher authorities in the Government. This Symposium should be able to make a convincing case for large-scale computerisation of the 1981 Census data, and it is not a day too soon for making such a case. If it is decided to go in for large-scale computerisation, even the format of the census questionnaire will have to be

changed and provision will have to be made right now for the hardware and software requirements involved in such computerisation.

The case for large-scale computerisation cannot, however, be taken for granted. After all, the Census is not involved in sophisticated statistical exercises. The prime requirement of the census is speedy presentation of a set of simple tables, suitably cross-classified by rural/urban residence, sex, broad groups, etc. This can be done by old fashioned mechanical devices in a decentralised manner. And there are cases (e.g., compilation of the total population figures), where even hand tabulations could be done without any loss of speed provided these operations are decentralised. Both in the 1961 and 1971 Censuses, the total population of India was given out in record speed and this was done without the help of electronic computers. On the other hand, it can be argued that detailed cross-tabulations of data collected through 20 or 22 questions and sub-questions for a population of about 700 million is an impossible task if electronic computers are ruled out.

But even on the electronic computers if one were to undertake 100 per cent cross-tabulation of data, the task will indeed be formidable. Recourse will have to be taken to sampling and there is no reason to believe that a properly drawn sample will give less accurate results than one hundred per cent tabulation, even if the latter were technically feasible.

The question which arises at this stage is : If it is decided to introduce sampling is it necessary at all to collect all data in respect of all persons or, in other words, can the census canvass a very small questionnaire for all persons and a detailed questionnaire on a sample basis at the enumeration stage? The other method would be to have the same questionnaire for all persons but having collected the data, go in for sampling at the tabulation stage? This issue must be thrashed out and the Symposium should come out with a clear answer. It is often argued that in view of the limitations of the census enumerator, sampling at the time of enumeration should be ruled out and in fact, such a sampling procedure instead of saving time and money will ultimately mean waste of time and money,

A third possibility is also there. The Census could ruthlessly cut down the questionnaire to say 5 to 6 questions and do the tabulation on 100 per cent basis and forget sampling altogether. The collection of detailed data, it is argued, could be handed over to other institutions and agencies like the Central Statis-

tical Organisation, National Sample Survey Organisation etc., who could be entrusted with periodic collection of data on a sample basis. This procedure would, of course, mean a drastic departure from past Census practices and also a dilution in the importance of the Census work. The critics of this approach, however, doubt the ability of the CSO, NSS and other organisations to so act as an alternative to the census in such a vast country, with their limited experience.

The upshot of the discussion so far is not that no innovation is possible and that the *status quo* should be maintained. There are more than 3 years left for the next Census, the new Census Commissioner has already taken over, and it is certainly possible to put our heads together and come out with concrete recommendations which should receive serious attention at the highest level. It is important to consider in this regard, the administrative and financial implications of the recommendations. There is no doubt that computer technology is a high-cost technology and the financial implications will have to be carefully worked out. I may repeat the suggestion, I made even before the 1971 Census was conducted, that the cost of the Census should be borne by the two five-year plans which cover the decade in question. The Census today is not merely an administrative tool as in the British days but, more importantly, a tool of economic and social planning and policy-making. There is no reason why the Census budget should be grudgingly approved by the Ministry of Home Affairs. The success of formulation, implementation, monitoring and evaluation of our plans, including rolling plans, would depend in a large measure on the reliability and adequacy of census data in respect of demographic, economic and social attributes of the population. The Census Department must, therefore, be strengthened. The technical staff, at all levels, must be substantially increased, the necessary infrastructure for a modern census created and the budget of the Census increased very considerably. This is possible when the Census gets linked to the planning process in the country.

Today there is widespread recognition of the fact that in a country like India, marked by great regional diversities, the macro picture can very often be misleading and the least that one could do in any analytical exercise would be to consider the district as a unit of observation. Playing with data for over 350 districts in India in respect of a large number of demographic, economic and social variables certainly calls for large-scale computerisation of Census data.

To sum up, I visualise the co-existence of hand tabulation, mechanical tabulation and electronic computer tabulation in the 1981 Census, I also visualise

that in view of the long and unbroken history of the Indian Census, continuity should be maintained and the 1981 Census questionnaire should continue to be on the lines of the extended questionnaire of the 1961 and 1971 Censuses with suitable modifications aiming at improvement. It also visualise complete collection of data in respect of all items from all persons but tabulation of data on a sample basis, and yet a vastly improved system of computerisation of data compared to the 1971 Census and I maintain that this improvement is possible only if the Government of India ungrudgingly sanctions more staff and more funds for the Census Organisation. I also suggest that an efficient data bank should be organised by the Planning Commission for its own use and by the ICSSR for use of recognised institutions and scholars. The financial implications of these data banks should also be worked out.

J. R. Rele

THE importance of this Symposium cannot be over-emphasized. The Indian Census has a very rich tradition, and among the official statistics available in India, the Census statistics are perhaps the best. However, for optimum utilisation of these statistics, it is imperative that the Census data be tabulated and made available to the consumers at the maximum speed. This would promote research activities and assist policy-makers. It should be possible to meet this objective through large-scale computerisation of the Census.

However, it should not be always taken for granted that computerisation will expedite data processing. Especially, *in* the initial stages of its introduction, computerisation would experience teething troubles. Moreover, for some of the simple tabulations, it may not prove to be the fastest. The computer may work fast, but it requires gruelling preparations before the data can be fed into the computer. It would require punching and verification of millions of cards, as well as cleaning and de-bugging of the tapes. It is this phase of the computerisation that often becomes quite crucial and demands some serious thinking. We need to carefully consider the necessary facilities and to ensure an optimum utilisation of the existing facilities available in the country.

The process of tabulation is much quicker after clean tapes are prepared. The availability of tapes based on 1 per cent sample for the 1971 Census has been a sign of progress from which we can move further. The tapes can be very useful to get a variety of tabulations for research. We also need to explore how the census publication can be expedited from the data available on the tapes.

One of the ways of improving the performance is through the use of sampling in the census, which is not new for the Indian Census. The possibility of making available, on a sample basis, some special tabulations, which are hitherto not available, may be explored. There is also the possibility of using sampling at the stage of data collection, wherein apart from some minimum basic information, all additional information can be collected through the use of sampling. The imminent danger in this regard lies in biasing the data through selectivity. However, this can be minimised through area sampling which appears most appropriate in the situation.

One of the important recent features of the Indian Census is the use of post-censal surveys to detect the errors of enumeration. The 1971 Census should be credited for presenting elaborate tabulations of the errors ascertained through the post-censal surveys. However, it is well known that the post-censal surveys of the previous censuses were probably themselves subject to gross errors. In this context, special effort should be made to organise the post-censal surveys more efficiently. It is necessary to ascertain the relative roles of sampling and non-sampling errors, in order to minimise the latter which appears to have been crucial in the previous post-censal surveys.

The age-sex data has been an important source of information available from the Indian Census. In view of the tremendous development of technical demography and methodologies for estimation of demographic parameters in recent years, the age-sex data available from the census have assumed even greater importance. Taking into account their utility for research, it is expedient to publish, and make available to research workers, the age-sex data at the national and sub-national levels at the earliest opportunity. It is customary to provide the age data by a number of variables only by broad age groups. To allow its full research potential on the basis of recent methodological advances, it would be extremely useful if the age data are presented in five year age groups by several variables such as religion, language, rural-urban residence and the like.

CENSUS provides a mine of information, whose analysis could throw light on real aspects of the economy and social life. But this requires one to go beyond the routine tabulations conventionally published by the Census, to derive supplementary tables of the census data. This need of supplementary tables was keenly felt in the preparation of the census monographs based on the 1961 Census data. Another need of supplementary tables arises from the fact that often the classification schemes are changed from one to another Census; for example, literacy and educational classification was given by the 1961 Census for the full range of age with 60+ as the terminal age group but by the 1971 Census upto the 35+ age group only. Similar problem of inter-censal comparability arises in respect of the tabulation of migration data, particularly the duration of residence classification. This need of generating supplementary tables is difficult to be fulfilled by manual operation on individual slips. It can be easily met by transferring the data to computer tapes and creating adequate facilities of running and re-running these tapes.

Asok Mitra has drawn attention to the poor quality of the census data, which limits the usefulness of attempting complex cross-tabulations except at certain high levels of aggregation. This should be duly recognised and all efforts should be made to improve the performance of the data collection system. However, at levels of aggregation like the district or the state, some cross-tabulations may throw up meaningful pictures but this too is possible only when these data are available on computer tapes.

Another defect that could be avoided in computer processing of the census data is the one relating to sorting errors, to which manual process is unavoidably subject. Further, manual operation permits only limited editing of the data and leaves little scope for cleaning them properly; both editing and cleaning of the data can be more efficient in computer processing. It is also possible to avoid printing errors if arrangements are made to print the tables directly from computer outputs.

Keeping in view the various problems raised above, I firmly believe that For any scientific analysis of the census data, the routine census tables alone will not be sufficient and it would be necessary from time to time to have supplementary tabulations. I also believe that the extent of errors in sorting and tabulation can be reduced considerably if this work is done with the help of computers. I, therefore, feel that it is the right time to decide about the manner in which the computerisation of the 1981 Census can be done.

Session II: Issues Relating to Sampling

PAPER

V. R. Rao

THE operations involved in the Indian Population Census are of such gigantic magnitude that sampling had to be resorted to for tabulation of data right from the 1941 Census. With the ever increasing need for a variety of information based on the census data, to meet the needs of planning and administration, the scope of sample tabulation has been greatly expanded in the later censuses. Most of the 1971 Census tables were, in fact, generated on a sample basis through electronic data processing (EDP). Sampling and EDP will continue to play a crucial role in the forthcoming census also in order to produce comprehensive and timely results.

Sampling in Past Censuses

2. In the 1941 Census, a 2 per cent systematic sample (known as the *Y* sample) was used for tabulation of age, means of livelihood, and industry. In the 1951 Census, a 10 per cent systematic sample of individual slips was used for tabulation of age and civil condition; and in 1961, a 20 per cent sample of household schedules was used for tabulation of all household economic data. In the 1971 Census a systematic sample of 20 per cent of individual slips in the urban areas and 10 per cent of individual slips in the rural areas was used for tabulating most of the tables on fertility, migration, and economic, social and cultural activities. A 20 per cent sample of census houses (from the house-listing schedule) was used to compile all the tables on housing. Tables based upon sample tabulation were prepared at the level of a district separately for rural areas, cities with population of 1 lakh and above, and the rest of the urban areas. Sampling has also been used for the post-enumeration check since the 1951 Census. A sample of roughly one tenth of one per cent of houses was taken for this purpose in the 1951 and 1961 Censuses. The sampling fraction was one twentieth of one per cent in the 1971 Census.

3. Only items on sex, age, religion, Scheduled Caste/Tribe, literacy, mother tongue and main, economic activity went into the tabulation based on full count. These were used for preparing the Primary Census Abstract for each village and each enumeration block in urban areas. Annex. I gives a more detailed description of the sampling followed in the 1971 Census,

Experience of Other Countries

4. The use of sampling as an aid to census has been in vogue for long in the advanced countries of the world. U.S.A. introduced sampling procedures for the first time in the 1940 Census, Since the 1950 Census, it has been collecting a major part of the census data on a 20 per cent sample. Canada and Israel have followed the U.S. example in the recent 1970 series of population censuses, while the U.K. carried out large-scale sampling studies in 1972 as a prelude to their use in the next census.

5. The important features common to all these cases were :

- (a) sampling was used at the stage of primary enumeration;
- (b) only the most essential information (age, sex, race, and civil condition) was collected on a universal basis;
- (c) information on complex, or low priority, items (employment, occupation, industry, class of work, education, income, other population characteristics and a number of items on housing) was obtained from a sample;
- (d) simple systematic sampling was adopted with the sampling fraction ranging from 20 per cent to 33 per cent. This was found adequate to provide the requisite tables at the level of a country with an average population of about 2,500.

In the U.S. Census of 1950 and the Canadian Census of 1971, the primary unit of sampling was the individual. U.S.A. subsequently changed over to the household/dwelling unit as the primary sampling unit in the later censuses. Israel also followed the same procedure in its 1972 Census. A detailed review of the sampling applications in population censuses of the U.S.A., Canada, Israel and U.K. is given in Annex. II.

Level at which Sample Estimates are Required

6. Before considering the various issues pertaining to the use of sampling in

the 1981 Census, it is necessary at the outset to settle the regional level at which the census tables based upon sample data are required. As noted earlier, the 1971 sample tables were produced at the level of the district. This may not be adequate in the 1981 Census in the light of present emphasis on detailed area planning extending down to the Community Development Block. All the census tables of the 1981 Census will, perhaps, be required for each C.D. Block to meet the needs of planning and development. The tables should be simultaneously available separately for towns with population of 1 lakh and above and other Urban areas in each district.

Census Items for Universal Coverage

7. It is convenient at this stage to identify the list of items which should be covered in respect of each individual of the population and which will find a place in the final results of the census. The experience of the 1971 Census shows that the items of universal coverage need be only a few. These may be confined to age, sex, religion, Scheduled Caste/Tribe, literacy, and broad economic activity, apart from identification particulars such as name and the relationship to the head of household. Mother tongue, which was included in universal coverage during the 1971 Census may not be considered important enough to be retained in this list during the 1981 Census. This item and all others relating to fertility, migration, economic activity and social and cultural characteristics can be conveniently handled through a sample. The actual choice of items in the sample list and the extent of detail in respect of each item depend upon the nature and stage of sampling, the size of the sample as well as the agency entrusted with the work. These are considered below.

Stage of Sampling

8. The most important issue relating to use of sampling in population census is whether sampling may be introduced at the stage of enumeration itself or whether it should be adopted at the stage of tabulation, as in the past censuses. Sample enumeration has obvious advantage of reducing the overall work-load and cost. For example, if the 20 per cent individual slips tabulated in the 1971 Census were alone enumerated for the sample list of items, there would have been a 30 per cent saving in the cost of printing and stationery (approximately equivalent to Rs. 20 lakhs) in addition to a substantial reduction in the enumeration time.

9. The question of sample enumeration was examined by the Registrar General in the past and was not favoured on the following three counts :

- (i) the primary enumerator will not be able to select the sample in clock-wise regularity which may vitiate the validity of the information;
- (ii) the households and persons included in the sample will be required to answer much larger number of questions than the rest; this additional imposition may cause resentment and non-cooperation;
- (iii) printing of individual slips (or household schedules) with different formats for universal and sample coverage will pose serious problems.

While there is some substance in these arguments, the anticipated difficulties of sample enumeration have perhaps, been somewhat exaggerated. The sampling scheme in any census operation has to be most elementary and it is not beyond the ingenuity of the census authority to devise a sampling scheme by appropriate choice of the primary sampling unit and the sampling fraction which can easily be understood and adopted by the primary enumerator. Objection to discriminatory treatment of households included in the sample is not borne out by past experience. The National Sample Survey and other sample survey organisations have been collecting information from sample households without any such problem and the N.S.S. questionnaire is many times larger than the census schedule. Finally, there is considerable experience available from other countries of printing census forms with differential coverage which would be of help in designing our own census schedules in a similar manner. In any event, this matter should be seriously considered in the light of the experience of past censuses, in which 80-90 per cent of the valuable information collected was practically left untabulated. It would be extravagant to collect information without any intention to process it, much less to make use of it.

Sampling Unit

10. Irrespective of whether sampling is used at the stage of enumeration or of tabulation, the question arises as to what is the most appropriate unit of sampling—the individual, the household, or a geographical entity such as the Village. The past censuses of India have, by and large, chosen the individual as the primary sampling unit mainly because of the manner in which primary information is maintained. The individual census slips were available with suitable arrangement according to sex, age and economic activity to facilitate samp-

pling of individuals. It may be recalled that the individual was the primary sampling unit also in the earlier censuses of the U.S.A. and the latest census of the Canada. However, the U.S.A. has changed over to the household/dwelling unit in the later censuses.

11. Choice of the primary sampling unit is dependent as much on prevailing field conditions as on statistical considerations. So far as India is concerned, the household is an integral economic unit and the long experience of the N.S.S. has shown that it is the most convenient unit for studying the demographic and other socio-economic characteristics of the population. It is also seen that the individuals in the census are approached through the household by first listing the households. A household would, therefore, appear to be the most convenient sampling unit under Indian conditions. Institutions in the non-household sector will, however, have to be treated in a slightly different manner.

12. Past studies of sampling applications in the Indian censuses have also favoured the household as the primary unit of sampling. Lahiri and Ganguly (1951), in a pioneering study on the nature and extent of sampling required for estimating population characteristics, came to the conclusion that although from statistical considerations the ideal unit of sampling is the individual person, the household should be preferred in practice as it is operationally more convenient and lends itself to a better control of the quality of primary data.

13. A geographical unit like village, or an equivalent entity, is not likely to be acceptable under Indian conditions as villages happen to be often exclusively inhabited by a particular ethnic or religious group and sampling of villages may not provide satisfactory estimates in respect of small communities and groups in tables compiled at the level of a C.D. Block or equivalent area.

Extent of Sampling

14. We may now consider the extent of sampling required for obtaining estimates of sufficient accuracy. The past censuses have apparently erred on the safe side and taken a rather large sample to ensure that the final estimates are as near as those based upon cent per cent count. What should be the minimum sample size of different primary sampling units for compiling the census tables, with the requisite cross-classification, has never been studied in great depth. The results obtained by Lahiri and Ganguly, however, throw some light on this

question. On an analysis of the data from 12,000 individual slips in an area of about 75 sq. miles, they inferred that a sample size of 22,000 individuals, roughly 2 per cent of the population-of a district of average size, will be required to compile the table on age-sex-marital status. The authors do not, however, indicate how much should be the sample size if households were the primary unit. It is necessary that studies of this type should be organised on a much larger scale to ascertain the extent of sampling required for estimating various population characteristics.

15. Even if one were to go roughly by the findings of the above study and assume that the results would be applicable with regard to other characteristics also, the extent of sampling required to compile tables at the level of a C.D. Block (or a city with population of 1 lakh or above), will have to be as high as 20 per cent of the individuals. If the household is taken as the primary sampling unit, (which is statistically less efficient than individual), the sampling fraction may have to be larger, say 25 per cent. These fractions are, of course, in the nature of tentative guesses and need to be substantiated or revised by field studies on the ground.

Strategy of Sampling

16. What should be the broad strategy of sampling in the 1981 Census in the light of the above discussion ? On the assumption that sampling will be used at the stage of enumeration, two possible approaches can be suggested : One is to canvass the small list of census items in respect of 75-80 per cent of households and the sample list of items (along with census items) in the remaining fraction of households. This approach implies that the households covered in the sample will have to answer a larger number of questions than the rest. In the second approach, intended to make the burden of inquiry fairly even among all households, the sample list of items may be divided into convenient and self-contained modules of approximately equal size, and each module canvassed in a separate sample of households. For this purpose the total population of households may be equally divided into as many samples as there are individual modules. One convenient way would be to divide the sample list of items into four modules relating to fertility, migration, economic activity, and housing and canvass each of them in a 25 per cent sample of households.

17. Either approach mentioned above has its advantages and drawbacks. In the first case, there is non-uniform coverage of information from household

to household, the sample households being subjected to more extensive interviews. In the second case the work-load is fairly evenly spread among the households but one requires to tabulate the entire population of households for compiling the requisite tables, while in the former case the detailed tabulation could be confined only to the single sample of households. The formatting of the schedules and the procedures for canvassing require careful designing in both cases.

IS. The first approach is quite advantageous if it is possible to employ an agency with higher calibre for collecting information in the sample households. It will then be possible to collect a more detailed information and introduce a large degree of sophistication in the census questionnaires. If as a result of field investigations suggested above, the sampling fraction turns out to be rather small, the cost involved in engaging an agency of better calibre may not be beyond the means of the census organisation.

19. Whatever be the final approach chosen, it seems unnecessary to collect at the house-listing stage all the information covered in the 1971 Census on the basis of a complete count. It should be adequate to confine the house-list to a few identification particulars of the census house and households. All other details relating to the structure of the census house, housing conditions, and size of the household could as well be obtained on a sample basis at the stage of house-listing. The sample questionnaire can possibly include many other useful items of specific concern to the household as a whole.

Preliminary Studies and Investigations

20. In order to take correct decisions on the several issues discussed in the foregoing paragraphs and determine the exact role that sampling should play in the 1981 Population Census, it is necessary to launch immediately a number of studies with the help of the primary data already available from the 1971 Census as well as the results obtained through specially designed field surveys. The main points of study should be :

- (i) the determination, of the sample size (of individuals or households) required for compiling the census tables at the level of Community Development Block;
- (ii) the problems, if any, encountered by adoption of sampling at the time of enumeration; and

(iii) the choice between the two approaches suggested above in the event of sample enumeration.

21. So far as the studies of the past data are concerned the procedures followed by Lahiri and Ganguly and the indices suggested by them (vide Annex. II) will provide a useful starting point. Model sampling studies taking the individual and household as the primary units of sampling may be carried out with the help of 1971 Census data from selected areas. The study should be broad-based enough to take into account various categories of census tables and it should cover representative areas of different parts of the country, both rural and urban. The field surveys may be organised in such a manner that census schedules designed separately according to the two approaches are canvassed in selected C.D. Blocks and towns of 1 lakh and above by simulating, as far as possible, the conditions obtaining in the regular census. The operations should be directed and observed by a superior agency which can assess and bring to light the nature of problems encountered in the field and indicate the manner in which they can be surmounted.

22. To conclude, sampling in censuses can play a crucial role in providing quick and fairly comprehensive information at very modest cost. The Indian censuses in the past have, no doubt, employed this device to some extent but its benefits do not seem to have been fully exploited. The 1981 Census should take advantage of sampling techniques to the maximum extent possible and it is time that we begin serious preparations for this purpose.

Annex. I: Sampling in the 1971 Population Census

The 1971 Census collected the data on a universal basis in respect of all the items on housing and population, but most of the data was tabulated only on a sample basis. The tabulation on complete basis consists of the General Population Tables, Primary Census Abstract (for each village and urban block) and a table on population classified by age, sex and worker (industrial categories) and non-worker. Of the 16 main items (33 sub-items) in the individual slip, the complete tabulation covered 7 items (8 sub-items) viz., age, sex, religion, scheduled caste/scheduled tribe, literacy, mother tongue, main activity—broad category, nature of industry, trade, profession or service. Sample tabulation accounted for the remaining 9 items, covering a total of 25 sub-items.

A systematic sample of individual slips, 20% for urban areas and 10% for rural areas, is selected for the sample tabulations. Since the tables are presented separately for rural areas, non-city urban areas and cities (towns with population of 1 lakh and above) of each district, separate samples are selected for these areas. Further to this sub-division of district, in rural areas, tehsj / taluk/ police stations have been treated as operational units for sampling. However, unduly large units were further sub-divided into smaller units and where the units were too small, contiguous units were combined to form operational units of at least 20,000 population each. The systematic sample was selected from each operational unit independently. Individual slips of males/females of each operational unit were arranged according to the 11 industrial categories and / within each category by age group, before drawing the sample. The systematic sample was drawn starting with the first industrial category proceeding from one age group to the next and then to the next industrial category and so on.

For the housing tables, a 20% systematic sample of census houses was selected from the house-list and data relating to these sample houses and households living in them was tabulated.

The details of sample tabulations are as below :

(a) *Economic Tables*

- (i) Females by age, marital status and industrial category,
- (ii) Population by age, educational level and industrial category,
- (iii) Workers by industrial classification.
- (iv) Workers by industrial classification and employment status,
- (v) Workers in non-household industry by employment status and age.
- (vi) Workers by occupation.
- (vii) Workers by occupation and employment status,
- (viii) Workers by occupation and industry,
- (ix) Workers by occupation and age.
- (x) Workers by occupation and educational level,
- (xi) Population by secondary work and main activity,
- (xii) Non-workers by type of activity and age.
- (xiii) "Other" non-workers by age and educational level.

(b) *Social and Cultural Tables*

- (i) Composition of households by relationship to head and age.
- (ii) Households by relationship of members.
- (iii) Population by age and marital status.
- (iv) Population by age and educational level.
- (v) Population by single year age.
- (vi) Speakers of language by mother tongue and subsidiary language.

(c) *Migration Tables*

- (i)-Population by place of birth and place of enumeration.
- (ii) Migrants by place of last residence, place of enumeration and duration of present residence.
- (iii) Migrants to cities by age, educational level and workers by occupation.
- (iv) Population of cities by place of last residence and duration of present residence.
- (v) Population by industrial category and place of last residence.

(vi) Migrants by age, marital status, place of last residence and duration of present residence.

(d) Fertility Tables

Currently married women and births of last year by :

- (i) Religion, present age and age at marriage.
- (ii) Age, education and age at marriage.
- (iii) Religion, age at marriage and duration of marriage.
- (iv) Education, age at marriage and duration of marriage,
- (v) religion, present age, education and age at marriage.

(e) Housing Tables

- (i) Census houses by use.
- (ii) Census bouses by material of wall and material of roof,
- (iii) Households by number of members and number of rooms occupied.
- (iv) Households by size and tenure status.

Annex. II: _ Review of Past Studies on Sampling Application in Population Censuses

The use of sampling in population census has been increasingly recognised in India and other developed countries like U.S.A.; Canada and the U.K. Sampling techniques are now regarded as essential tools for reducing the work-load of enumeration, data processing and time-lag between the census count and the production of end results. Sampling also enables exercising better control over census operations (by virtue of reduced work-load) and thus ensures greater accuracy in the final results.

2. Several field experiments and desk studies were undertaken in the past both in India and abroad to determine the nature and extent of sampling in the population census. The following is a brief description of the studies/experience of sampling applications in India and abroad.

A. Indian Experience

3. The main object of the population census is to build up tables showing the distribution of the population classified according to various characteristics.

If sampling is to be adopted for obtaining information on, say, age-sex-marital status of individuals, the search for an ideal sampling strategy will have to be directed to pick up, among the various feasible and practicable strategies, the particular one which provides the distribution of the population according to the age-sex-marital status classification which does not differ significantly from the corresponding table based on cent per cent count. The major effort in this direction was made by Lahiri and Ganguly (1951) who suggested for this purpose an overall measure of the precision of the sample table since the entries in various cells of a sample table (which are estimated and computed from the same sample) are generally not mutually independent and as such the percentage standard errors of the separate estimates may not reflect the 'accuracy' of the sample table as a whole. Making use of the set of deviations of the estimated numbers in the various cells of the sample table, from the corresponding cell frequencies of the census table, Lahiri and Ganguly considered the following measures :

$$\delta_1 = \frac{1}{F} \sum |f' - f| \quad (i)$$

$$\delta_2 = \frac{1}{c} \sum \frac{|f' - f|}{f} \quad (ii)$$

$$\delta_3 = \sqrt{\left\{ \frac{1}{F} \sum \frac{(f' - f)^2}{f} \right\}} \quad (iii)$$

where f' and f are the corresponding cell frequencies of the sample and census tables, $F = S$ is the total population of the area concerned as per census count, c is the number of 'cells' in the census table and \sum denotes summation over these c cells.

4. To examine the relative merits of these measures and choose the one which is more preferable, the authors carried out a model sampling experiment using systematic sampling on the 1941 Census slips (one slip for each individual) of an area of about 75 square miles with a population of about 12,000 in the district of Hazaribagh in the State of Bihar. The slips, first arranged according to the serial numbers of the enumerator's blocks, circles, etc., were further arranged according to the census house number within each block and individual slips within households according to age of the individuals. Then systematic samples were extracted by choosing every 120th slip. The whole material was exhausted

by taking all the 120 samples. Age-sex-marital status tables were constructed for every one of these samples and the values of the three δ 's were worked out. Also, the values of the δ 's for the sample tables omitting unimportant cells (i.e., those having less than 1 percent of the total population) were computed. Then laying down the criterion that a good index of the level of disagreement should be such that an ordering of the set of sample tables on the basis of the index should not be very much altered if some minor alterations like ignoring the unimportant cells are made, the relative performance of the three δ 's was studied using the above all cell- δ and important cell- δ values. It was found that δ_1 is superior to δ_2 and δ_3 in that the proportion of cases in which the two series of rankings of the sample tables made on the basis of all cell- δ_i and important cell- δ_i agreed was the highest (about 91 percent) for δ_1 compared to 68 percent for δ_3 and 58 percent for δ_2 . In regard to the validity of the δ 's the authors assumed that every one of the δ 's is most probably a tolerably good measure of divergence and tested one against the others. It was found that δ_1 is more closely correlated with δ_2 and δ_3 , than either δ_2 with δ_3 and δ_1 , or δ_3 with δ_1 and δ_2 . Also, by using the concept of level of disagreement (or agreement) the validity of the all cell- δ_1 was tested against important cell- δ_2 and δ_3 . It was noted that the special aspects of divergence indicated by δ_2 and δ_3 have found fairly adequate reflection in δ_1 . Thus, it is deduced from the particular model sampling experiment that δ_1 is more reliable and valid than δ_2 and δ_3 .

5. The composite-error δ_1 is essentially positive and its behaviour is unbiased in the sense that the cell frequencies of the sample table are unbiased for the corresponding frequencies of the census table. So it tends to be unlike that commonly met with in the case of the error of an unbiased estimate of a single variable, for the reason that the δ_1 's need not concentrate near zero even though the errors in the individual cells are concentrated about zero. The authors had conducted model sampling experiments using systematic sampling to study the distribution of the composite error δ_1 . It was seen that the δ_1 's are concentrated about a point which is different from zero and as the size of the sample gradually increases the point of concentration keeps on approaching zero, till it is reached when the sample coincides with the population. Also, the standard deviation of the distribution decreases as the size of the sample is increased in such a manner that the coefficient of variation remains practically constant. Taking these results into account, the authors consider it necessary to use both mean (δ_1) and the S.D. of Σ_1 , i.e., $\sqrt{\{\text{mean}(\delta_1^2)\}}$, for deciding the choice between two alternative sampling schemes to be adopted.

6. As the coefficient of variation of δ_1 was almost constant for the age-sex-marital status table, the authors used mean (δ_1) for dealing with the question of relative efficiencies of the household and the individual as sampling units when the problem is to obtain the age-sex-marital status distribution of the population. A model sampling experiment was carried out using systematic sampling with individuals and households as sampling units. In the latter case, systematic sampling was applied to the totality of households in the given area without making any arrangement, and also after arranging them in the order of their size. Then as usual all possible samples were drawn, sample tables were prepared from each sample using unbiased method of estimation in the case of sampling individuals, and unbiased and ratio method of estimation in the case of household sampling, and mean (δ_1)'s were computed and compared. It was inferred by this study that under the system of census-taking obtaining in India, where every person has a separate census slip, the individual as a sampling unit is to be preferred. In case the age-sex-marital status enquiry is to be made concurrently, with just a complete count of the population then household sampling with the ratio method of estimation may be preferable. Further, it is argued that with the individual as a sampling unit the enumerator will possibly have a greater scope of substituting one individual for another since the order of enumeration within a household will depend, at least to a certain extent upon his choice, whereas he will really have no choice if the household is the sampling unit, since the house-listing is done prior to the census proper. Another, probably more important, reason is that the quality of the returns on age and marital status, particularly in a country like India, where people have very vague notions about their own ages, is likely to be much better if the household is taken as a sampling unit as it will provide the enumerators with some check on the reliability of the return on a person by comparison with those of the other inmates of the household.

7. Restricting the study to the case where the individual is the sampling unit, the authors have established an empirical relationship between mean (δ_1) and the sample size and drawn inference on the basis of this relationship. It has been noted that a two percent sample of the size of 22,000 (at district level) appears to be rather high even if the number of age groups is 10.

8. Some studies were also carried out by the Office of the Registrar General of India prior to the 1971 Census. These were dealt with by Majumdar (1973), who described the considerations in favour of deciding a systematic sample of

20 per cent in urban areas and 10 per cent in rural areas for the tabulation of the low priority items in the 1971 Census.

B. *Experience of Other Countries*

9. The U.S.A. introduced sampling procedures for the first time in the 1940 Census. Canada and Israel have followed the U.S. example in the recent 1970 series of censuses, while the U.K. carried out large scale sampling studies in 1972 as a prelude to the next census.

10. In the 1950 U.S. Census, a fairly large proportion of the information collected was obtained either from a 20 per cent sample of the population or a 20 per cent sample of dwelling units concurrently with the items obtained on a 100 per cent basis. Tabulations of sample data were generally made for areas consisting of 2,500 *population or more*, but cross tabulations only for much larger areas. The determination of whether an item was to be on the 20 per cent sample or the complete census depended upon the precision and the amount of detail with which the information was needed, in addition, account was taken of the costs involved and the expected accuracy of response for some of the questions. Population counts by age, sex, race and residence and some of the basic dwelling unit characteristics, for example, were regarded as fundamental and needed with high precision for all areas, small as well as large, and these were obtained on complete enumeration basis. Some questions were retained on a 100 per cent basis either because the data were comparatively inexpensive for a census (for example, citizenship) or because they were required for detailed tabulations and for comparatively *-small areas*. The items canvassed among the 20 per cent sample of persons included, among others, income, education and migration. The housing questions obtained on a sample basis covered such items as year built, heating equipment and so on.

11. The population and housing samples are both basically systematic samples, the sampling unit being a person and a dwelling unit respectively. Each population schedule consisted of 30 lines, one for each person enumerated. The sample information was obtained for the persons enumerated on **every** fifth line of the population schedule. The housing sample was set up differently. The procedure used was to group the sample questions into five sets and to get information on one set for each dwelling unit. Each schedule provided for 12 dwelling units. Studies made prior to the 1950 Census indicated that some problems could be expected of a systematic sample if variations were not introduced

in-the systematic pattern. This is primarily due to enumeration practices which result in cyclical patterns of listing the persons with heads of households enumerated on the first line of a schedule and so on. In order to reduce the effect of the systematic pattern, five different versions of the schedule were used. As a result, each line of the population schedule was part of the sample on one and only one of the five versions. Similarly each housing sample group was listed first on only one of the five printings. Within each census enumeration district, the schedules were divided approximately equally among the five versions.

12. The basic schedule design used for the 1950 Census had very important effects on the kind of data published. For example, for most characteristics of the population obtained on a sample basis, information was not available for the entire family. This meant that some cross classifications of sample data which might be desirable from an analytical point of view could not be made. In a similar manner, cross classifications for most sample housing items also could not be provided. The feasibility and efficiency of using the household instead of the individual as a sample unit was subsequently tested. The feeling was that the use of the household as a sample unit had advantages as well as disadvantages. One of the major problems was the difficulty of providing simple instructions to enumerators for dealing with institutions and other large quasi-households, which could-not be sampled on a household basis. However, a household sample permits the compilation of family statistics on a sample basis.

13. The 1960 U.S. Census of Population and Housing represented a major advance as compared with the 1950 Census. The research and evaluation work on the 1950 Census, the increased use of sampling and the application of electronic computer in large scale statistical operations were major factors influencing the new developments introduced in the 1960 Census. In this census, data on age, sex, colour, race and marital status and on some housing items were collected on the basis of complete enumeration, whereas the data on employment status, occupation, industry, class of worker, education, income and place of work and other population items and a substantial number of housing items were collected for a 25 per cent systematic sample households. Because a household sample was likely to give slightly less sampling efficiency for most items than a sample of individuals of the same size, the 1960 sampling rate was increased to 25 per cent as compared to the 20 per cent rate of sampling in the 1950 Census. The data collection was generally done in two stages. In the first stage, the enumerators collected the data on the complete enumeration items and they left sample questionnaires in every fourth household with the request to fill-in and return it

by mail. During the second stage, the enumerators contacted those sample households which either did not return the filled-in, or sent incorrectly filled-in, questionnaires. In areas, which were sparsely populated, the work on both the stages was, however, done in one operation.

14. In the 1970 U.S. Census, the samples were again single stage systematic sample of units selected at *one* of three rates, 20 per cent, 15 percent and 5 per cent. For persons living in housing units, the housing unit including all its occupants was the sampling unit; for persons in large group quarters, the sampling unit was the person. Actually the 20 per cent sample was further split into two panels of 15 per cent and 5 per cent each receiving a somewhat different form. This split panel approach was used to accommodate a longer list of questions without loading every sample respondent with the full set. Quality control procedures were used throughout the census process and where there was clear evidence that the sample procedures were not properly followed, enumerator's assignments were returned to the field for resampling.

15. In the 1972 Population Census, Israel followed a pattern very close to that of the U.S. In this census 100 per cent data were collected only in regard to a limited number of items and information on the rest of the items was obtained from a sample.

16. While in the 1961 Canadian Population Census, no sampling was used at the data collection stage, the evaluation programme of the 1961 Census indicated that the reliability of the census statistics based on a one-in-three sample would be comparable to that of a full count census. As a result most of the 1971 Census data was collected from a one-in-three sample.

17. Field tests were carried out in Britain in selected areas in 1972 to test the feasibility of a system of "inter-locking samples". Basic demographic questions were asked of all households; the remaining questions were distributed between 10 per cent samples, with some questions appearing in more than one of the samples. Thus there were ten different census forms each of about the same length and in one trial these were inter-leaved so that they could be handed out sequentially in the order one to ten. The basic aim of the test was to establish whether the questionnaire length could be half that of the 1971 Census without reducing the total number of topics covered, recognising, of course, that same questions would be asked of only a sample of households.

References

1. **Lahiri, D.** and Ganguly, A., 1951, An overall measure of precision of a sample table with applications in the study of relative efficiencies of different sampling units in population census, *Bulletin of International Statistical Institute*, 33(IV), 55-74.
2. **Majumdar, H.**, 1973, *Analytical Studies Relating to Sampling Procedure*, presented in the Second Data Base Seminar held in Poona.
3. **U.S. Bureau of Census**, 1956, *Sampling in the 1950 Census of Population and Housing*, Working Paper No. 4, Washington, D. C.
4. _____, 1960, *The Accuracy of Census Statistics with and without Sampling*, Technical Paper No. 2, Washington, D. C.
5. **Waksberg, J.** and **Hansen, R.**, 1965, *Sampling Applications in Census of Population and Housing*, Bureau of the Census Technical Paper No. 13, Washington, D. C.
6. _____, _____ and **Roupane, P.**, 1973, Estimation and presentation of sampling errors for sample data from the 1970 US Census, *Bulletin of International Statistical Institute*, 45(3), 66-32.
7. **Fellegi, T. P.** and **Sunter, A. B.**, 1973, Balance between different sources of survey errors, *Bulletin of International Statistical Institute*, 45(3), 334-354.
8. **Nathan, G.**, 1973, Utilisation of information on sampling and non-sampling errors for survey design—Experiences in Israel, *Bulletin of international Statistical Institute*, 45(3), 393-407.
9. **Redfern, P.**, 1973, The different roles of Population Censuses and Interview Surveys, particularly in the U.K. context, *Bulletin of International Statistical Institute*, 45(3), 191-210.

K. C. Seal

THE paper is very interesting and contains many useful suggestions. What is being advocated here is essentially a two part census procedure comprising the canvassing of small list of census items to about 75-85 per cent of households and detailed list, including the census items, to the remaining fraction of households to be selected through a simple method of sampling. While there is scope as well as justification for introducing sampling at the enumeration stage of the census, in order to make census more cost-effective, the critical question is how to operationalise it in the field.

2. Sampling at the tabulation stage, as was done in the 1971 Census, did not pose much problem as it was done by trained persons at the Census Headquarters by sampling systematically every 10th individual in the rural area and every 5th individual in the urban area. In replacing this by sampling at the enumeration stage, several problems are likely to crop up. Individual as sampling unit at enumeration stage, would involve, that first all the individuals are canvassed under the universal schedule and then a sample of them is taken up for detailed schedule. There does not appear to be any simple method by which a representative sample of individuals could be drawn in the field by persons not trained in sampling procedure.

3. In the case of treating the household as sampling unit at enumeration stage, we may not have an up-to-date list of households to start with but we would have a list of houses obtained from the house-listing operation which precedes a census by some 6-8 months. Some of the houses may, however, be establishments (non-household) and in preparing a frame of households at the time of census enumeration, the establishments would have to be weeded out. Rao's paper would appear to suggest (see para 9), that some sampling at house-listing stage is envisaged to get data on structure of houses, housing condition, etc. Does it mean that there would be two sampling operations, one at house-

listing stage based on list of houses and another at enumeration stage based on households ? Admittedly, the household sample drawn from the list of households prepared at the time of housing census conducted several months earlier would introduce some bias. On the other hand, if the sample of households is to be drawn at the time of the main census from the list of houses enumerated during the housing census, assistance of trained statistical investigator might be necessary,

4. A sample fraction of 20% if individuals or 25% if households, is suggested in the paper. If detailed characteristics are to be estimated at CD block level, this sample size would be huge on the population base of India. The results may not be commensurate with cost as the CD characteristics are likely to undergo fairly rapid change with time. The area for which detailed characteristics are to be tabulated can be a district, as in the past, while universal items alone need be tabulated at the CD block or lower levels. This might reduce the sample fraction.

5. In the 1971 Census tabulation by district, we have tables for (i) rural areas, (ii) cities with population one lakh and above, and (iii) non-city urban. If the sampling approach indicated in the paper is adopted we would get the data for CD block and cities with one lakh and above population only. Does it mean that non-city urban areas be left out in the process ?

6. In regard to the strategy of sampling, two methods have been indicated - one of which is the module method, in which a specific module with fertility or migration or economic activity or housing is canvassed in 25 per cent of households, while the second envisages that a single comprehensive schedule comprising all these items is canvassed to a 25 per cent sample of households. Apart from difficulties mentioned in the paper, obviously, the module method is not advantageous especially when co-relative studies (such as between migration and economic activity) are to be undertaken.

7. Along with the types of problems likely to be encountered in sampling as indicated above, the practical difficulties which the Registrar General has pointed out in adopting sampling at enumeration stage have also to be seriously considered. Any sampling scheme would remain a non-starter unless it is demonstrated to be operationally feasible keeping in view both technical and operational difficulties involved. If random samples are to be drawn *in the field*, trained

statistical investigators will have to be engaged in canvassing the census schedule. This does not seem possible at this stage unless the census operation becomes either a joint venture of the Office of the Registrar General and the Department of Statistics or the entire work of population census is entrusted to the Department of Statistics. Such a change in structure and organisation in the near future would not be feasible. Canvassing of a universal census slip to all the persons and a more detailed slip along with the universal slip to a random sample of the population may, however, be possible if certain census blocks were selected through area sampling with appropriate sample design. Both the universal and the detailed slip need to be canvassed in the selected census blocks only. For this purpose, the census enumerators of higher calibre may be specially chosen and given more intensive training for canvassing the universal and detailed census slips. This should ensure collection of quality data in respect of sampled census blocks. This is also important in view of the fact that doubts have been expressed in the past about the quality of data on certain socio-economic characteristics collected in the past censuses through enumerators of limited calibre. It might be worthwhile to pay a higher honorarium to these trained census investigators used for canvassing the universal and the detailed slips in the sampled census blocks than what is allowed to ordinary census enumerators. Such a procedure should overcome the genuine difficulties pointed out by the paper. The large number of trained investigators required for canvassing the sampled census blocks could alternatively be taken on deputation from the Bureaus of Economics and Statistics of various State Governments. The cluster sampling of census blocks will no doubt provide less precise estimates than other sampling methods using the household or the individual as the sampling unit. However, the loss in precision may not be of great concern since the number of census blocks to be sampled by the aforesaid area sampling method will be very large; as many as 0.2 million, in case 20% of census blocks are sampled.

8. Regarding the coverage of items for universal canvassing, it is felt that it should include marital status along with age, sex, religion, scheduled caste/scheduled tribe and economic activity. This suggestion is made because marital status is now an important policy variable and schemes pertaining to population control and family welfare would require this information for the smallest area. The detailed information to be collected under the sample may cover mother-tongue, migration, detailed socio-economic characteristics and, in respect of currently married females, also age at marriage, number of children born and number of surviving children.

9. The paper mentions that 80-90% of the information collected in the past censuses was practically left untabulated. If this is true, it is a matter of serious concern. It is essential that some sampling procedure is introduced at the enumeration stage to avoid such wastage of resources in collection of unutilized data in future. If sampling of households is not found feasible, at least the alternative sampling method advocated above in para 7 should be tried out in the 1981 Population Census.

Session III: Problems Relating to Computer Hardware, New Computer Needs and Problems of Software

PAPERS

N. Seshagiri

THE National Informatics Centre can provide processing facilities on its large computer proposed to be installed in the second half of the financial year 1978-79.

2. To the extent required, computer programmes for the 1981 Census data analysis, can be provided by the National Informatics Centre. In case it is decided that NIC would be assisting the census computerisation, NIC should be involved in the design of formats and data structures for the 1981 Census data preparation.

3. Multiple key-to-tape units which are proposed to be installed at the NIC during the first half of 1978, can be demonstrated to those responsible for the computerisation of the census data, so that they can take a decision by the middle of 1978 whether such equipment would be useful for them to purchase for their operations.

4. Capture and computerisation of census data is in the nature of a "peak-load planning" from the larger national perspective. The load would be mainly spread around 1981 and 1982 as far as the data preparation and processing are concerned. One of the peak-load planning procedures, ideally suited for the present case, is the following. Certain growing organizations like NIC are identified who have an approved growth profile during the years 1983 and 1984 requiring the purchase of a number of multiple key-to-tape units and employment of data preparation and data processing personnel. By a prior approval of the Planning Commission and the Ministry of Finance, it would be possible to pre-

pone the purchase of the equipment and recruitment of personnel by about two years and take up the census data preparation and processing during 1981 and 1982. NIC is one such organization. A few others of this kind may also be identified. During 1981 and 1982, the equipment purchased and personnel recruited on this preponed basis, can be engaged fully in the preparation and processing of census data. By 1983, these equipment and personnel would be released just at the time the inhouse requirement of these organizations reach up to the level adequate to utilise them.

5. It is recommended that as far as possible, preparation of cards could be avoided and direct entry devices used in large numbers. The use of optical mark page readers can be thought of as trained and educated persons may be required to collect the data on optical mark pages. One may consider employing this method in the cities. In any operation of this colossal magnitude, no single approach would be optimal in all places and circumstances. Therefore, there is a need to work out the optimal mix of various types of data preparation methods, like optical mark page method, optical mark card method, direct entry method, card punching method and manual operation.

A, Balasubramanian

1. Electronic Data Processing, (EDP) has found increasing roles in supporting census operations in India since the 1971 Census. Whilst data was computerised through the media of punched cards (largely relying on 'hand-punches') and then being transferred on to tapes, the processing of the data was largely on a 16K Character Memory Machine with sequential tape file handling capability. Processing of the 1971 Census data was partly carried out manually and partly on the computers. The main Census Schedules (100%) and the rural samples (100%), were carried out manually whilst the other processing of urban samples (20%), house-list data (20%), advance tabulations (1%) etc. were carried out on the computers.

2. In the context of the EDP equipment that is in use in the country, the 1981 Census could be planned on the basis of the assumption that much of the data being collected is duly formatted in machine transformable form (80 column images) and in some cases calling for manual scrutiny, in a machine readable form using mark-sensing techniques. Once the information is collected, the digesting of the information and statistical summarization would enable infor-

mation being made available in a variety of different forms. Select combinations of data are normally tabulated in a form that somebody conceives as being useful. But it is quite obvious that many questions that arise for planners and analysts on census data are not answered directly by the census data and one may have to process the original raw data to answer such questions. It should also be possible to provide representative sample of census information, after removing identification information, to enable users to study and tabulate correlations at will.

3. It will also be desirable to hold data in random access files containing multiple indices tabulations and summaries for answering many questions being asked.

4. Current Plans for Computers

The major computer facilities that can be expected to be installed in the Government sector and which facilities could be made use of in compilation of data or for processing are listed in Appendix 'A'. Besides these systems a number of computers and data entry facilities will be installed in industrial and commercial organisations all over the country which can be utilised on a shared basis in a planned manner. Such organisations include the Reserve Bank of India and the KIC at Bombay, the Bharat Heavy Electricals at Tiruchirapalli (Tamilnadu), Hardwar and Delhi, Bharat Electronics at Bangalore, the Electronics Corporation of India at Hyderabad, Steel Plants in Bihar and the Regional and the State Government Computer Centres which are also being continually upgraded. Facilities at state capitals such as Bhopal, Chandigarh, Ahmedabad, Hyderabad, Bangalore, Bhubaneshwar, Shillong etc. could be expected to be upgraded and provided more elaborated data conversion facilities.

5. Data Entry

In census operations data coding, checking and entry constitute a very major effort. A typical estimate is 70,000 man weeks on a single shift basis for coding and checking and about 40,000 man weeks for key-boarding. On a three shift, six month time period basis, the number of key-to-tape key-boards needed would be of the order of 2500. A fair degree of data validation would be done at the stage of manual coding and again on the computers. The use of key-to-tape data entry devices should be largely made to avoid wasteful expenditure on cards. A multiple key-board direct data entry system would provide for veri-

fication, validation and sorting without having to use the main computer for such jobs. A sophisticated direct data entry system with 16 terminals would cost around Rs. 12 lakhs at current international prices. Indigenous key-to-tape and key-to-cassette, key-to-diskette systems would also be available in the market. In view of the need to make large numbers of such key-boards available for the census operations and *also to* ensure continued utilisation of such data entry devices, we should plan for the building up of about 2500 key-board stations, distributed amongst the state data processing centres, NIC cells and NSS units of 8/16/32 key-boards. An investment of the order of Rs. 12.5 crores could be envisaged for this requirement.

6. Use of Mark-Sensing

Mark sensing as an easy means of transforming data into machine readable form could be used to good purpose. In activities requiring checking, quite often there is a need for group totals to be recorded on sets and groups of documents which can be subsequently summarised on Taluk/District/State basis. Mark-sense cards can be used for such purposes. The National Centre for Software Development and Computing Techniques (NCSDDCT) Bombay has established indigenous sources for mark-sense cards which cost about the same as the normal punched cards. However, there will be a need to identify areas of application, where mark-sensing could be used with economy and consistently with the requirement of accuracy in transfer of data to the computer.

7. Storage of Data

Whilst all data will be stored on tapes, we should standardise on 1600/3200 BPI, 9 track tapes for minimising storage volumes. Programs and procedures for analysis should be based on direct access, disk media. It is to be noted that facilities would be available in all major cities and almost all state capitals. At Delhi, the use of major facilities at NIC, Planning Commission, Department of Statistics could be made. The IBM-1401 system currently with the Census could be replaced by an intelligent terminal linked up to one of the larger centres. The terminal could have provision for direct data entry, mark-sense card reading, optical recognition facilities, tape peripherals and printers, for enabling local data conversion, file development on major system, handling of enquiries, and also assembling data as may be identified by various users.

8. Views of Public Accounts Committee Related to Use of Computers for Census Operations

The PAC has advised in the 221 st report to the Fifth Lok Sabha that for one time applications (such as collection of census data) work should largely get executed through centralised agencies as the Government Computer Centre and the design of the system and procedures should be based on an analysis and evaluation related to the actual needs,*

9. Conclusion

The earlier experience in the use of computers to support census operations has confirmed the need for planning the 1981 operations to be more integrally supported by computer based operations. Computerisation of census data would support any enlargement of the scope of the volume of data to be collected to meet any specific objectives or directions for the census. The volume of operations and the size of effort to be mounted in relation to pragmatic time scales for compiling data would call for decentralisation of the coding, checking, and data entry operations on an all India basis.

Such decentralisation would require coding and systems procedures and norms to be evolved centrally and training and operations implemented in a coordinated basis. Increasing use of direct data entry devices, minimising expenditure on punched cards as a media, should be resorted to. High density (1600/3200 BPI), 9 track, tapes should be adopted as a standard. Data analysis should be based on direct access media. Investment would be required for data entry devices and media (tape, cards, disks). To ensure continued utilisation, such investments should be coordinated with data collection organisations such as NIC, NSS, Regional Computer Centres, State Data Centre. Use should be made of major computer facilities for the analysis and dissemination of data.

7. The computer *facilities* at the Registrar General (Census) should be made more adaptive to enable coordination of the overall programme.

*221st Report of the Public Accounts Committee of Lok Sabha, para 2.46 and 2.47.

Appendix—A

<i>State</i>	<i>Institution</i>	<i>System</i>	<i>Configuration</i>
(0)	(2)	(3)	(4)
I. Delhi			
	National Informatics Centre	Large System (\$ 3.5M) with cells at Bombay (TDC-316), Calcutta and Madras	
	Planning Commission	INTERDATA 8/32	Memory 256 KB , Disk 300 MB 2 CR, 2 LP etc.
	Department of Statistics (proposed)	—	Memory 256 KB , Disk 300 MB, 6 Tape , 1 CR, 1 LP etc.
	Institute for Agriculture Research and Statistics	Burroughs 4771	Memory 150 KB , Disk 200 MB , 4 Tapes, 2 CR, 1 LP etc.
	Indian Institute of Technology (proposed)	ICL 2960	Memory 1.5 MB , Disk 1200 MB , 2 Tape, 1 CR , 1 LP etc.
II. Maharashtra			
Bombay	National Centre for Software Development Computing Techniques	DEC 1077	Memory 256 KW (36 bits) , Disk 400 MB, 2 CR , 2 LP, etc.
Bombay	Regional Centre (one planned by CMC)	-	Memory 512KB, Disk 400 MB, 4 Tape, 1 CR, 1 LP, Large Data Entry facilities
Poona	Regional Computer Centre	ICL 1904	Memory 128 KW (24 bits). Disk 246 M Character, 4 Tape, 2CR, 2LP, etc.
III. West Bengal			
Calcutta	Regional Computer Centre	B 6748	Memory 393 KB (to be expanded) . Disk 400 MB, 4 Tape, 2 CR, 2 LP, etc.
Calcutta	Indian Statistical Institute	R 1033	Memory 256 KB, Disk 120 MB, 4 Tape , 2 CR, 2 LP, etc.

Appendix A (contd. on page 79)

(1)	(2)	(3)	(4)
IV. Calcutta	VEC Project	IRIS-80	
Tamilnadu			
Madras	Indian Institute of Technology	IBM 370/1 55	Memory 512Kbyte, Disk 400 MB, 3 Tapes, 2 CR, 2 LP
	State Govt. Centre (to be replaced)	IBM 1401 TDC-312	
V. Uttar Pradesh			
Kanpur	Indian Institute of Technology (proposed)	--	Memory 512 KB/784 KB, Disk 400 MB, 3 Tapes, 2 CR, 2 LP etc.
Lucknow	Govt. Computer Centre (proposed)	---	Memory 256 KB, Disk 280 MB, 2 CR, 2 LP
Varanasi	Banaras Hindu Vishvavidyalaya	TCL-1904S	Memory 128KW (24 bit), Disk 248 Men., 24 Tape, 2 CR, 2 LP
Roorkee	Roorkee University	DEC-2050	Memory 256 KW (36 bits), Disk 400 M byte, 4 Tape, 1 CR, 2 LP
VI. Punjab-Haryana			
Chandigarh	Regional Computer Centre	DEC-2050	Memory 256-KW (36 bite), Disk 400 M byte, 4 Tape, 1 CR, 2 LP
VII. Karnataka			
Bangalore	Indian Institute of Science (proposed)	—	Memory 512KB, Disk 200MB, 3 Tapes, 2 CR, 2 LP, etc.
Bangalore	State Govt. Centre (proposed)	—	
VIII. Gujarat			
Ahmedabad	Physical Research Laboratory	IBM 360,44	Memory 256 KB, Disk 15 MB, 2 Tape, 1 CR, 1 LP
Ahmedabad	State Govt. Centre	ICL-I901A	
Anand	National Dairy Development Board	ICL-2950	Memory 1 MB, Disk 400 MB, 4 Tapes, 2 CR, 2 LP
IX. Kerala			
Trivandrum	Vikram Sarabhai Space Centre	IBM-360/44	Memory 256 KB, Disk 15 MB etc.

Abbreviations. LP—Line Printer, CR—Card Reader, and Tape—Magnetic Taps.

J. Roy

LACK of adequate data preparation equipment will be a major constraint on computerisation. It will, therefore, be prudent to restrict the use of computer only to complex tabulations on a sampling basis. The responsibility for formulating an efficient sampling design for this purpose may be entrusted to an expert group.

Even the limited amount of computerisation that is envisaged should be done on a decentralised basis at regional centres, where data should be transcribed on to magnetic tapes and 'cleaned' through a process of scrutiny and editing. Preliminary tabulation of these regional data should also be done at regional centres to provide control figures. These magnetic tapes from different regional centres should be collated at a central data processing centre, located at Delhi, where final all-India tables should be prepared. For this purpose, it is necessary to identify as early as possible the regional data processing centres and examine the compatibility of hardware and software available at these centres.

A major problem will be standardisation of the scrutiny and data-editing procedures at different centres. Specifications for scrutiny and editing will have to be drawn up by an expert group which will also examine whether available hardware and software are powerful enough to meet these specifications. If they are not, modification of the existing software, or even preparation of new software, will have to be undertaken. For this purpose, it may be desirable to harness the expertise available all over the country and seek cooperation of different research institutions and other agencies.

The above recommendations are for the 1981 Census only, but preparations must start immediately for *full* scale computerisation of future censuses. For this purpose, it will be essential for the Registrar General to have a technical group organised on a permanent basis. It may be necessary to acquire as soon as possible key-to-tape data recording machines, which can be tried on an experimental basis during the 1981 Census.

A. Ghosh

IN computerising the census data, the main problem is to put the scrutinised data on tapes or discs. In comparison, the problem of tabulation or cross-tabulation on computer is much simpler and can be done on any medium-sized computer available in the country.

The use of key-to-tape devices will no doubt greatly facilitate the task of putting the data on tape—thus eliminating not only the use of cards but separate data validation runs on the main computer (as data keyed-in are edited by multi-terminal data entry systems). But acquiring a large number of these data entry systems, assuring their proper maintenance, training the personnel to use them, will, in my opinion, pose a gigantic problem, keeping in view the time available before the 1981 Census.

So it appears that for speedy presentation of a set of simple tables the old fashioned, but quite effective, devices should be maintained. At the same time a sincere attempt should definitely be made to fully computerise the census data by acquiring sophisticated key-to-tape devices, so that even if the computerisation of the whole census data is not feasible in the time available, the experience gained will be of immense value. After all, large scale computerisation is inevitable in the long run.

Symposium Participants

1. Brig. A. Balasubramanian,
(Author of Paper for Session III),
Department of Electronics,
New Delhi.
2. Dr. Ashish Bose,
(Discussant Session I),
Head, Demographic Research
Centre,
Institute of Economic Growth,
University Enclave,
Delhi-110007.
3. Shri K.K. Chakravorty,
Assistant Registrar General,
Office of the Registrar General,
India,
2-A, Mansingh Road,
New Delhi-110011.
4. Shri R.C. Chanda,
Director, Computer Centre,
Planning Commission,
Yojana Bhavan,
New Delhi-110001.
5. Shri P. Chandrasekhar,
Senior Systems Scientist, I.P.A.G.,
C-4/41, Safdarjang Development
Area,
New Delhi-110016.
6. Shri S.N. Chaturvedi,
Deputy Director, Data
Processing,
Office of the Registrar General,
India, West Block I, Wing-I,
R.K. Puram,
New Delhi-110022.
7. Prof. P.B. Desai,
President,
Indian Association for the Study
of Population,
Institute of Economic Growth,
University Enclave,
Delhi-110007.
8. Dr. Stan D'Souza,
(Author of Paper for Session I),
Director,
Indian Social Institute,
Lodhi Estate,
New Delhi-110003.
9. Shri V. Elangovan,
Computer Group,
Electronics Corporation of India
Ltd.
Hyderabad.

10. Dr. Ambrish Ghosh,
(Discussant Session III),
Professor of Mathematics,
Indian Statistical Institute,
203, B.T. Road,
Calcutta-700035.
11. Shri M.S. Gokhale,
Director, Computer Centre,
Department of Statistics,
East Block, R.K. Puram,
New Delhi-110022.
12. Shri R.P. Goyal
Treasurer,
Indian Association for the Study
of Population,
Institute of Economic Growth,
University Enclave,
Delhi-110007.
13. Shri S.K. Gupta,
Director, Data Processing Division,
National Sample Survey
Organisation,
25-A, Shakespeare Sarani,
Calcutta-700017.
14. Dr. M. Holla, (Rapporteur)
Joint Director,
Central Statistical Organisation,
Sardar Patel Bhavan,
Parliament Street,
New Delhi-110001.
15. Prof. Asok Mitra,
(Author of Paper for Session I
and Chairman Session IV),
Centre for the Study of Regional
Development,
Jawaharlal Nehru University,
New Delhi-110057.
16. Prof. A.L. Nagar,
(Chairman Session II),
Delhi School of Economics,
University of Delhi,
Delhi-110007.
17. Dr. D.D. Narula,
Director,
Indian Council of Social Science
Research,
Indraprastha Estate,
Ring Road,
New Delhi-110002.
18. Shri K.S. Natarajan,
Senior Research Officer,
Office of the Registrar General,
India,
4/19, Asaf Ali Road,
New Delhi-110002.
19. Shri P. Padmanabha,
(Author of Paper for Session I),
Registrar General, India,
Office of the Registrar
General, India,
2-A, Mansingh Road,
New Delhi-110011.
20. Dr. M.K. Premi,
(Discussant Session I),
General Secretary,
Indian Association for the Study
of Population,
Centre for the Study of Regional
Development,
Jawaharlal Nehru University,
New Delhi-110057.

21. Shri M.P.Rao,
Assistant Director (DP),
Office of the Registrar General,
India,
West Block-I, R.K. Puram,
New Delhi-110022.
22. Shri V.R.Rao,
(Author of Paper for Session II
and Chairman Session I),
Director,
Central Statistical Organisation,
Sardar Patel Bhavan,
Parliament Street,
New Delhi-110001.
23. Shri Ravinder Kumar,
Electronics Commission
(I.P.A.G. Cell),
C-4/41, Safdarjang Development
Area,
New Delhi-110016.
24. Shri P.M. Reddy,
Statistician,
Operations Research Group,
Race Course Road,
Baroda-390007.
25. Prof. J.R. Rele,
(Discussant Session I),
Director,
International Institute for
Population Studies,
Govandi Station Road, Deonar,
Bombay-400088.
26. Prof. J. Roy,
(Discussant Session HI),
Computer Division,
Indian Statistical Institute,
203, B.T. Road,
Calcutta-700035.
27. Shri J.S. Sarma,
Additional Secretary,
Department of Statistics,
Sardar Patel Bhavan,
Parliament Street,
New Delhi-110001.
28. Dr. R.K.S. Saxena,
Indian Meteorological
Department,
New Delhi.
29. Dr. K.C. Seal,
(Discussant Session II),
Advisor, Labour & Employment,
Planning Commission,
Yojana Bhavan,
Parliament Street,
New Delhi-110001.
30. Dr. S. Selvaratnam,
Economic and Social Commission
for Asia and the Pacific,
Bangkok.
31. Dr. N. Seshagiri,
(Author of Paper for Session III),
Director, (I.P.A.G.),
Electronics Commission,
C-4/41, Safdarjang Development
Area,
New Delhi-110016.
32. Shri R.N. Sharma,
Assistant Director,
Central Statistical Organisation,
Sardar Patel Bhavan,
Parliament Street,
New Delhi-110001.

33. Shri U.P. Sinha, Reader,
International Institute for
Population Studies,
Govandi Station Road, Deonar,
Bombay-400088.
34. Smt. G. Suguna Kumari,
Deputy Director,
Central Statistical Organisation,
Sardar Patel Bhavan,
Parliament Street,
New Delhi-110001.
35. Shri K. Sundaram,
Computer Group,
Electronics Commission of
India Ltd.,
Hyderabad.
36. Brig. V.M. Sundaram,
(Chairman Session III),
Army H.Q., Electronics Data
Processing Centre,
- Signals Enclave,
New Delhi-110010.
37. Shri R.P. Tyagi,
Institute of Economic Growth,
University Enclave,
Delhi-110007.
38. Shri K.R. Unni,
Assistant Director (D.P.),
Office of the Registrar General,
India,
West Block I, R.K. Puram,
New Delhi-110022.
39. Dr. (Mrs.) Vatsala Narain,
Secretary,
Indian Association for the Study
of Population,
International Institute for
Population Studies,
Govandi Station Road, Deonar,
Bombay-400088.