Reprinted from Sankhyā : The Indian Journal of Statistics, Series A, Vol. 23, Part 4, 1961.

ANALYSIS OF ERRORS IN CENSUSES AND SURVEYS WITH SPECIAL REFERENCE TO EXPERIENCE IN INDIA*

By P. C. MAHALANOBIS and D. B. LAHIRI

Indian Statistical Institute

I. Introduction

1. The demand for statistical information is growing increasingly and rapidly, and survey organisations are hard pressed to supply results with the required speed. They must assess the quality of the data and caution the users that the results are not perfect, but they must avoid creating an undue impression that the data are so unreliable as to be of no use. The manner in which the existence of errors can be presented would depend on the level of statistical maturity of the country, not only of a select group of survey experts but also of the direct users, the policy makers, and to some extent the general public who may be affected by any policy decision based (at least partially) on the statistical information. There, however, is general agreement among survey experts that survey reports should supply some idea of the reliability of the results.

2. The study of errors in surveys has two purposes. First, to guide the user in interpreting the results. Secondly, in improving the quality of future surveys. Even when great care has been taken to set up important controls it is still necessary to get an assurance that the controls were effective and results with desired accuracy[1] have been obtained.

3. In India a number of techniques are being used for evaluation of survey results. All depend on comparisons among alternative (independent) estimates. Some of these may on *a priori* grounds be assumed to be more reliable than the others.

4. An illustration of this is provided in the evaluation of the results supplied by (a) (so-called) complete enumeration and (b) by sample survey of the outturn of jute in Bengal against (c) statistics of jute trade obtained subsequently which are known to be very reliable. This is a rather rare example where opportunity had occurred of comparing three sets of estimates for two consecutive years.

5. In some situations there are no *a priori* grounds for preferring one estimate to another. In this case the divergence between two or more independent estimates may provide a basis for a mental appraisal of the margin of uncertainty. This is

* This paper originally presented at the 32nd Session of the International Statistical Institute is being reprinted here with the kind permission of the Editor, *Bulletin of the International Statistical Institute.*

1 For practical purposes the desired accuracy can be often thought of or specified as the margin of 'permissible error' in the sense that any decision to be taken on the basis of sample estimates would remain the same within the limits of the permissible error.

made possible if the results are based on interpenetrating net-work of (sub) samples (IPNS) each of which is surveyed and/or processed by different but comparable operational units. When two (or more) samples are drawn from the same population and covered according to the same survey design, the results based on the different samples are equally valid, even though they are derived by different operational units; and divergences between the different sets of estimates supply directly some idea of the margin of uncertainty. It will be noticed that the operational differentials find a reflection in this manner of assessing the margin of uncertainty.

6. The interpenetration at the operational level may be of various types. For example, the field work is conducted by the same agency but the processing is done by different agencies.[2] Or, the field work is by different agencies but processing by the same agency; for example, in the survey of landholdings in India (reported in this paper) the entire processing work was done in the Indian Statistical Institute, even though the field work was conducted by both the Central and State agencies. Again, there may be complete interpenetration where both the field and processing work are done by completely different agencies; this is the form which the Indian National Sample Survey (NSS) is gradually assuming with the participation of the different States of India. Even for the same agency our normal practice is to arrange field work with interpenetration in respect of parties of investigators. Study of party and agency differences finds a place in our section on landholdings. There are also some marginal cases where even for the same agency the processing is done at different centres, or by different tabulating units, or at different points of time.

7. It should, however, be pointed out that, strictly speaking, a complete absence of *a priori* preferences is not always a reality. For example, if it is known that a particular agency has had a long experience in a particular field then the results thrown up by it may be accepted to have higher validity. Or, when one survey is carried out by temporary *ad hoc* staff and another survey by a whole-time permanent statistical staff then some may be inclined to accept the results thrown up by the latter to have higher validity. For the same reason one may accept an intensively supervised well conducted sample survey by qualified, experienced and well-trained investigators to have higher validity compared to a complete enumeration conducted under usual census conditions. Examples of this situation will be found in our sections on spot-check of crop census records and sample verification of livestock census.

8. One difficulty of evaluation is that the assessor may not be in possession of full background information about the agency, or about the conditions under which the census data, for example, may have been collected or processed. Sometimes comparisons between census and sample check or those between interpenetrating

---

[2] This is the practice for the pre-harvest crop-acreage survey where the field work is conducted by *patwaris* under the State authorities. The processing is done by three agencies each covering two sub-samples, the agencies being (a) the State authorities, (b) the Directorate of National Sample Survey, Government of India, and (c) the Indian Statistical Institute.

samples covered by different agencies may provide corroborative evidence in support of certain 'feelings' based on previously available background information which may even be of a vague and inadequate nature.

9. It may be worth-while stressing an obvious point which, however, is sometimes not kept in mind. For proper evaluation of results it is essential to take into consideration all available information, not merely the internal evidences supplied by the survey alone; not merely the quantitative external evidences but also even vague non-quantitative evidences. There is usually a residual element of subjective appreciation even when decisions are sought to be made in an objective way. The aim must be to utilise the whole of the available evidence in such a way that a maximum amount of agreement can be reached among competent assessors. This is the basic approach of science.

10. In some methods of evaluation the emphasis is on unitary check where the purpose is to evaluate the quality of data collected at the primary unit of enquiry. Evaluation of results for small administrative units is also possible. These are illustrated by the spot-check of crop census records and sample verification of the livestock census. Such checks can supply not only material for the study of errors of ascertainment, unit by unit, but if performed on a random basis, may also provide a means for separation of total *net* error of an aggregate estimate into ascertainment error and errors of coverage and compilation. (See our section on livestock census).

11. Another useful technique is to break up the survey period, for example, one round[3] of the National Sample Survey (NSS) into a number of sub-rounds, and to compare the estimates for each sub-round. Differences in such sub-round estimates, if any, would give valuable information for proper interpretation of the data. This is an example of a survey plan where the work of the same agency can be evaluated against its own work conducted under somewhat different conditions. (See our section on population growth).

12. The above idea of self-evaluation has been used in other forms. For example, in crop cutting experiments, to estimate the yield of crops per hectare, the crop is harvested separately, at each sample point, in the form of two or three concentric sample cuts. To what extent the work has been done under control can be then studied from the magnitude of the divergence between different estimates, each based on a different size of cut.

13. A similar device which has been found to be of value, is using more than one reference period of time for the collection of data by the interview method. In some designs the data are collected for a long period (for example, one year) in such a way that tabulations by shorter reference periods (e.g. one month) are possible. Comparisons of results based on different periods of reference may reveal factors

[3] NSS is a multi-purpose survey, a specified group of subjects is covered in an integrated manner in a single "round" to be completed in a specified survey-period; different rounds have usually a different survey plan depending upon the group of subjects chosen and upon the relative emphasis placed on different aspects of the survey.

like 'recall lapse.' An illustration of this is provided in the survey design for the estimation of birth, death and growth rates.

14. Another technique is to compare estimates obtained from one method of collecting the data against another more reliable method for estimating not the same character but a related one which sets a lower (or upper) limit to the former character. An illustration is provided in the comparison of death rates obtained by the interview method against a method of keeping a watch (by the method of re-enumeration) on a sample of individuals. Comparisons may also be made of results obtained by a less intensive enquiry against those based on a more intensive one. (See section on population growth).

15. As the Indian National Sample Survey (NSS) is carrying out surveys round after round, it is possible to compare the results based on two or more rounds.[4] 'Consistency' over rounds adds to the confidence with which the results may be accepted.

16. Another technique which appears to have good potentialities is comparison over 'space' as distinguished from 'time' (or rounds) which we have just described. Illustrations of similarity of the nature of divergence from one area to another are to be found in our sections on landholdings and sample verification of livestock census.

II. JUTE PRODUCTION IN BENGAL 1944-45 AND 1945-46

17. It is not always that one gets an opportunity of evaluating the results obtained on the basis of complete enumeration and sample survey against a third, but extremely reliable, figure. Such an opportunity was availed of in two consecutive years in regard to the 1944-45 and 1945-46 Jute Crop of Bengal.

18. Jute being a cash crop of international importance, accurate export trade figures are maintained and become available about 15 months after the harvesting season. Being a crop of such importance there is naturally a great demand for accurate statistics as early as possible. The official forecast based upon plot-to-plot enumeration attempted to meet that demand. Sample surveys were also conducted by the Indian Statistical Institute. These were objective methods of enquiry where

---

4 For example, an assessment is sometimes made in regard to listing (of households for sample selection) by repeating the same sample areas (villages or urban blocks). A partial analysis of such data (NSS 13th round) shows that in two States (rural areas only) there has been under-listing of households by 0.7% and 1.0% respectively. Another example is provided by our sample survey of manufacturing industries where data on capital, labour, input, output and other related information are annually collected. In the current round data are being collected separately for the last two years. The data for the year before last were also collected in the previous round. As the larger establishments are being covered in both the rounds we have here an opportunity of examining the consistency of the information obtained in the two rounds for these common (about a thousand) establishments. A preliminary examination shows that this technique is likely to prove useful. Taxes of the order of 12.5 million rupees paid on account of excise duties have not been taken into account in the previous round. In another instance an amount of 5.6 million rupees was left out from the input entries in the earlier round.

a random sample of fields was taken up for actual physical examination for acreage estimation; and harvesting of a random sample of crop cuts followed by necessary weighments provided the yield rates. There were two interpenetrating sub-samples (IPNS) which were covered by different parties of investigators.

19. Table 1 gives the relevant information. It will be noticed that in both the years the official forecasts based on complete count were both very much out whereas the sample survey estimates were quite close to the trade figures. The two sub-sample (IPNS) estimates in both years agreed with the trade estimates within roughly 3 per cent while the estimates based on the so-called complete count differed from the trade figures by 27.2 per cent in 1944-45 and 16.6 per cent in 1945-46.

TABLE 1. COMPARISON OF OFFICIAL (COMPLETE COUNT) AND SAMPLE SURVEY ESTIMATES OF JUTE PRODUCTION WITH TRADE FIGURES, BENGAL, 1944-45 AND 1945-46

particulars about jute crop	quantity (thousand bales)[a]	
	1944-45	1945-56
(1)	(2)	(3)
1. consumption during the season		
1.1 in jute mills (actual)	6000	6308
1.2 exports (actual)	1050	2213
1.3 in villages (estimate)	600	600
2. total	7650	9121
3. consumed from previous year's stock	324	697
4. jute crop in other provinces	598	862
5. balance : Bengal crop, trade figures	6728	7562
6. complete count : Bengal crop, official forecast	4895	6304
7. sample survey : Bengal crop, Indian Statistical Institute		
7.1 sub-sample 1[b]	6836	7734
7.2 sub-sample 2[b]	6518	7773
7.3 full sample	6686	7755
8. discrepancy of (6) on (5)	−27.2%	−16.6%
9. discrepancy of		
9.1 (7.1) on (5)	+ 1.6%	+ 2.3%
9.2 (7.2) on (5)	− 3.1%	+ 2.8%
9.3 (7.3) on (5)	− 0.6%	+ 2.6%

[a] 1 bale = 400 pounds, 1 *maund* = 82.2857 pounds;

Trade figures are reported in bales but sample survey carried out on the basis of *maunds* which have been converted into pounds and then to bales, (the approximation 1 bale = 5 maunds sometimes used would give somewhat lower figures).

[b] Sub-sample estimates of production have been obtained by multiplying sub-sample estimates of crop acreage by the full sample yield rates at district level, (IPNS estimates of yield rates being not easily available now).

III. Spot check of crop census records, 1937, 1949-50 and 1950-51

20. For a very large portion of the cadastrally surveyed area of the country the statistics of acreage under important crop are obtained every crop season by the census method. The data are not collected by interviewing the cultivators. Actual visits to all the fields in a village are to be made; and therefore the data should

be relatively free from sizeable ascertainment errors, especially, when these are collected by the *patwari*, a village official, who is well informed about the crops grown in the locality.

21. The *patwari*, however, is a land revenue official and the crop records are primarily maintained for this purpose; and it is pertinent to enquire whether or not the data fulfil the more exacting requirements of crop statistics. He is moreover burdened heavily with multifarious duties and therefore actual visits to the fields may not always be possible.

22. A spot-check of the census records was therefore undertaken in the *Rabi* season, 1949-50, the work being under the control of the Department of Economic Affairs of the Ministry of Finance. The Government secured the services of a very senior officer with more than 20 years experience in land revenue work. Moreover, this officer had himself set up a crop census organisation in one of the States.

23. A batch of specially experienced investigators made visits to a sample of fields in the presence of the *patwaris*, and after securing their agreement noted the actual utilisation of the fields (whose areas were accurately known); the corresponding entries in the census records were also noted. A similar spot-check was again conducted during the *Rabi* season of 1950-51. A summary of the findings in respect of wheat, barley, gram, *arhar*, *mator*, mustard and linseed is given in Table 2.

TABLE 2. COMPARISON OF CROP ACREAGES AS REPORTED BY PATWARIS AND CHECKERS, 1949-50 AND 1950-51

name of crop[a]	number of villages	number of comparisons[b]	acreage		sum of		discrepancy as % of col. (5)	
			patwari	checker	positive discrepancies	negative discrepancies	absolute	algebraic
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			1. 1949-50					
1.1 wheat	391	8097	3492.0	3030.9	965.8	−504.7	49	15
1.2 barley	391	4144	1185.7	1117.8	468.5	−400.5	78	6
1.3 gram	391	6519	3010.3	3027.1	866.1	−883.0	58	− 1
1.4 arhar	391	1120	298.9	494.1	57.9	−253.2	63	−40
1.5 mator	391	1340	195.7	260.4	54.2	−119.0	67	−25
1.6 mustard	391	3071	293.7	693.6	115.8	−515.8	91	−58
1.7 linseed	391	1145	134.0	269.2	59.3	−194.6	94	−50
			2. 1950-51					
2.1 wheat	167	1170	1005.2	863.6	205.3	− 63.7	31	16
2.2 barley	167	277	91.7	81.3	23.5	− 13.1	45	13
2.3 gram	167	1819	1739.0	1301.0	602.8	−164.7	59	34
2.4 arhar	167	610	183.9	255.4	82.7	−154.2	93	−28
2.5 mator	167	357	105.5	189.9	15.6	−100.0	61	−44
2.6 mustard	167	247	58.1	38.0	26.3	− 6.2	85	53
2.7 linseed	167	130	174.1	26.7	151.1	− 3.6	580	552

[a]Crops in mixture allocated to components in 1949-50, but completely excluded in 1950-51.
[b]Does not include where neither *patwari* nor checker reports the specified crop.

24. It should be pointed out at the outset that the 1949-50 results are not comparable with those of 1950-51 as mixed crops were excluded in the later year but not in the previous one. There are also differences in geographical coverage, and in the method of selection of sample villages and sample fields (plots). In the earlier check the choice of the villages was conditioned by the selection of *patwaris* made by their supervising *kanungos* and also by the avoidance of villages where harvesting had already started. Also in the selection of the plots within a sample village attempt was made to cast the sample haphazardly so as to cover all portions of the village. In the later spot-check both the villages and the fields were selected at random. However, a substantial number of these villages could not be covered by spot-check as the *patwari* could not be contacted, or even when contacted, he had not completed the crop records due to other preoccupations like the population or the livestock census. There were also some villages which did not fall under the jurisdiction of any *patwari*.

25. Growing of crops in mixture is fairly common during the *Rabi* season. In the case of fields with mixed crops, the proportion of the field under mixed crop was shown separately. Also the acreages under the different constituents were generally separately recorded on the basis of eye-estimation. However, for the 1949-50 data there were a few cases where this allocation was not recorded and the total area for these cases was divided equally among the constituent crops at the tabulation stage. For the 1950-51 data this question did not arise because the analysis was restricted to crops grown singly. The 1949-50 comparisons as given in the Table show large absolute discrepancies ranging from 49 per cent for wheat to 94 per cent for linseed. Part of the discrepancy is undoubtedly due to rather defective[5] instructions to the *patwari*. For example, he is asked to ignore all minor constituents and report the entire area under the major component. The effect of this for a crop like linseed, which is known to be grown extensively as a minor component in a mixture, is obvious. Although as a component it may be minor its total contribution is very large compared to the acreage under pure (unmixed) linseed. The large negative algebraic discrepancy (−50%) for 1949-50 must be, at least partly, due to this factor. The position is similar with mustard. Some adjustments for the above shortcomings are reported to be made before the publication of official acreage statistics; obviously, the basis for such cannot be the conditions obtaining in the year for which the statistics are reported. The situation calls for drastic steps and not mere adjustments.

26. A second source of discrepancy for which all *patwaris* may not be fully responsible is the unavoidable adoption of a rule of allocation to components of mixed crop at the tabulation stage (1949-50). The checker was asked to record separately the acreages of the constituents of a mixed crop. But *patwaris* in certain States were, according to official instructions, not expected to show the separate allocations. In such States for purposes of official acreage statistics, allocation of

[5] In relation to agricultural statistics, but presumably not so for land revenue purposes.

the area to different components is made at the level of district each comprising about 2000 villages on the average. This means that the total district area under a particular type of crop mixture is allocated to the components according to the ratio[6] fixed long ago, usually at the time of last settlement of land revenue, such settlement operations taking place at intervals of 20 or 30 years. This is obviously unsatisfactory. (Incidentally the problem of completely satisfactory procedure of allocation of mixed crops is still awaiting solution).

27. The magnitude of the discrepancies observed in 1949-50, however, is so large that it is hardly possible to explain this to be solely due to causes for which the *patwari* is not directly responsible. To obtain a clearer grasp of the role played by the *patwari* the 1950-51 data were analysed only for plots (and sub-plots) for which crops in intermixture were not reported. For such comparisons the discrepancies must be attributable to *patwari* performance. There is unmistakable evidence that the census records were far from accurate; the absolute discrepancy ranged from 31 per cent for wheat to 580 per cent for linseed.

28. The 1949-50 spot-check was not on a random sampling basis, and the 1950-51 check, although planned on a random basis, could not be executed as such; and therefore it is strictly not possible to assess the *net* effect of the ascertainment error on the estimation of total acreage under a crop. The random components of the ascertainment errors of individual units may balance to some extent on aggregation. To study the extent of this balancing we are presenting below summary results of one of our older studies conducted in 1937.

29. A compact area, called *thana*, 78 square miles in area, (comprised of 10 'units' covering a total of 108 *mauzas* or villages) was completely covered by a census; the acreage under jute was recorded plot by plot, for more than one-sixth of a million fields. A sample of 11 *mauzas* was selected at random, and a second complete enumeration was independently carried out in these *mauzas* by a different set of investigators. Again a systematic sample of every 20th plot with a random starting point was independently enumerated in all the mauzas. (There were other types of re-enumeration which need not be stated here). The survey was conducted by the Director of Agriculture, Bengal, and the statistical analysis of the data was undertaken by the Indian Statistical Institute.

30. The average discrepancy between the jute acreages according to the two enumerations are given below for the different units of increasing average sizes, namely, plot (0.29 acre), village (462 acres), union (7.8 sq. miles) and *thana* (78 sq. miles). At the plot level the average discrepancy is about 54 per cent of the average jute acreage (according to census). The percentage discrepancy varied considerably from one sample village to another—the median value is 26 per cent and the five middle most values ranged from 12 per cent to 33 per cent. As far as can be judged from comparison of duplicated observations on an average of about 859 plots per

[6] All these ratios were not (and still are not, in spite of best efforts) known to us at the time of analysis of data.

union the median percentage discrepancy at the union level comes to about 30 per cent; and for the *thana* the percentage discrepancy on the basis of 8586 plots (random systematic) works out to nearly 18 per cent.

31. From the magnitude of the discrepancies noted above it appears that a census may not provide accurate estimates for small administrative areas. The random component of the non-sampling error may balance considerably for very large administrative areas with the result that for these large areas the results may not differ appreciably from those obtained by a carefully conducted sample survey if the bias component is comparatively small. But care should be taken not to interpret that what holds good for a large area also holds good for each (or majority) of its constituent (small) administrative units.

IV. Sample verification of livestock census, 1956

32. In most of the States of the Indian Union, the eighth quinquennial livestock census was conducted during February to April 1956 and the data were subsequently brought up to the reference date of 15 April 1956. The enumeration was done by village officials like *patwaris*. The collection and compilation of these statistics were the responsibility of the State authorities. The Central Government in the Ministry of Food and Agriculture desired an independent sample verification by a different agency, and, as a consequence, the National Sample Survey (NSS) Organisation undertook this work during June-July 1956.

33. For rural areas 1624 sample villages, selected from among those just covered for socio-economic enquiries in the tenth round of NSS were taken up. Similarly a sample of 340 urban blocks (1951 Population Census Enumeration districts) were selected. For detailed enquiry 20 households were selected at random from each of the sample villages; and in urban blocks all the households were covered. In the sample units all non-household livestock establishments, which were very few in number, were also covered.

34. Data were collected about the livestock in possession of a household on the date of survey together with information about changes since 15 April 1956, so that the number as on the census date of reference could be obtained. The census registers were consulted only after the collection of these data, and census entries corresponding to the sample households were copied on the sample verification schedule. This was done in the sample village or urban area itself in order to minimise 'matching' difficulties.

35. It must be pointed out that the NSS investigators (enumerators) are wholetime quasi-permanent staff employed on a continuing basis, and are especially trained and experienced in the collection of data (including those on livestock) by the interview method, and their work is under intensive supervision. The data thrown up by them may therefore be considered to be of higher validity, and the census results may be evaluated against the results of the sample verification as standard.

36. Before passing on to the actual comparison of the results as shown in Table 3 we may refer to certain general features. Household by household enquiry revealed that in nearly one-fourth of the households having livestock the census enumerator did not make an actual visit. This fraction varied considerably from State to State. It did not however mean that the census enumerator made completely arbitrary entries for these households; he might have ascertained from neighbours, or, being a local man, the recording might even have been from his personal knowledge.

37. It was not always possible for the NSS investigators to secure the census figures for the sample households for various reasons. For example, the village might not have been covered in the census (at least up to the sample verification time), 'non-availability' of census records in the village (which may either mean that census had not taken place, or that for some reasons the records had been removed to some other place), 'clubbing' together two or more households in the census records, etc. Cases of omission to record the census entry against the verification figure might also have been caused by 'matching' difficulties, but such a contingency is likely to be very rare because the census enumerator (a local person usually the head-man or village *patwari*) from whom the census records were collected must have helped in identifying a sample household in his records unless, of course, when it was a case of omission on his part. In about 9.43 per cent of the rural and 5.61 per cent of urban sample households census figures were not recorded. A careful scrutiny of the investigators' 'remarks' showed that there were very strong grounds to conclude that some (at least one-eighth) of these cases were due to census omission.

38. We now pass on to the comparison of sample verification estimates with the census figures. We restrict ourselves to rural areas which account for by far the larger proportion of total livestock. Table 3 gives the necessary information not only for all-India but also for the major States. Manipur, Orissa and West Bengal are omitted because no census took place before the sample verification.

39. It will be noticed that there are highly significant differences in the all-India estimates of total cattle and total buffaloes. There is serious under-estimation in census figures, an increase of about 15 per cent is needed to bring up the census figure to the level of sample verification estimates. At the State level also there are significant differences.

40. Another set of estimates called the census-sample (cs) estimates has also been obtained in order to have some idea of the effect of different sources of error. Table 4 gives the necessary information for all-India (rural) in respect of total animals, total males, total females, working bullocks, and cows in milk, separately for cattle and buffaloes. The sample verification (sv), the census-sample (cs) and census (c) estimates are all shown in this Table.

41. In making the census-sample estimates we have made complementary use of census as well as survey data. From the census we have taken the counts of head of livestock for the sample households, and from the survey we have taken

TABLE 3. COMPARISON OF CENSUS COUNTS AND SAMPLE VERIFICATION ESTIMATES OF LIVESTOCK NUMBERS (MILLIONS), 1956

Rural sector: All-India and Major States

state[a]	sample surveyed		total cattle					total buffaloes				
	villages	h.hs.	census[b]	sample verification	difference	s.e.	diff. s.e.	census[b]	sample verification	difference	s.e.	diff. s.e.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. Andhradesh	88	1721	5.75	7.38	−1.63	0.64	2.55*	4.02	5.15	−1.13	0.42	2.69**
2. Assam	41	801	5.27	7.25	−1.98	0.81	2.44*	0.52	0.48	0.04	0.11	0.36
3. Bihar	192	3505	15.11	14.86	0.25	0.87	0.29	3.52	3.19	0.33	0.18	1.83
4. Bombay	119	2312	11.39	13.31	−1.92	0.80	2.40*	4.02	3.92	0.10	0.26	0.38
5. Hyderabad	80	1548	9.48	11.08	−1.60	0.76	2.11*	2.39	2.50	−0.11	0.16	0.69
6. Madhya Bharat	36	682	6.27	7.08	−0.81	0.49	1.65	1.91	2.39	−0.48	0.29	1.67
7. Madhya Pradesh	95	1866	15.35	15.83	−0.48	1.50	0.32	2.48	2.97	−0.49	0.39	1.26
8. Madras	139	2705	10.16	9.80	0.36	0.71	0.51	2.20	2.23	−0.03	0.22	0.14
9. Mysore	48	907	4.82	6.38	−1.56	0.65	2.40*	1.17	1.57	−0.40	0.16	2.50*
10. Punjab	55	1083	4.27	5.13	−0.86	0.52	1.65	2.76	3.29	−0.53	0.29	1.83
11. Rajasthan	66	1270	11.26	12.52	−1.36	0.82	1.54	3.20	3.59	−0.39	0.40	0.98
12. Uttar Pradesh	278	5526	22.52	27.92	−5.40	0.82	6.59**	9.57	11.47	−1.90	0.51	3.73*
13. All-India[c]	1434	27729	134.32	154.86	−20.54	2.97	6.92**	41.11	46.78	−5.67	1.11	5.11**

[a]States prior to 'Reorganisation'.

[b]Adjusted, whenever necessary, to States prior to Reorganisation on the basis of district census figures; magnitude of adjustment quite negligible compared to total.

[c]Excludes Manipur, Orissa and West Bengal, in which 4, 72 and 96 villages were respectively covered by NSS.

*Significant at the 5 per cent level.

**Significant at the 1 per cent level.

the counts of the number of households in the sample villages. In view of the fact that for some of the sample households census counts were not recorded for reasons explained earlier we proceeded as if those households were really taken into consideration by the census enumerator and assumed further that the average characteristics of those households were the same as those for which census data were available. (The latter assumption holds broadly for sample-verification data).

TABLE 4. COMPARISON BETWEEN SAMPLE VERIFICATION (sv), CENSUS-SAMPLE (cs) AND CENSUS (c) ESTIMATES : 1956

Rural sector : All-India[a]

number of sample villages : 1334 number of sample households : 27729

livestock category	sample verification[b] (sb)	census sample[b] (cs)	census[b] (c)	adjustment factor: total sv/c	adjustment factor: coverage-compilation cs/c	adjustment factor: enumeration sv/cs
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. *cattle*						
1.1 total cattle	154.9	148.8	134.3	1.152	1.108	1.040
1.2 total males	86.2	83.2	73.8	1.168	1.126	1.037
1.3 total females	68.6	65.7	60.4	1.136	1.087	1.045
1.4 working bullocks	61.8	60.3	53.2	1.162	1.135	1.024
1.5 cows in milk	15.9	17.7	17.1	0.930	1.038	0.896
2. *buffaloes*						
2.1 total buffaloes	46.8	44.5	41.1	1.138	1.082	1.052
2.2 total males	12.0	11.2	10.5	1.136	1.068	1.064
2.3 total females	34.8	33.2	30.6	1.139	1.087	1.048
2.4 working bullocks	5.2	4.9	4.9	1.053	1.002	1.051
2.5 cows in milk	10.1	11.3	10.8	0.932	1.046	0.891

[a] Excludes Manipur, Orissa and West Bengal.

[b] Figures in millions.

42. The three estimates are subject to errors from various sources which may be broadly divided into the following classes : (1) coverage error—omission (or duplication) of villages or blocks, (2) listing error—omission (or wrong inclusion) of households in selected villages or blocks, (3) errors in enumeration (and categorisation) of livestock for the households concerned, (4) errors in compilation and editing etc. The sample-verification and census-sample estimates are in addition subject to sampling error. The listing error may be regarded as coverage error excepting in the situation described below. In this census there is reason to believe that in some instances the enumerators deliberately omitted to record households without livestock. To the extent this is done correctly the failure to list does not contribute anything towards coverage error. When, however, non-possession of livestock happened to be a wrong assumption such omission should strictly be regarded as enumeration error.

43. It is believed that sample-verification (sv) estimates are subject to comparatively small amount of errors of the above type. We may take the sample verification estimate as standard against which to compare the other estimates.

44. It is clear from the method of obtaining the census-sample (cs) estimates that these are at par with sample-verification (sv) estimates, excepting for errors of enumeration of livestock and therefore the ratio sample-verification estimate/census-sample estimate or sv/c may be taken as an index of (or an adjustment factor for) the net census-enumeration error. The last column in Table 5 gives estimates of enumeration adjustment factor.

45. Again one would expect agreement (excepting to the extent sampling error comes into play) between the census (c) estimates and census sample (cs) estimates if census figures were relatively free from coverage, compilation (including editing) errors. We can therefore use the ratio (census-sample estimate)/(census estimates) or cs/c as an index of (or adjustment factor for) *net* coverage-compilation error. The coverage-compilation adjustment factor ranges from 1.038 to 1.135 for the cattle categories noted in Table 4. The corresponding range for buffaloes is 1.002 to 1.087. Generally the coverage-compilation error affects the aggregate figures more seriously than enumeration errors.

46. An index or adjustment factor for the total error is given by the ratio sample-verification estimate/census estimate, or sv/c. It will be noticed that we have defined the adjustment factors in such a manner that total adjustment factor is equal to the product of the coverage-compilation adjustment and the enumeration adjustment factor.

47. Leaving apart cows in milk for the present the total adjustment factor is around 1.15 for the various cattle categories, and nearly 1.14 for the various buffaloe categories (excepting working bullocks for which it is 1.05).

48. But for cows in milk, in spite of the fact that coverage-compilation adjustment factor is greater than unity, there is so much overenumeration in census that the total adjustment factor becomes less than unity. It is to be remembered, however, that placement in the category cows in milk is likely to be subject to considerable ascertainment error specially when the cow is nearing the end of its lactation period. Moreover, there is some amount of seasonal variation which might have added to the difficulty of ascertaining the position as on the census date; (both the census and the sample survey were spread over a fairly long period before and after this date respectively).

49. The total adjustment factors are themselves subject to errors of sampling. We have estimated these for the all-India categories, total cattle and total buffaloes,—the standard errors are 1.92 per cent and 2.36 per cent of the respective adjustment

TABLE 5. FREQUENCY DISTRIBUTION[a] OF TOTAL (t), COVERAGE-COMPILATION[b] (c), AND ENUMERATION[b] (e) FACTORS OF ADJUSTMENT : LIVESTOCK CENSUS AND SAMPLE VERIFICATION, 1956

adjustment factor		cattle														buffaloes													
		total cattle			total males			total females			working bullocks			cows in milk			total buffaloes			total males			total females			working bullocks			cows in milk		
		t	c	e	t	c	e	t	c	e	t	c	e	t	c	e	t	c	e	t	c	e	t	c	e	t	c	e	t	c	e
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)
1.	1.50–	—	—	—	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—
2.	1.40–1.50	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1	—	—	1	—	—	1	—	—	—	—	—	—	—
3.	1.30–1.40	2	1	—	2	3	—	—	1	—	2	2	—	1	1	—	1	—	—	4	1	—	1	—	1	1	—	—	1	2	—
4.	1.25–1.30	1	1	—	1	—	—	2	—	—	1	1	1	1	—	—	2	—	1	—	—	—	—	—	—	1	—	1	—	—	1
5.	1.20–1.25	3	—	1	1	—	1	4	1	2	1	1	—	1	1	—	—	2	—	2	1	1	2	2	1	1	3	—	—	—	—
6.	1.15–1.20	4	3	1	6	4	—	1	2	—	4	—	—	—	1	—	3	1	1	—	2	2	4	1	2	—	—	2	—	—	—
7.	1.10–1.15	2	4	—	1	2	1	4	3	1	1	3	1	1	2	—	2	1	2	3	2	2	2	—	—	1	—	2	1	1	—
8.	1.05–1.10	—	1	3	—	1	2	1	1	1	2	2	—	1	—	—	—	3	6	—	1	2	2	2	4	3	—	2	—	1	—
9.	1.00–1.05	1	1	4	1	2	6	1	2	5	2	1	6	2	3	2	4	—	2	1	—	5	—	1	4	—	2	4	—	3	—
10.	0.95–1.00	2	1	4	2	—	3	1	2	4	1	1	4	1	1	4	1	—	—	1	—	—	2	1	—	—	—	—	4	1	4
11.	0.90–0.95	—	—	—	—	—	—	1	—	—	—	—	1	2	—	1	2	2	1	3	1	1	1	3	1	1	1	1	3	2	2
12.	0.85–0.90	—	—	—	—	—	—	—	—	—	—	—	—	1	2	2	—	1	—	—	1	—	—	1	—	1	—	1	1	—	4
13.	0.80–0.85	—	—	—	—	1	—	—	—	—	—	1	—	1	1	3	—	1	—	—	—	—	—	—	—	1	—	—	3	2	—
14.	0.75–0.80	—	1	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1	2	—	—	—	—	3	3	—	—	—	2
15.	0.70–0.75	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	1	—	—	1	—	1	—	—	—	2	—	1	1	—
16.	0.60–0.70	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—
17.	0.50–0.60	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—
18.	–0.50	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—
19.	total	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13	15	13	13

[a]Each case corresponds to one of the following 14 States (rural) : Andhradesh, Assam, Bihar, Bombay, Hyderabad, Madhya Bharat, Madhya Pradesh, Madras, Mysore, Orissa, Punjab, Rajasthan, Uttar Pradesh and West Bengal, the 15th case being all-India (rural) which includes other (minor) States.

[b]For coverage-compilation and enumeration adjustment factors Orissa and West Bengal could not be included as Census took place after the sample verification.

TABLE 6 DISTRIBUTION OF SAMPLE HOUSEHOLDS BY THE NUMBER BY WORKING BULLOCKS (CATTLE) REPORTED AS PER LIVESTOCK CENSUS AND SAMPLE VERIFICATION SURVEY : 1956

Rural sector : All-India
number of sample villages : 1434[a]

working bullocks as per sample survey	working bullocks as per census																									
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	20	21	24	26	28	39	40	not available	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
0	13110	199	260	8	13	3	5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1135	14733
1	240	2308	261	19	12	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	327	3168
2	337	196	5650	177	106	7	5	—	1	—	1	—	—	—	—	1	1	—	—	—	—	—	—	—	—	793	7275
3	17	7	117	472	62	8	3	2	2	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	66	757
4	23	4	102	52	849	33	27	—	2	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	99	1192
5	—	—	10	2	31	91	17	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	6	159
6	5	—	8	5	19	7	172	5	12	—	3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	13	249
7	—	1	—	—	—	—	7	26	3	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	38
8	1	—	1	—	5	1	14	2	44	2	3	—	3	—	—	—	—	—	—	—	—	—	—	—	—	5	81
9	—	—	1	—	1	1	1	1	1	7	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	15
10	—	—	—	—	1	—	1	—	2	1	12	—	1	—	—	—	—	1	—	—	—	—	—	—	—	3	22
11	—	—	—	—	—	—	—	—	—	—	1	5	1	—	—	—	—	—	—	—	—	—	—	—	—	—	7
12	—	—	1	—	—	—	1	—	—	—	2	—	4	—	—	—	—	—	—	—	—	—	—	—	—	—	8
13	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1
14	—	—	—	—	—	—	—	—	—	—	—	1	1	—	5	—	2	—	—	—	—	—	—	—	—	—	9
15	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	2
20	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1	1	—	—	—	—	—	1	4
21	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1
22	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	—	—	—	—	—	—	2
24	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	1
26	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	1
27	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1
28	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—	—	—	1
30	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	1
40	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	1
total	13734	2716	6411	735	1099	152	253	37	68	10	25	6	10	1	5	2	4	1	3	2	1	1	1	1	1	2450	27729

[a]Excludes Manipur, Orissa and West Bengal.

factors. For enumeration adjustment factors the percentage standard errors are roughly half of those shown above.

50. All the adjustment factors have also been calculated for the rural areas of 12 major States. The census took place in Orissa and West Bengal subsequent to the sample verification. For these States it was possible to calculate only the total adjustment factor. We have presented in Table 5 a consolidated picture of the situation. Since the sample size for most of the States was comparatively small not much significance can be attached to individual adjustment factors. But taken together they reveal significant tendencies.

51. It will be noticed that except for cows in milk there is general tendency for adjustment factors to be greater than unity. This tendency is more marked in case of cattle. Although no great significance need be attached to specific cases one may conclude that there are in fact a few States where the actual number in certain categories exceeds the census number by 25 per cent or more.

52. In order to obtain a more detailed case by case picture of enumeration error we have obtained a few two-way tables showing the distribution of all-India (rural) sample households by number of head of livestock as reported by the NSS investigator and as recorded by the census enumerator. The table for total cattle would have been preferred but to save space we are presenting, by way of illustration of general features, Table 6 for working bullocks (cattle) only.

53. Out of nearly twenty-eight thousand households covered by NSS the comparison was possible for 91 per cent of cases. Out of these comparable cases in 38 per cent of the households there was no cattle according to NSS. But in nearly 5 per cent of these no-cattle households, the census recorded one or more head of cattle, the average being more than two. Among those with cattle, the census recorded complete absence in about 5 per cent of cases, and in two-thirds of former households there was complete agreement between the two agencies. In spite of this agreement the average discrepancy between the census and NSS entries was considerable in relation to the number of head of cattle possessed by a household. In fact, the standard deviation of discrepancy was as much as 56 per cent of the average number of head of cattle per household. The picture for working bullocks (cattle) is very similar, the corresponding standard deviation being of the same order. The *gross* enumeration-error was therefore very high.

54. We now come to two important points which are sometimes made in favour of census as against sample survey. First, the census can provide reliable counts for small administrative units (say, villages, each with 100 households on the average), and secondly, accurate measurement of change between two fairly close points of time are furnished by the census. It is obvious from the nature and magnitude of the enumeration and coverage compilation errors that the livestock census has hardly satisfied these objectives.

V. Survey of landholdings, 1954-55

55. In the eighth round, 1954-55 of the National Sample Survey we had for the first time an opportunity to organise the field work in a completely independent interpenetrating system. The entire sample for a State was broken up generally into 12 independently drawn sub-samples, and 4 of these were covered by the Central agency and the remaining 8 by the State agency. In a few States there were additional sub-samples which were not taken into account in any of the Tables and Charts excepting Table 11 and Chart 3. The data thrown up opened the possibility of examining the relative biases of the two agencies. It should be pointed out here that the Central agency (Directorate of NSS) is a quasi-permanent agency purely for statistical surveys, the investigators (enumerators) being whole-time employees with considerable experience of data collection by the interview method. The State agencies participated on an *ad-hoc* basis, and employed their normal staff, usually of the Land Revenue and Agricultural Departments, for this purpose. An enquiry into landholdings and several other socio-economic enquiries were taken up simultaneously by the Central agency but the State agencies took up the first enquiry only.

56. It may be pointed out that even when observed differences are 'statistically significant', whether this is of any practical importance or not requires to be judged against the 'permissible error' for the purpose in view. The present enquiry, for example, was made primarily to collect information to decide broad policies of land redistribution; for such purposes the concentration curves (shown in Charts 1, 2, and 3) can supply very useful information. The sample holdings were arranged in ascending order of size and accumulated; and the estimated cumulative percentage of holdings is shown on the horizontal scale; the estimated cumulative percentage of the area under holdings is shown on the vertical scale. In Chart 1, which relates to 'household ownership holding' three concentration curves are shown separately, one for the Central sample, and two for State samples for which the information was collected respectively by party 1 and party 2 of investigators. Chart 2 is similar to Chart 1 but relates to 'household operational holding,' (which is defined as 'area owned' plus 'area leased in' minus 'area leased out' by household). In Chart 3, the two State samples have been pooled together; and only two concentration curves are shown respectively for the Central and the State sample for 'agricultural holding' in the upper pair of curve, and for '(total) operational holding' (which is constituted of all land under one distinct technical and economic unit so that in certain cases more than one household may be associated with the same 'total operational holding' and the same household may be associated with more than one '(total) operational holding') in the lower pair of curves. The real issue is whether the policy decision would remain the same by using *either* of the two concentration curves based respectively on the Central and State samples. In the present case, the divergence between the Central and the State concentration curves is so small that policy decisions would remain practically the same whether the Central or the

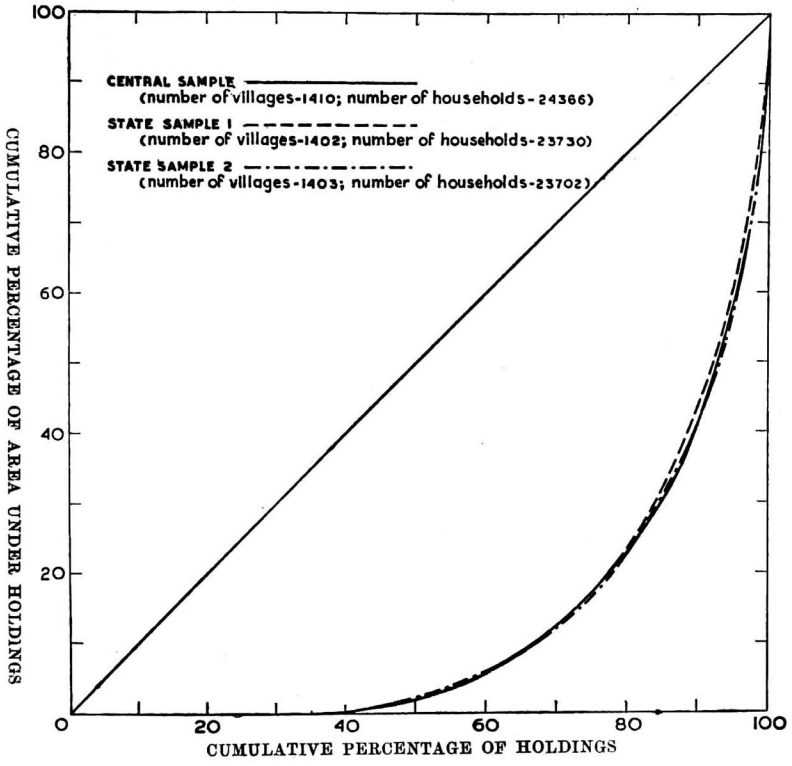

CHART (1) : CONCENTRATION CURVES FOR HOUSEHOLD OWNERSHIP HOLDINGS

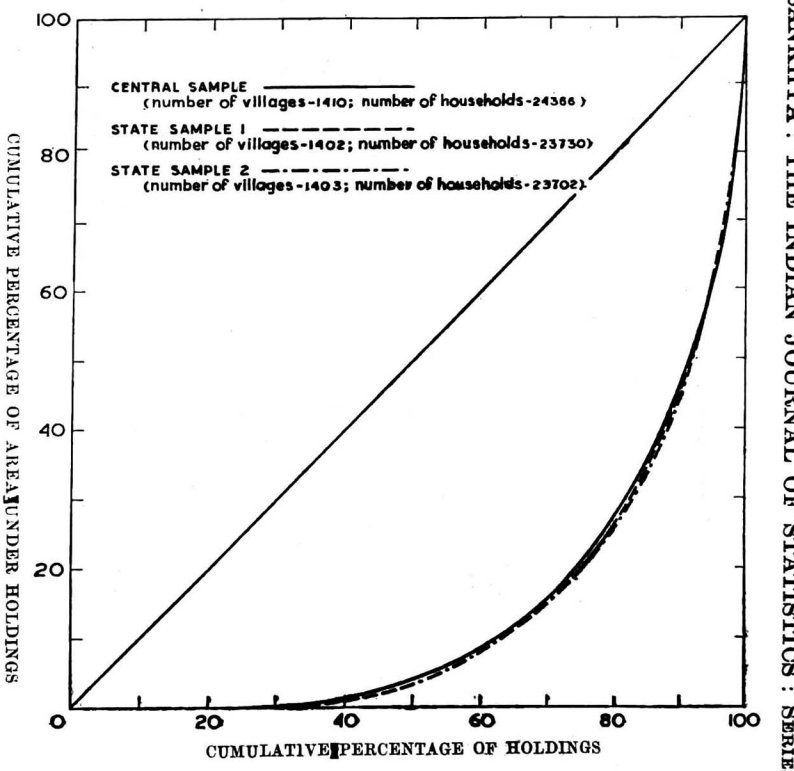

CHART (2) : CONCENTRATION CURVES FOR HOUSEHOLD OPERATIONAL HOLDINGS

State concentration curve is used; the divergence between the two (Central and State) concentration curves may therefore be considered to be well within the limits of permissible error.[7] It may be added that the converse case may also arise. The observed divergence between concentration curves or other results based on two sub-samples of an interpenetrating network of samples may not be statistically

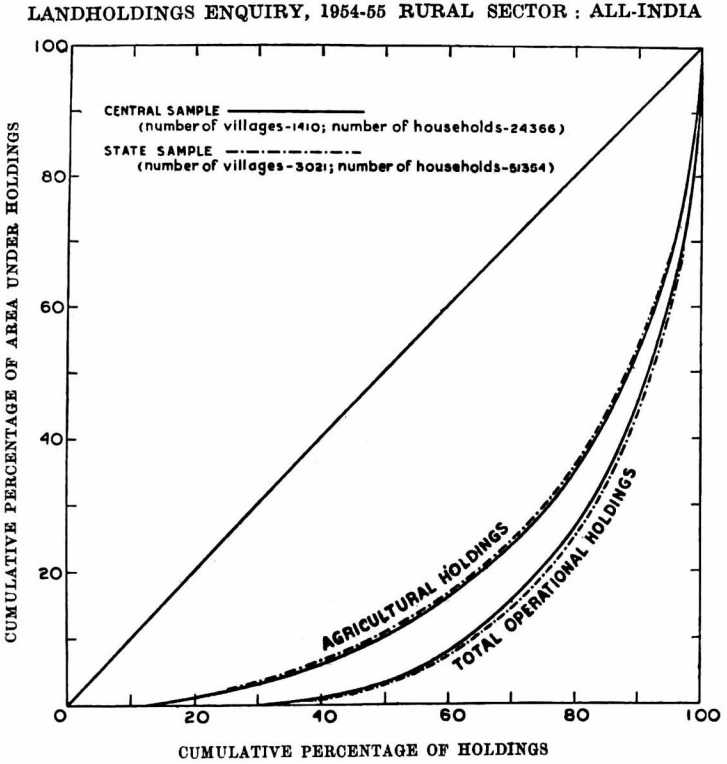

CHART (3) : CONCENTRATION CURVES FOR AGRICULTURAL HOLDINGS AND (TOTAL) OPERATIONAL HOLDINGS

significant and yet be wider than the limits of 'permissible error' for any given purpose. The statistically 'significant' differences which will be found in some cases in the following paragraphs, while indicating some lack of control, should be judged, however, in the light of the observations made here.

57. We are giving in Table 7 the all-India (rural) estimates provided by the Central and State agencies for some basic characters. The corresponding standard errors are also shown. For every State separate estimates were obtained for each of the sub-samples; these facilitated the estimation of the standard errors. It will be noticed that the Central estimate is significantly larger than the State estimate in

[7] The 'pooled' concentration curve which would lie 'between' the other two curves is not shown in any of the charts in order to avoid confusion arising from very close superimposition of several curves.

respect of 'number of households'[8] as well as the 'acreage operated'. In case of 'area owned' although the former estimate is larger the difference is not statistically significant. We arrive at the same conclusion by a simpler method to be explained immediately.

TABLE 7. COMPARISON OF CENTRAL AND STATE ESTIMATES FOR SOME BASIC CHARACTERS : LANDHOLDINGS ENQUIRY, 1954-55

Rural sector : All-India

number of sample villages : central sample—1410; state sample—2805
number of sample households : central sample—24366; state sample—47432

character	central		state		difference		std. error of diff.(6)[a]	diff./ std. error
	estimate[a]	std. error[a]	estimate[a]	std. error[a]	actual[a] (4)-(2)	p.c. (6)/(2) %		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. number of households	64.863	0.848	62.892	0.504	1.971	3.0%	0.986	2.00*
2. total acreage operated by rural households	348.151	7.553	330.402	4.934	17.749	5.1%	9.022	1.97*
3. total acreage owned by rural households	307.037	7.148	304.992	4.780	2.045	0.7%	8.599	0.24

* Significant at 5 per cent level.

[a] Figures in millions.

58. There are other characters for which comparisons between Central and State estimates are possible. For some of these comparisons simpler methods have been adopted. The general nature of the discrepancy between the Central and State estimates may be surmised by considering (1) the variation of the sub-sample estimates, and (2) the consistency, if any, in the nature of the set of discrepancies for a group of territorial divisions.

59. We have in Tables 8 and 9 provided all the 12 sub-sample all-India (rural) estimates, 4 for the Central and 8 for the States—for five classes of characters, namely, (1) aggregates, (2) average size, (3) rates per capita, in Table 8; and 4 distribution of households by size of holding, and (5) the area operated for each holding size class, in Table 9.

60. It may be pointed out incidentally that a few basic sub-sample estimates provide a simple means of obtaining sub-sample estimates of derived statistics, and thus the simple t-test which we have applied to the basic characters is also easily applied to the derived statistics. Thus we have given the results for the derived characters—total acreage leased in, the average sizes, and the rates per capita. Although not shown here, one can test out, from what is given, the significance of the difference in respect of average size for each holding size class.

[8] The Central and State samples were drawn quite independently, with probability proportional to 1951 population, and there were a few, usually large, common villages, (villages were the first stage units); all the households in a village were first listed independently before the sample households were selected for detailed enquiry. A direct comparison of the exact number enumerated is therefore possible. The Central agency recorded on the average 11 households against 10 recorded by the State agencies.

TABLE 8. COMPARISON BETWEEN CENTRAL AND STATE SAMPLE ESTIMATES: AGGREGATES, AVERAGE SIZES AND RATES PER CAPITA: LANDHOLDINGS ENQUIRY, 1954-55

Rural sector: All-India[a]

number of sample villages: sub-sample size—318 | number of sample households: sub-sample size—5416

specification of estimate	aggregates						average size				area per capita	
	number (in millions)			area (million acres)			household (number of persons)	operational holding (acres)	household operational holding (acres)	household ownership holding (acres)	operated (acres)	owned (acres)
	households	persons	operational holdings	operated	owned	leased in (net)						
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. *central estimates*												
1.1. sub-sample 1	61.0	298.7	59.5	294.7	270.7	24.0	4.90	4.95	4.83	4.44	0.987	0.906
1.2. sub-sample 2	58.8	285.3	58.0	323.0	274.0	49.0	4.86	5.57	5.50	4.66	1.132	0.960
1.3. sub-sample 3	59.9	292.8	59.4	313.6	268.4	45.2	4.89	5.28	5.24	4.48	1.071	0.917
1.4. sub-sample 4	56.5	280.1	55.7	290.4	257.4	33.0	4.95	5.22	5.14	4.55	1.037	0.919
2. *state estimates*												
2.1. sub-sample 1	56.3	286.7	54.8	278.2	254.9	23.3	5.09	5.08	4.94	4.52	0.970	0.889
2.2. sub-sample 2	56.7	283.5	54.9	281.0	248.4	32.6	5.00	5.11	4.95	4.38	0.991	0.876
2.3. sub-sample 3	58.3	295.0	57.2	269.9	246.1	23.8	5.06	4.72	4.63	4.22	0.915	0.834
2.4. sub-sample 4	56.8	280.5	48.4	301.6	276.3	25.3	4.94	6.23	5.31	4.86	1.075	0.985
2.5. sub-sample 5	58.8	301.1	57.3	300.4	269.9	30.5	5.12	5.24	5.11	4.59	0.997	0.896
2.6. sub-sample 6	57.2	285.3	55.7	289.1	271.0	18.1	4.99	5.19	5.05	4.74	1.013	0.950
2.7. sub-sample 7	56.1	285.0	53.6	300.0	271.9	28.1	5.08	5.59	5.35	4.85	1.052	0.954
2.8. sub-sample 8	56.1	280.2	48.6	277.5	270.5	7.0	4.99	5.71	4.94	4.82	0.990	0.965
3. *pooled estimates*												
3.1. central	59.0	289.2	58.1	305.4	267.6	37.8	4.90	5.25	5.17	4.53	1.056	0.925
3.2. state	57.1	287.2	53.8	287.2	263.6	23.6	5.03	5.34	5.03	4.63	1.000	0.918
3.3 difference (3.2–3.1)	−1.98	−2.05	−4.33	−18.23	−4.02	−14.2	+0.13	+0.09	−0.14	+0.10	−0.056	−0.007
3.4. difference (%) (3.3/3.1)	−3.4	−0.7	−7.4	−6.0	−1.5	−37.6	+2.7	+1.7	−2.7	+2.2	−5.3	−0.8
3.5. t	2.43*	0.44	2.30*	2.23*	0.61	2.51*	3.79**	0.36	0.93	0.80	1.73	0.24

* Significant at 5 per cent level. ** Significant at 1 per cent level for $d.f. = 10, t_{.05} = 2.228$, $t_{.01} = 3.169$.

[a] Some States have been omitted for convenience; only the States listed in Table 10 are included.

TABLE 9. COMPARISON BETWEEN CENTRAL AND STATE SAMPLE ESTIMATES : DISTRIBUTION F HOUSEHOLDS AND AREA OPERATED BY HOUSEHOLD OPERATIONAL HOLDING SIZE-CLASS : LANDHOLDINGS ENQUIRY, 1954-55

Rural sector : All-India

number of sample villages : sub-sample size—351 number of sample households : sub-sample size—5983

specification of estimate	number of households (million) in household operational holding size class (acres)						area operated (million) acres by household operational holding size class (acres)				
	0.00	0.01-0.09	0.10-0.99	1.00-4.99	5.00-9.99	10.00 & above	0.01-0.09	0.10-0.99	1.00-4.99	5.00-9.99	10.00 & above
1. *central estimates*											
1.1. sub-sample 1	4.08	14.12	8.72	19.95	10.39	8.93	0.36	4.0	53.8	72.5	201.8
1.2. sub-sample 2	4.25	13.01	8.64	20.07	9.14	10.02	0.34	4.1	55.0	65.2	244.2
1.3. sub-sample 3	3.68	16.05	7.29	20.17	9.63	8.93	0.43	3.5	54.1	67.2	229.6
1.4. sub-sample 4	3.39	13.46	8.21	19.23	9.05	9.05	0.43	4.0	52.6	64.4	215.0
2. *state estimates*											
2.1. sub-sample 1	8.05	10.58	7.71	17.60	9.18	8.64	0.30	3.7	47.4	64.7	204.2
2.2. sub-sample 2	8.22	9.56	8.70	18.35	9.02	8.35	0.27	4.1	49.0	64.4	201.4
2.3. sub-sample 3	8.29	10.49	9.66	18.65	8.82	8.77	0.33	4.5	50.8	62.5	196.3
2.4. sub-sample 4	8.51	9.02	8.51	18.57	8.65	9.47	0.25	4.0	51.0	61.4	232.9
2.5. sub-sample 5	9.17	9.85	7.95	20.44	8.63	8.89	0.29	3.8	54.4	62.1	220.1
2.6. sub-sample 6	8.16	10.28	7.87	18.98	8.86	9.06	0.29	3.7	50.3	62.9	223.4
2.7. sub-sample 7	8.29	9.06	6.84	10.36	8.55	9.30	0.29	3.3	50.4	61.1	218.9
2.8. sub-sample 8	8.13	10.17	8.01	18.17	9.43	8.34	0.29	3.7	48.2	66.0	206.1
3. *pooled estimates*											
3.1. central	3.89	14.15	8.21	19.85	9.58	9.18	0.39	3.90	53.9	67.3	222.7
3.2. states	8.35	9.87	8.16	18.17	8.89	8.86	0.29	3.86	50.2	63.1	212.9
3.3. difference (3.2–3.1)	+4.46	−4.28	−0.05	−1.08	−0.69	−0.32	−0.10	−0.04	−0.37	−0.42	−0.98
3.4. difference(%) (3.3/3.1)	+114.7	−30.2	−0.6	−5.4	−7.2	−3.5	−25.0	−1.0	−6.9	−6.2	−4.4
3.5. t	20.04**	7.82**	0.12	2.36*	2.56*	1.38	5.27**	0.20	3.26**	2.74*	1.09

* Significant at 5 per cent level. ** Significant at 1 per cent level; for $d.f.=10$, $t_{.05}=2.228$, $t_{.01}=3.169$.

61. It will be noticed from Tables 8 and 9 that the Central and State estimates are not always in agreement with each other. It is found that the State sample provides lower estimates of aggregates compared to the Central sample. There appears to be significant differences in regard to the estimation of 'total number of households', 'total number of operational holdings', 'total area operated' and 'total (net) area leased in' by all rural households from urban households. For these aggregates the State sample under-estimates by 3.4 per cent, 7.4 per cent 6.0 per cent and 37.6 per cent respectively.

62. It is of interest to note that the State sample provides a significantly lower estimate of 'total number of households', but the 'average household size' for that sample is significantly higher compared to the Central sample, and the product of the two, namely, 'total population' is not significantly different. It is difficult to interpret the result with confidence. One of the possibilities is that the State sample failed to cover fully the very small households, and the contribution of those small households to the total population size is small compared to the sampling fluctuations so that the above test failed to detect the discrepancy. Another possibility is that the State agencies in preparing the list of households in the sample villages leaned more on the list of cultivators which they had for purposes of land revenue, and moreover, they may not have always taken into account any recent partitioning of households which might really have taken place, if recording of such information did not make any material difference in regard to collection of land revenue.

63. It is only logical to compare the characteristics for each State separately, because agencies differed from State to State. It is to be noted, however, that the reduction of the sample size at the State level may make the detection of differences more difficult. It may on the other hand be pointed out that when pooled to all-India level the State bias (and errors) will balance to some extent and this may also make detection at all-India level difficult, even though sampling error may be reduced. We have therefore found it desirable to present a summary picture separately for the more important States also. For this purpose we have chosen the characteristic 'per capita area operated' the Central and State estimates of which do not differ significantly at the all-India level. We have also chosen for further examination the characteristics, 'number of households', 'number of operational holdings', 'total area operated' and 'average household size', for which the two estimates differ significantly.

64. It will be noticed in Table 10 that in majority of the States the sign of the difference between the Central and State estimates are the same as at the all-India level (shown in an earlier Table) for characteristics for which the difference is significant at the all-India level. Thus out of 14 comparisons the similarity holds in 10 cases for 'total number of households', 11 cases for 'total number of operational holdings', 12 cases for 'total area operated', and 10 cases for 'average household size.' It is interesting to note that although the 'per capita area operated' estimates do not differ significantly at the all-India level, yet in as many as 13 cases

TABLE 10. PERCENTAGE DIFFERENCE BETWEEN CENTRAL AND STATE SAMPLE ESTIMATES : LANDHOLDINGS ENQUIRY, 1954-55

Rural sector : Principal States[a]

state	sub-sample size		no. of h.hs.		no. of opl. holdings		total area operated		average hh. size		per capita area operated	
	villages	h.hs.	diff.%[b]	t	diff.%[b]	t	diff.%[b]	t	diff.%[b]	t	diff.%[b]	t
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. Andhradesh	18	344	−0.54	0.19	−7.91	0.97	−3.78	0.34	+5.49	3.43**	−8.07	0.82
2. Assam	10	229	−8.00	0.86	−10.41	1.15	−33.09	2.45*	−1.86	0.22	−29.84	3.39**
3. Bihar	36	557	−0.39	0.06	−3.10	0.48	−4.02	0.59	+0.76	0.60	−4.01	0.79
4. Bombay	36	577	−1.29	0.43	−10.44	2.12	−4.09	0.66	−0.38	0.14	−2.47	0.49
5. Hyderabad	18	335	+7.26	1.23	+1.34	0.17	+24.08	2.29*	+4.59	1.67	+10.88	1.50
6. Jammu & Kashmir	19	229	+7.97	0.49	+8.92	0.54	+8.24	0.44	+5.12	3.06*	−4.52	0.53
7. Madhya Pradesh	27	403	−2.88	1.01	−11.39	1.41	−2.13	0.30	+1.11	0.35	−0.69	0.13
8. Madras	24	389	+0.19	0.07	+0.74	0.24	−1.78	0.23	−0.21	0.80	−1.93	0.26
9. Mysore	10	220	−8.96	2.29*	−12.12	3.06*	−8.13	1.13	+8.25	2.25*	−6.74	1.15
10. Orissa	18	377	−4.72	0.86	−3.57	0.71	−7.30	0.76	−0.22	0.08	−2.61	0.37
11. Punjab	15	275	−11.39	1.51	−26.90	2.15	−16.96	1.16	+9.20	2.73*	−11.03	1.05
12. Rajasthan	18	244	+1.10	0.28	−7.34	0.92	−13.18	1.11	+0.39	0.16	−13.83	1.48
13. Uttar Pradesh	51	873	−3.83	1.04	−5.97	1.53	−11.72	2.36*	−6.12	2.77*	−13.08	2.80*
14. West Bengal	18	364	−18.07	2.18	−18.85	2.23*	−19.04	2.72*	+5.33	2.04	−8.89	1.13

* Significant at 5 per cent level. ** Significant at 1 per cent level; for d.f. = 10, $t_{.05} = 2.228$, $t_{.01} = 3.169$.

[a] Only States with more than 200 households per sub-sample are included.

[b] Difference : (State estimate—Central estimate) expressed as percentage of Central estimate.

TABLE 11. PERCENTAGE DIFFERENCE BETWEEN CENTRAL AND STATE ESTIMATES IN REGARD TO AGRICULTURAL HOLDINGS: LANDHOLDINGS ENQUIRY, 1954-55

zone	number of sample villages		number of sample households		number of agricultural holdings			area under agricultural holdings			average size of agricultural holding			proportion of agricultural holdings to total operational holdings		
	central	state	central	state	central (000)	state (000)	% difference[a]	central (000 acres)	state (000 acres)	% difference[a]	central (acres)	state (acres)	% difference[a]	central	state	% difference[a]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
1. north	204	407	3484	6997	9237	8064	−12.70	44571	38997	−12.51	4.83	4.84	+0.21	.8026	.7451	−7.16
2. east	331	696	6402	12868	13865	11800	−14.89	59014	51383	−12.93	4.26	4.35	+2.11	.8048	.7517	−6.60
3. south	223	503	4255	9651	8802	8333	−5.33	43922	42452	−3.35	4.99	5.09	+2.00	.6650	.6431	−3.29
4. west	168	372	2820	5991	3936	3899	−0.94	49778	49578	−0.40	12.65	12.72	+0.55	.6696	.6789	+1.39
5. central	244	524	3771	8532	6807	6444	−5.33	85572	89394	+4.47	12.57	13.87	+10.34	.6916	.6674	−3.50
6. north-west	240	519	3634	7315	4768	4336	−9.06	63594	56423	−11.28	13.34	13.01	−2.47	.7714	.7180	−6.92
7. all-India	1410	3021	24366	51354	47415	42876	−9.57	346451	328227	−5.26	7.31	7.66	+4.79	.7423	.7039	25.17

[a]% difference : (State estimate—Central estimate) expressed as percentage of Central estimate.

the Central estimate exceeds the State estimates. This is, of course, a significant result.

65. Going back to the earlier Table 9 we note the sub-sample estimates of the distribution of households by the six size-groups of household operational holdings; (the sizes were obtained correct to two places of decimals, so that 0.00 means less than .005 acres)—the size classes being 0.00, 0.01–0.09, 0.10–0.99, 1.00–4.99, 5.00–9.99 and 10.00 & above. The corresponding acreages are also shown for each group (excepting for 0.00). It will be seen that the differences are significant for all the groups excepting 0.10–0.99 and 10.00 & above. There is considerable difference between the two estimates for the two lowest size classes, 0.00 and 0.01–0.09. (The State sample has registered a very large proportion in the 0.00 class). But the difference ceases to be significant when the two lowest size classes are merged together. It is difficult to offer any very convincing explanation of the phenomenon without further probing analysis. It is not known whether the State investigators were inclined to record more approximate figures (so that rounding off to 0.00 would be more frequent), or whether thinking that the main purpose of the enquiry is to collect information about *agricultural* holdings they have not paid adequate attention to recording areas under house-site.

66. Some of the total landholdings may be used exclusively for non-agricultural purposes. Excluding these we obtain holdings each of which is wholly or partly put to agricultural use. Zonal[9] distribution of such agricultural holdings is shown in Table 11.

67. It will be noticed that relative to the Central sample the State sample under-estimates the number of such holdings (by 9.57 per cent). This feature is to be found in all the six zones. We have previously seen that taking total landholdings (which are comprised of all agricultural as well as non-agricultural lands) we have a similar phenomenon (under-estimation by 4.33 per cent) and it is pertinent to enquire whether or not the position regarding agricultural holdings is merely a reflection of the other phenomenon about total landholdings. It will be noticed from Table 11 that not only the State sample under-estimates the (total) operational holdings but even the proportion of agricultural holdings to (total) operational holdings is practically uniformly lower in the State sample. This may happen if the State sample behaves in the following manner—(1) larger omission of agricultural holdings, possibly the smaller ones, and/or (2) undue failure to record any agricultural utilisation, again, possibly in some of the smaller holdings. An examination of the data (not reproduced here) shows that in the lower size-classes the State sample registers a lower percentage of agricultural holdings. It will also be noticed in Table 11 that the average size of agricultural holding is greater in the State sample—possibly because of omissions of some of the smaller holdings.

68. The survey design described at the beginning of this section is incomplete in one respect. This is in regard to the assignment of different sub-samples to

[9] In the 1951 Population Census the States were suitably grouped to form six Population Zones.

different parties of investigators. It will be recalled that there were four sub-samples in the Central sample; two of them were covered by the first Central party and the remaining by the second Central party. There were similarly two parties of State investigators. In our earlier discussions we have ignored the question of party differences.

TABLE 12. ANALYSIS OF VARIANCE OF AVERAGE AREA OWNED PER HOUSEHOLD; LANDHOLDINGS ENQUIRY, 1954-55

state	sum of squares			mean square			F	
	agency $d.f.=1$	party $d.f.=2$	error $d.f.=8$	agency	party	error	agency	party
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Uttar Pradesh	0.252	0.176	0.150	0.252	0.088	0.019	13.26**	4.63*
2. Bihar	0.004	0.246	0.919	0.004	0.123	0.115	28.75(r)	1.07
3. Orissa	0.001	0.381	1.864	0.001	0.190	0.233	233.00(r)	1.23(r)
4. West Bengal	0.583	0.249	0.998	0.583	0.124	0.125	4.66	1.01(r)
5. Assam	1.480	0.050	17.776	1.480	0.025	2.222	1.50(r)	88.88*(r)
6. Andhra	0.064	0.381	2.740	0.064	0.190	0.342	5.34(r)	1.80(r)
7. Madras	0.059	0.080	1.230	0.059	0.040	0.154	2.61(r)	3.85(r)
8. Mysore	0.254	0.0016	7.281	0.254	0.0008	0.910	3.58(r)	1137.5**(r)
9. Travancore & Cochin	0.441	0.631	2.395	0.441	0.316	0.299	1.47	1.06
10. Bombay	0.175	0.919	3.778	0.175	0.460	0.472	2.70(r)	1.03(r)
11. Saurashtra	15.072	5.941	71.547	15.072	2.970	8.943	1.69	3.01(r)
12. Madhya Pradesh	0.037	1.673	4.935	0.037	0.836	0.617	16.68(r)	1.35
13. Madhya Bharat	0.134	3.266	9.278	0.134	1.633	1.160	8.66(r)	1.41
14. Hyderabad	4.673	2.422	12.922	4.673	1.211	1.615	2.89	1.33(r)
15. Vindhya Pradesh	6.120	1.703	16.394	6.120	0.852	2.049	2.99	2.40(r)
16. Rajasthan	1.649	4.819	45.394	1.649	2.410	5.674	3.44(r)	2.35(r)
17. Punjab	1.397	3.945	9.973	1.397	1.972	1.247	1.12	1.58
18. PEPSU	0.771	4.990	25.181	0.771	2.495	3.148	4.08(r)	1.26(r)
19. Jammu & Kashmir	0.0004	0.083	1.334	0.0004	0.042	0.167	417.50*(r)	3.98(r)

$F_{.05}(1,8)=5.32$. $F_{.01}(1,8)=11.26$. $F_{.05}(8,1)=238.9$. $F_{.01}(8,1)=5982$.
$F_{.05}(2,8)=4.46$. $F_{.01}(2,8)=8.65$. $F_{.05}(8,2)=19.37$. $F_{.01}(8,2)=99.37$.

(r) indicates error/party or error/agency in F-ratio.

* Significant at 5 per cent level. ** Significant at 1 per cent level.

69. In order to test whether there is any significant difference between the party estimates (and between Central and State estimates) the technique of analysis of variance has been applied on the character 'area owned per household' as estimated from the survey. The results for each of the 19 States (having 12 replications or sub-samples each) are shown in Table 12. The total degrees of freedom (11) has been split up as follows—between agency (1), between parties within agency (2), and error (8).

70. It will be noticed that there was significant difference between parties within agency for the States of Uttar Pradesh, Assam and Mysore. Agency differences were significant for Uttar Pradesh and Jammu & Kashmir.

VI. SAMPLE SURVEY FOR ESTIMATION OF RATE OF GROWTH OF POPULATION 1958-59, 1959-60

71. It is not an uncommon practice in the NSS to have a built-in system of checking in the survey design. There is in the first instance the system of completely independent field work as well as processing—one by the Central agency and the other by State agencies. This interpenetrating arrangement provides two valid but independent estimates of all characters under study.

72. There are other systems which provide comparisons, the elements of which are all provided by the same agency. We shall illustrate one such method used in our current plan in regard to the estimation of rate of growth of population which is obviously of fundamental importance in any planning for national development. The survey, conducted by the interview method, is restricted to the rural areas in the first instance. Our vital registration system being rather unreliable we had to explore the possibilities of sample survey methods.

73. Control of sampling and other errors is sought to be achieved in two ways. First, control of sampling error by having a fairly large number of sample villages and canvassing necessary information from each and every household in them. It is also believed that the *prima facie* acceptability or non-acceptability of vital rates for the sample villages obtainable from their complete coverage has helped to reduce and control errors of omission etc. Again complete coverage has helped cross-checking of dates of births and deaths in the neighbouring households by studying the time sequence and the interval between those events; the deaths to single member households, no longer existing at the time of enquiry, are also obtainable only on the complete enumeration of villages by enquiring about such cases in, say, every 10th household. Also the village as a sample unit has definitely better technical and practical advantages over a sample of households in re-enumerating the population, narrated later. Secondly, by making the reference period sufficiently long and sampling fluctuations are reduced further. This step naturally increased the magnitude of ascertainment biases due mainly to recall lapse.[10] A short reference period would increase the sampling error itself beyond desirable limits as the resources at our disposal did not permit us to increase the number of sample villages. Moreover, there were reasons like relatively larger border bias if too short a reference period was chosen.

74. By border bias we mean undue (net) inclusion or exclusion of vital events actually occurring at points of time around the end-points of the reference period. In the NSS the normal practice is to have as reference a period of specified length

10 It is the general practice in NSS to introduce in its earlier rounds some studies of a 'pilot' nature before taking up the subject in question in a full-fledged manner. These studies are not small scale *ad-hoc* ones, but are extensive surveys so that the effect of large-scale operational conditions may be adequately reflected in the results. For similar reasons the exploratory nature is not usually fully articulated. The experience gained will help in reducing errors and biases in subsequent surveys. Earlier studies in respect of collection of data on vital events showed that under-reporting increased considerably with increase in recall period.

immediately preceding the date of interview, and such border bias is likely to be more serious near the remoter end-point.

75. It has been planned to collect data with a 2-year reference period and have such additional information on the time of actual occurrence that one can work with any shorter reference period at the tabulation stage. It is felt that although there would still be border errors for any smaller reference period yet it will mainly be of a random nature with the result that the net effect (bias) would be smaller than what it would be if the reference period of collection was identical with reference period of analysis. Conscious efforts made by the investigator and/or informant for the correct placement of a vital event in relation to the (remoter) cut-off point is believed to result in larger biases than what would result from a cut-off at the tabulation stage. It should be pointed out here that the rural population has usually no precise knowledge of the exact date of occurrence of a birth or death.

76. After establishing all these precautions it is necessary to have a self-evaluating system. In this connection we shall take up several points one by one. NSS investigators have acquired some experience in collection of birth and death data in earlier rounds. But it is not known whether they have reached a 'steady' state so that there is no further learning-effect. Evaluation in this regard is necessary. If the learning-effect is still found to be present then caution should be exercised in interpreting the results. For this and other purposes the NSS investigation is spread uniformly over one-year round. To be more specific the entire round has been broken up into six two-months sub-rounds and data have been separately tabulated for each sub-round. Sub-round comparisons will be of help in studying the learning-effect.

77. It is considered desirable for the sake of accuracy to enumerate each and every individual in all the households in the sample villages. This will help us to have not only a correct picture of the number in different age-sex-marital status classes but would also help us to record more correctly all births and deaths occurring to a member of the household canvassed. But as a multi-purpose survey organisation the NSS has to explore the effect of adopting a less time-consuming plan so that resources for other enquiries may be found. It has therefore been decided that in the 14th round only the first two sub-rounds would have a detailed individual by individual enumeration, whereas in the later sub-rounds summary information on a household basis would be collected. A comparison of the results thrown up by the two methods of varying intensity will therefore be of value in assessing the results thrown up by the less intensive enquiry.

78. There is also the basic problem of evaluation of the very approach of estimating birth and death rates as described earlier. A second more direct approach, particularly in regard to deaths, may be helpful in studying the efficacy of this approach. In the second approach an account has been taken of the whereabouts, in the fifth and sixth sub-rounds of the 14th round, of each and every individual

enumerated in the first two sub-rounds,[11] and this resulted in recording deaths if the person in question had died. Cases of migration, both in and out, have been also noted. Additions due to births and cases of omissions in either sub-rounds are also recorded. It may be pointed out that babies born immediately before the date of interview, even if omitted in the first two sub-rounds, (which may not be a very unlikely event), are less likely to be omitted at the re-enumeration stage when they have somewhat grown up (provided they have not died in the intervening period).

79. At the re-enumeration stage an attempt has also been made to collect data about births and deaths during the reference period. It will be seen from what has been said about that the second approach (in which, strictly speaking, the first approach is a part) is expected to be less liable to omission of deaths and to some extent of births as well. But both of them are likely to suffer from under-enumeration of cases of births and deaths of very short-lived infants. However, it may be pointed out that as our principal objective is to determine the rate of increase such cases would not affect (excepting in a marginal way) the results.

80. It must be emphasised, however, that exploring the possibility of estimation of birth and death rates fairly accurately has been also kept in view. It therefore becomes necessary to provide for a method of assessing the magnitude of the recall biases. Tabulation of data by varying lengths of recall periods will throw some light, but in addition a more direct approach appears desirable.

81. The evaluation plan takes into account not a single annual round but two consecutive rounds (the 14th and 15th rounds of NSS). According to this plan the same set of villages are being covered in these rounds. It will be recalled that our reference period is 'last two-years' so that, what is 'last year' in the earlier round becomes 'year before last' for the next round. We shall therefore have two sets of data for the same reference period, but with different recall periods. The 15th round is currently going on and it is hoped that the two sets will provide a means for evaluation of the accuracy of such data collected by the interview method. It will also provide a method of adjustment of data subject to recall lapse.

82. We shall now present some of the preliminary results obtained so far. Table 13 sets down the estimates of (1) average household size, (2) sex-ratio, (3) percentages of the population falling in the three age groups : 0–14, 15–44, 45–, (4) birth-rate, (5) death-rate and (6) rate of natural increase for each of the six sub-rounds of the 14th round (the first round for the 2-round enquiry). There are two parties of investigators; in every stratum one set of six villages is to be covered by an investigator belonging to the first party, and a second set of six by an investigator belonging to the second party. It is thus possible to obtain two valid and independent sets

[11] To ensure careful re-enumeration work, certain fictitious names have been entered in the origianl listing. And a new item of information viz., 'days sick during last month' has to be entered (in the fifth and sixth sub-rounds), so that on the one hand the investigator has to ask pointed questions about an individual earlier enumerated, and this would on the other hand incidentally help him to secure indirectly the delicate information about death which may have taken place.

of estimates. These are also shown. The reference period is 'last year' for births and deaths.

83. It appears on examination of all possible long *versus* short schedule comparisons that the shorter schedule gives rise to larger estimates of average household size. In regard to sex-ratio, there is no obvious difference. For the age groups, particularly the two older groups 15–44 and 45–, there is clear evidence of difference in the two schedule types. Under-estimation of the older groups and over-estimation of the middle group in the short schedule is obvious. For the younger group, 0–14, within party (or within sub-sample) comparisons show that the short schedule gives underestimates; 14 out of 16 possible comparisons support this.

TABLE 13. COMPARISON OVER SUB-ROUNDS, SCHEDULE TYPES AND SUB-SAMPLES OF CERTAIN DEMOGRAPHIC VARIABLES; POPULATION ENQUIRY, 1958-59

Rural sector : All-India

sample size : 218 villages/sub-sample/sub-round

2616 villages during entire round

sub-round	schedule type	sub-sample or party	average household size	sex ratio[a]	percentage in age-group			rates per 1000 persons		
					0–14	15–44	45–	birth	death	growth
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	long	1	5.07	104.12	40.70	42.84	16.46	38.37	18.98	19.39
		2	5.01	103.97	40.54	42.79	16.67	38.56	19.12	19.44
		pooled	5.04	104.05	40.62	42.82	16.56	38.47	19.05	19.42
2	long	1	4.97	101.50	40.90	42.65	16.45	38.76	19.50	19.26
		2	5.11	103.86	41.03	42.77	16.20	40.45	19.89	20.56
		pooled	5.04	102.74	40.97	42.72	16.31	39.66	19.70	19.96
3	short	1	5.16	102.61	40.45	45.16	14.39	37.49	19.82	17.67
		2	5.15	103.84	40.13	45.36	14.51	35.00	18.81	16.19
		pooled	5.16	103.17	40.31	45.25	14.44	36.36	19.36	17.00
4	short	1	5.21	104.94	40.65	45.36	13.99	41.63	21.57	20.06
		2	5.21	105.47	40.83	45.01	14.16	38.65	19.08	19.57
		pooled	5.21	105.20	40.74	45.19	14.07	40.13	20.32	19.81
5	short	1	5.14	103.13	40.62	44.60	14.78	36.33	20.04	16.29
		2	5.15	103.59	40.33	44.70	14.97	37.33	17.24	20.09
		pooled	5.15	103.37	40.46	44.66	14.88	36.85	18.60	18.25
6	short	1	5.28	103.60	40.83	45.13	14.04	38.13	16.28	21.85
		2	5.16	103.61	40.29	45.18	14.53	37.65	17.22	20.43
		pooled	5.22	103.61	40.53	45.16	14.31	37.87	16.79	21.08
1–6	long and short	1	5.14	103.30	40.68	44.31	15.01	38.50	19.47	19.03
		2	5.13	104.07	40.54	44.28	15.18	38.02	18.58	19.44
		pooled	5.14	103.69	40.61	44.29	15.10	38.26	19.02	19.24

[a] Males per 100 females.

84. Similar analysis of birth and death rates does not reveal any marked difference between the two schedule-types excepting possibly for birth-rate which is higher (13 out of 16 possible comparisons) for the long schedule.

85. It is also necessary to study the learning or familiarity effect. The difference, if appreciable, will be revealed by an examination of the differences between sub-rounds other conditions remaining the same. We shall study this in respect of birth rates. Concentrating on the first two sub-rounds we note that both the parties have registered a higher rate at the second sub-round showing perhaps some improvement in enumeration of births. To secure a little more confirmation we have examined this in relation to the five (administrative) Zones into which the country has been divided; 8 out of 10 possible within-party comparisons support this. It will be recalled that from the third sub-round onwards the short schedule has been used. The fourth sub-round again shows a higher birth-rate compared to the third not only for all India but for 9 out of 10 possible within-party zonal comparisons. In case of death-rate, however, there is no such phenomenon. It would appear therefore that there is a learning-effect in regard to births but nothing very much in respect of deaths.

86. The existence of learning-effect brings in the question whether the difference in the schedule types could not be explained by the learning-effect. We would be inclined to feel that the nature and magnitude of the differences are such that it cannot be explained in that way.

87. Preliminary analysis of the re-enumeration data collected in the fifth and sixth sub-rounds show that the annual (the interval between enumeration and reenumeration was roughly 8 months[12] (number of deaths etc., has been proportionately increased to 12 months) death rate among persons enumerated earlier comes out 1.71%. This may be seen against a death rate of 1.77% obtained by the previous approach for the part of the enquiry conducted during the same survey-period (fifth and sixth sub-rounds). It will be noted that inclusion of infants born after the first enumeration and the deaths among them (before re-enumeration) would have made the two figures comparable. As far as one can make out from our estimate of birth rate viz., 3.82%, and the infant death proportion of 99 per 1000 live births (proportion of infants born alive and dying in the reference year) estimated from this enquiry, the comparable death-rate comes to about 1.95%. It would appear therefore that the method of accounting in re-enumeration has resulted in a more complete enumeration of deaths. Further probing analysis is necessary before one can be on surer grounds regarding this evaluation of the results.

88. It is necessary to examine the magnitude of recall-bias. A direct evaluation of this bias will be possible, as explained earlier when the 15th round data (yet to be collected and analysed) are seen against the 14th round data for the same villages and households. We can however study the effect in a less direct way from the 14th round data. It is found that at the all-India level, the 'year before last' birth rate was 81.7% of the 'last year' birth rate. In regard to death rate the corresponding percentage was substantially lower, viz., 52.6%. While the effect of the

[12] An eight month gap is not entirely satisfactory, one year would be preferable, and the data currently being obtained in the 15th round will remove this defect.

recall-lapse is quite appreciable in case of the above two rates it is interesting to estimate the effect of this on the growth-rate, because there is a certain amount of compensation as both are under-estimates and the one (death rate) subject to higher (%) recall-lapse has a lower absolute value. The net effect is over-estimation of growth rate by 10.5% if 'year before last' replaces 'last year'.

89. The above however is a partial appraisal of the situation because the question of recall-lapse within the course of the year has been overlooked. One can, however, reasonably expect from what has been said above that within year recall-lapse effect on the growth-rate cannot be very large. An actual study on the 7th round data showed that a slight increase of about 2 per cent in the growth rate may perhaps be allowed for recall-lapse within the course of the one-year reference period.

90. It would appear therefore that the 14th round NSS estimates of growth-rate is perhaps subject to a slight downward bias. It is however not quite clear at this stage about the effect of a 2-year reference period, introduced for the first time in the 14th round, and how far the conditions obtaining in the 7th round are obtainable in the 14th round. Deeper analysis of 14th and 15th round material may help in clearing up the issues.

VII. Acknowledgements

91. In the preparation of this paper some of our colleagues have helped us with necessary data. We should like to acknowledge our obligation to all of them and particularly to A. Ganguly, R. P. Saha, R. K. Som and S. Raja Rao.

RÉSUMÉ

En développant les techniques pour l'évaluation des résultats d'inspection, l'on considère le fait suivant avec une importance qu'il mérite. Le résultat final dépend non seulement de la "vraie" valeur de l'unité pour l'observation, mais aussi de la méthode d'obtenir de telles renseignements, unité par unité, et aussi de la méthode d'arranger des données.

Le résultat pêut même dépendre de la choix de l'unité de certitude ("ascertainment unit"). La technique fondamentale est donc l'établissement des comparaisons entre deux (ou un plus haut nombre) estimations de la même quantité obtenue sous les conditions opératoires et/ou techniques, partiellement ou entièrement différentes.

L'on pêut évaluer l'inspection contre les données obtenues par une méthode qui est, a priori, plus valable. "Unitary checks" qu'il est possible d'introduire quand les unités de certitude sont identiques, mettent en jour les individuelles erreurs de certitude. La comparaison entre "totaux" évalués—et c'est possible quand même les unités de certitude sont différentes—donne l'erreur totale derivée de tous les sources. L'évaluation des unités choisies aléatoirement nous fournit un moyen de juger de l'effet de l'opportunité amoindrie de compenser les erreurs émanées des sources différents, où il s'agit des estimations de 'totaux' ou quelque semblable sommaire de "renseignements" des unités territoriales (plus petites).

Cette évaluation est importante, car c'est possible que la petitesse de l'erreur totale quand les unités territoriales sont plus grosses, n'indique la même chose pour les petites unités. L'on développe une technique pour séparer la certitude totale ou l'erreur d'énumeration, de l'erreur totale de "compilation-editing," en évaluant un "total" ou l'information sommaire semblable.

Dans le cas d'échantillons entrepénétrants, quand l'on tire deux (ou un plus haut nombre) échantillons de la même population et si l'on les traite par le même dessein d'inspection, les resultats basés sur les différents échantillons sont également valables,—quand même ils sont obtenus par différents rechercheurs et/ou par les différentes unités d'arrangement. Les différentielles opératoires corréspondent à la divergence entre différentes estimations. De plus, l'estimation de l'étendue d incertitude est basée sur cette divergence.

Une technique de la "propre évaluation" (self-evaluation) parait aussi d'être utile. Pour achever ce but il faut accumuler les données sur les unités de certitude ayant différents "dimensions",—les unes étant les parties des autres—par exemple, une petite coupe dans une plus grosse coupe (des expériences de faucher). Les données pour une longue période de référence, accumulées en une telle manière que les réductions en tables pour une plus courte période de référence sont aussi possible, donnent, en effet, des différents dimensions de la période de référence. Une comparaison des estimations que l'on puisse obtenir des différents dimensions facilite l'évaluation.

La comparaison d'éstimation, sur différents pointes ou périodes (de temps) est une autre méthode. Les comparaisons des estimations pour la même période de référence (mais aux différentes périodes d' inspection) sont aussi avantageuses. Faire des recherches sur l'accord dans les propriétés de la divergence entre différentes sections territoriales, en vaut la peine.

Toute les techniques discutées sont nanties des résultats obtenus dans les domains variés—moisson, bétail, propriété et population.