

S. P. Jain

Census Single Year age Returns and Informant Bias*

DISTORTIONS in single year age returns in the censuses have attracted serious attention of actuaries, who produced Indian life tables since 1881. For their study, single year age returns of 100,000 persons were made available. They consistently exhibited heavy concentration at ages having 5 and 0 in the unit's place. The Census Commissioners made attempts to improve the accuracy of data showing population characteristics tabulated by quinquennial age groups. The possible groups are 0-4, 1-5, 2-6, 3-7 and 4-8. The first one has the authority of international usage and has been adopted in India except in 1931, when 2-6 was adopted in accordance with the Actuary's recommendation. It was found to give a smoother progression of group totals than 0-4. The grouping 1-5 suffers from the defect of having the preferred unit digit 5 at the end. The grouping 3-7 deserved consideration from the fact that preferred digit and 5 comes at the midpoint of the group interval, even though it gives a see-saw distribution. The group 4-8 did not receive any serious attention, as the resultant group totals did not show a smoother run than that shown by 2-6. It was Shri R. A. Gopalswamy, who conceived the idea of drawing a 10% sample of 1951 census slips for producing population characteristics classified by age. The precise instructions for selecting the sample slips were as follows :

Break each pad (of 100 slips) and stack the slips of the Pad; and cut the stack as in, a card game. Place the lower portion above the upper portion and deal the slips into the pigeon holes. You should deal the slips with the pigeon holes

The original appeared in Registrar General's Newsletter.

in the order 1,2,3,4,5,6,7,8 and 9 respectively. At the same time, you should watch the slips of displaced persons. If you come across any slip of a displaced person deal it in the pigeon hole of displaced persons. The slips in the pigeon hole marked *S* will make up the 10% sample. Random number table, by which statisticians swear, was not used. This procedure made the results suspect.

Prof. Mahalanobis was able to persuade the Government of India to finance N. S. S. for producing single year age tables for 1941 from what he christened as *y*-sample after the name of Mr. Yeatts, the Census Commissioner for 1941 Census. Shri Asok Mitra preferred to produce complete single year age tables for 1961. The writer, in his capacity as the Census Actuary, studied the data of single age returns for each constituent State and published a report on his findings in a Census Paper. The conclusions reached are as follows:

- (i) Besides concentration at 0s and 5s the data unmistakably show smaller heapings at 2s and 8s, Heaping at 6s was not found to be marked. Five is a major centre of attraction, which counteracts aversion to returning ages in odd numbers. The population count at odd numbers are in serious deficit.
- (ii) Ages 10,12 and 18 in the span 10-19 are major centres of attraction. In some states this fact makes 15 as an age in deficit.
- (iii) Age 5 has a minor peak in contrast to the marked peaks at 25, 30, 35, etc.
- (iv) Age 0 is in marked deficit where as 1 shows a peak. The Census counts at ages in the range 0-7 do not show marked variations of the type observed at subsequent ages. The deficit at age 0 is probably due to the lack of understanding on the part of the informant of the fine distinctions between age *last* birthday, age next birthday, age nearest birthday.
- (v) After age 30, heapings at decennial ages are greater than at quinquennial ages, so much so, that after age 60 the age data show much greater heaping at 60, 70 etc. The result is that age returns after 60 are wholly unreliable except for reflecting 10 yearly group totals.
- (vi) The magnitude of distortions in age returns does not substantially differ between States.

Biases in age returns are by no means absent from the data of advanced countries, where people celebrate their birthday and hence have a very clear idea of their exact age. The distortions noticed there are attributed to aversions and preferences for certain unit digits of age, or a tendency to appear younger or older. In India, most people do not remember their ages. There are very few

occasions, when people have to recall their ages. For entry into service and other civic purposes, whatever comes to their mind is recorded. The recorded age can be changed by an affidavit signed later by a magistrate, which is easy enough. From common sense, it seems that the factor that causes age distortion is estimation and not deliberate bias, as in the West. Enquiries on age by census enumerators not infrequently bring forth the suggestion to record what the enumerator thinks it to be. The poor enumerator has to make his best guess by making the informant recall the major events which he had witnessed in his life time. In his training he is advised to keep handy for reference a schedule of major national events. The writer seems to have been the first to record formally in his report on 1961 Age Tables that distortions in age in India are due to problems of estimation.¹ This was independently stated subsequently by Coale in U. N. Manual IV². The extent of distortion varies from country to country, as will be seen from the index of distortion for³ the different countries mentioned below :

<i>Rank</i>	<i>Index of Distortion</i>	<i>Rank</i>	<i>Index of Distortion</i>
I India (1961 Census)	75.16	V Thailand	5.02
II Iran	47.38	VI Singapore	4.72
III Philippines	24.30	VII Hong Kong	4.64
IV Malaya	18.36	VIII Korea	3.04
		IX Japan	2.74

*Myer's index of preference.

The above ranking tallies with what is stated by Coale in Manual IV. He stresses the importance of distortions in age returns and says that "In case the degree of distortion is mild the estimates of vital rates vary within a narrow range but when it is extreme the range is extensive". The method of estimation which is involved here is based on the Stable Population Theory applied to enumerated age distribution in quinquennial age totals. The practical meaning of Coale's statement is that vital rates so obtained are not precise and are just like other estimates. Indian researchers swear by these estimates and consi-

1. Census of India—Paper No. 2 of 1963—1961 Age Tables.

2. 'Methods of Estimating Basic Demographic Measures from Incomplete Data' U.M., 1067.

3. Jain, S. P. "Smoothing of Incomplete A-age Structure"—A paper contributed to International Seminar organised by the Indo German Foundation.

der those obtained by other methods, like differencing two age ceasal distributions, as less worthy of their attention.

The writer has been keen to pursue the subject of accuracy of Indian age data, Shri Padmanabha, Registrar General, kindly agreed to get the requisite tabulations made from 1971 tape records of Punjab, covering 134615 persons, which form the basic data of this paper. Since the informant is the pivotal person, it may be expected that his relationship to the person enumerated and his educational level play an important role in determining the age return. The informant in the case of census is the head but in cases of practical difficulty in contacting the head, it cannot be ruled out that the enumerator obtains his data from another elderly member of the household or even an outsider. The enumerator obtains his information from the head and not from the person concerned, if he or she is not the informant. As it is, census does not record the name of the informant. The percentage of informants, who are also head of household, as shown by careful sample surveys, is given below :

<i>N. S. S.</i> ⁴	(21st Round)	<i>Percentage</i>
1. Punjab	(a) Rural	85
	(b) Urban	82
2. Uttar pradesh	(a) Rural	88
	(b) Urban	85
<i>A. and V. Survey</i> ⁵	(a) Rural	75

These percentages show that inspite of the care to obtain the data from the head, in the actual practice in 20-25% cases they are collected from persons other than the head.

It is assumed in the census that the head knows the ages of the members of the household, which usually consist of his (or her) spouse, married and unmarried children, brothers and sisters and others relatives as also unrelated persons. Naturally, it is of interest to study whether the index of accuracy, as shown by the census age data varies for these groups. In view of the fact that the mind of an educated person is trained, and that he is likely to appreciate the impor-

4. My thanks are due to Dr. K. C. Seal, Director, Central Statistical Organisation for granting permission to see N, S. S. schedules for the purpose.

5. Ambannvar J. P. and Visaria P. M. "Influence of Literacy and Education on **Quality Of Age Returns**", *Demography-India IV* (0, 1975).

tance of the census better, education level of the informant may be expected to make a difference to the quality of the age data. In fact, the sample survey conducted by Ambannavar and Visaria in the rural areas of Maharashtra covering 10260 persons in two districts shows that it is so. Without being aware that the topic has been studied, the writer made a request to the Registrar General for the necessary tabulations. However, the present paper makes an advance in that accuracy has been studied also by informants' relation to the person enumerated. Further, urban areas also have been covered.

Educational levels of informant have been categorised as (a) illiterates together with mere literates without schooling, and (6) those who have been to school covering those who had passed any of the levels primary to matric and above. Ambannavar and Visaria adopted the categories illiterate and literate irrespective of the fact whether they had been to school or not. To this extent, our results pertain to different categories. The decision to place illiterates and mere literates without having gone to school in the same category is based on the consideration that mere literacy, which is defined in census as ability to read and write a simple letter in any language, should not make any difference to the accuracy of age return. There should be a recognised level of education to make any impact.

The census has adopted for each sex 12 categories of household relationships for tabulation purposes. They can be combined as follows :

- I. Head (male or female)
- II. Head's spouse
- III. Head's father
- IV. Head's mother
- V. Head's male descendants (son, son's son etc.)
- VI. Their wives
- VII. Head's fraternal male relations (brothers, cousins etc.)
- VIII. Their wives
- IX. Head's female relations' husbands
- X. Head's female relations (e.g. daughters, sisters)
- XI. Head's other relation e.g. son's daughter, brothers
- XII. Unrelated persons such as domestic servants.

For our purpose, all the heads' children and grand children, whether male or female, were grouped under category V, for the reasons that these children were born in the family. Census categories IX, X XI and XII were recast into our category IX to represent 'other male members' of household and X 'other

female members' of the household. The relative percentage strength of the tea categories so recast is given below :

	Urban			Rural		
	Illiterate Heads	Educated Heads	Both	Illiterate Heads	Educated Heads	Both
I. Head	14.83	24.81	18.741	6.08	17.31	16.28
II. Head's spouse	15.90	10.65	13.85	14.39	3.67	12.60
III. Head's father	0.57	0.25	0.45	0.65	0.15	0.56
IV. Head's mother	2.44	0.26	1.58	2.34	0.15	1.98
V. Head's children and grand children (Lineal descendants)	30.70	31.12	30.87	29.68	45.06	32.22
VI. Wives of male children and grand children	2.59	3.63	2.99	4.16	3.85	4.11
VII. Head's Male fraternal relations	1.53	2.57	1.93	2.24	3.05	2.38
VIII. Their wives	0.54	0.56	0.55	0.84	0.53	0.79
IX. Other male members	7.16	4.37	6.07	9.25	7.51	8.98
X. Other female members	23.74	21.78	22.97	20.37	18.72	20.10
	100.00	100.00	100.00	100.00	100.00	100.00
Total Persons	19588	12511	32099	85467	17049	102516
No. of Households	2907	3105	6012	13745	2951	16696

Category V has the highest percentage strength and is therefore discussed separately.

Categories I, II, IX and X are the remaining ones, which come next in the numerical strength. Indices of Distortion for each them and all ten categories combined were calculated. Two such indices were calculated: one is Myer's Index, which measures overall inaccuracy in the age span 10-59 and the other, Whipple's Index which shows preference for 0s and 5s in age span 23-62. Higher the index, greater is the distortion. Other groups are too small for meaningful analysis.

The Results

Lineal Descendents. This category forms about 30% of the total population of the households in the sample. The indices of distortion of the age returns of this category in the age span 0-39 were as follows :

	Urban		Rural	
	Whipple's Index	Myer's Index	Whipple's Index	Myer's Index
Illiterate head	134	17.08	134	18.89
Educated head	118	18.70	112	25.04

These indices are distinctly much below those for other categories of relationship to the head which are shown in the next table. Myer's Index for illiterate heads in the urban and rural areas are lower than those for educated heads, but it is the other way as shown by Whipple's Index. However, differences are small. There is no clear rural-urban differential.

Age returns for children below the age 8 are of special interest, because these children were born a few years back and the memory of the head regarding their ages cannot go very much wrong for the reasons already elaborated earlier. In the age span 0-7, the pattern of distortion is much the same as in other age intervals, but the magnitude of distortion is much smaller. The extent of distortion is shown by the deviation of the enumerated population from the expected population at the age. We can take the expected population at an age as the average of the population at the adjoining ages. This is the usual expectation on the assumption that the run of figures is smooth. Nothing to the contrary has been shown by most populations. At no age deviation in terms of percentage of the expected came to more than a deficit of 2.4% or an excess of 9.8%. Age 1 is an exception in that the number of persons counted at this age is in a large deficit of 28.5%. It is at par with that observed at ages having 1 in the unit place in other age groups. For instance, the deficit at age 11 is 35.3%. Age 13 shows a high excess of 102.5%. In the age span 10-18, the deficit at age 14 was found to be 4.1%, which is the lowest in the span.

Myer's Index is more comprehensive. It measures overall inaccuracy taking all digits into account. Whipple's Index measures concentration at digit 0 and 5 only in the age range 23-62. The data given above show the following features.

(i) In the case of illiterates Whipple's Indices for the rural and the urban areas

do not differ substantially, but the difference between illiterate and educated is marked both in the rural and the urban areas. This is true for each relation

INDICES OF DISTORTION IN AGE RETURNS 1971 CENSUS (PUNJAB)

All	Whipple's Index					Myer's Index				All
	Relation to Head					Relation to Head				
	I	II	IX	X		I	II	IX	X	
Rural Areas										
351	387	321	274	312	(i) Illiterate	112.98	110.94	51.49	49.57	56.85
280	301	239	178	150	(ii) Educated	72.75	63.09	54.11	22.21	53.21
Urban Areas										
356	385	331	330	338	(i) Illiterate	85.84	85.12	61.32	34.87	87.35
241	259	220	168	168	(ii) Educated	59.53	47.67	19.03	25.40	44.40

I—Head self; II—Spouse; IX—Other male members; X—Other female members; All—All household members.

category also. It is surprising that the index is the highest for the category of head. He has a greater tendency to return his own age in fives and zeros. The reason seems to be that the head is usually an elderly person about the age of 50 and is ignorant of his own age. His return of the age of other relations including the spouse is less distorted. It is not easy to explain this except by some wild guesses. Another surprising feature is that the return of age of other male members is even less distorted.

(ii) Myer's index shows much the same position. The following further remarks may be made :

- (a) The indices given by Ambannavar and Visaria are 39.0 for rural literate informant, and 49.2 for rural illiterates. They are distinctly lower than the census figures. Punjab indices are a little better than those for all-India, which had the male index 78.1 and female index 80.3 in 1971.
- (b) In both areas, rural and urban, illiterate heads show a higher index than what it is in the case of educated heads for the individual categories and 'All'.
- (c) The indices for categories IX and X are smaller than those for categories I and II.

(4) The indices for spouse of both illiterate and educated heads in the rural and urban areas are a little lower than those for the corresponding 'head* category.

(e) The index for spouse is the highest for illiterate heads in the rural areas. It is the lowest in the case of educated heads in the urban areas.

(iii) There is no clear evidence of rural/urban differential in distortion among illiterates. Among literates, age returns in the urban areas are less distorted than in the case of rural areas.

As a final conclusion, it may be said that the enumerator should request the head of the household to talk with each other about the correct age of each member at the time he goes to the household for preparing a rough inventory for houselisting purposes. This operation is carried out much in advance of the actual enumeration, hence, it is necessary that the enumerator should, before filling the census slips enquire whether in fact such a talk took place. If not, he should request that it may be done before he obtains the necessary particulars. He may request that as many members as can be collected may be associated in the process of arriving at the correct age and in fact other particulars too. Certain events in the household and in the general population can be recounted before the members to help them in recalling their ages so as to ensure that the age recorded is as correct as possible. During census training the census enumerator is advised to keep a calendar of important national events with him. He should make use of this calendar to help the head and the members in giving correct ages.

As education spreads, age returns will improve in quality, but of course, this is a long process. The Directive principles of Indian Constitution providing for compulsory education remain a distant dream. Every education minister is only too conscious of his duty in this regard but somehow his achievement falls much short of his profession.

It is needless to emphasise the importance of conducting an analysis of the type attempted in this paper at every census for each State. Several techniques have been used to graduate the age data but they are hypothetical. A much simpler and convincing method is to correct the quinquennial group totals by the percentage of smoothed age distribution to the unadjusted count as obtained from the data of a neighbouring area.

The subject is discussed further in Part II with the help of the results of analysis made from the individual N.S.S. Schedules. As already indicated, these schedules show whether the informant was the head of the household or some one else. This part is expected to throw light directly on the point whether the

age returns obtained from an educated head are more accurate than **those from** an illiterate head.

PART II

In this part accuracy of age as recorded by the National Sample Survey investigators in the 21st round (1972-73) in Punjab is studied. The households covered are not a random sample. As many households as could be handled by 3 persons copying the particulars in the course of 8 hours of working are included in this analysis. The first three bundles of each of rural and urban schedules that were made available were taken up for transcribing the requisite particulars which were recorded in code numbers. The index of distortion given here is Whipple's index based on all ages and not only 23-62, The numbers of heads falling in the different categories of classification are shown below. They are rather small and hence the conclusion cannot be taken as firm.

Family Religion	Rural Areas		Urban Areas	
	Illiterate head	Educated head	Illiterate head	Educated head
Hindus	116	22	130	152
Sikhs	179	40	44	63
Others	46	1	9	8
	341	63	183	223

Here only such heads as were also informants have been included.

(ii) *Findings.* The following table shows the Whipple's Index.

	Rural Index for		Urban Index for	
	Head	Spouse	Index	Spouse
<i>I NSS Schedule</i>				
Hindu Illiterate	325	200	300	220
Sikh Illiterate	385	215	345	250
Indu Educated	320*	300*	215	200
Sikh Educated	295	240	205	200
<i>// Census Data</i>				
Illiterate Head	387	321	385	331
Educated	301	239	259	220

*Based on small numbers.

It seems that religion does not affect distortion in age returns.

Census figures have been repeated to show comparability. The differentials between illiterate and educated heads shown by N.S.S. data are consistent with those indicated by the Census data excepting that according to census data the index for the spouse's of illiterate heads is markedly higher both in the urban and the rural areas.

It is seen from the NSS schedules that the head gives the age of the spouse lower by 1 or 2 years upto age 25. From 25 to 45, the difference is usually 1,2 or 5 years and above 45 it is usually 5 or 10 years but of course, other figures also occur. If the head returns his age mostly in fives and zeros, the age of the spouse is spread over other digits including odd ones. Thus, there is no universal aversion to give age in odd numbers. The quality of age returns will be greatly improved if age is recorded in years and months. This will give the age last birth day more correctly, particularly of infants and children.

The study, which covers both rural and urban areas, shows that five categories, 'Head* 'Spouse', 'Lineal Desendents,'Other Male Members', and *Other Female Members' exclusive of fraternal relations are numerically important groups in households. Distortions in the age returns of the different groups and all groups combined in case of educated heads are much less than in the case of illiterate heads. It is the highest in the return of the Head himself. The category Spouse follows closely in this regard. A noteworthy pointer is that distortion is distinctly less than the categories of 'all other male' and 'all other female*' members. Rural/Urban differentials are weak in the case of illiterate heads, but they are marked in the case of literate heads. Distortion in the age returns of lineal descendants, who are mostly below 30, is very much less, presumably because they were born in the family within head's knowledge. Differentials by rural/urban or literate/illiterate heads are not well marked.