REPORTING PATTERN OF AGE DATA: A STUDY BASED ON CENSUS DATA FOR NORTH-EASTERN STATES OF INDIA

Jatin Phukan[1], Krishnarjun Bora[2] and Bipin Gogoi [3],*

Department of Statistics, Dibrugarh University ,Assam, India

Corresponding Author: bipin.gogoi@gmail.com[3]

Abstract

Background: Age is an important variable not only for demographers but also for many other branches of social science research, epidemiological and community-based study. The quality of age data distorted due to various type of errors whether consciously or unconsciously. The present study attempts to address the issues regarding the quality of age data reported across the North-Eastern states of India and to find out the potential demographic factors associated with digit preferences or avoidances.

Methodology: The present study used graphical method of evaluating age data and Mayer's blended index which reflects preferences or avoidances for each of the ten digits from 0 to 9.

Results: All North-eastern states of India highly preferred ages ending with digit 0 and 5. Illiterate people of Northeast India, highly prefer terminal digit 0 and 5 respectively and avoid rest of the digits where strongly avoided few terminal digits like 3 and 7 while reporting age data. Even educated people also use to prefer terminal digits 0 and 5 while reporting age. Overall age reporting is batter in Mizoram (5.4) among all of the North-eastern states of India followed by Sikkim (6.7) and Meghalaya and worst age reporting performed by Assam (24.9).

Conclusions: Gender of people, place of residence and educational level of the respondents has found highly potential demographic factors associated with digit preferences or avoidances in age reporting in all the North-eastern states of India.

Key words: Census of India, Digit preference, Mayer's Blended Index, Northeast India.

1. INTRODUCTION

Though conceptually age is a simple concept, the importance of age is well known if it is correctly available. The quality of data collected on age either through population census or survey, is far from satisfactory as various type of error may distort the age data. The first type of error is the coverage error i.e. individuals of a given age may get omitted or missed by the investigator or erroneously included in it (e.g. count twice). For example new born babies are many times forgot to report as a member of the household or may not reported servants living in the house as a member of the household. The second type of error are arise due to failure to record age, where most of the people are not aware about their birth dates or birth dates of other member's living in the same household. The third type of error may arise due to the misreporting of ages either because of ignorance of age or deliberately misreport it. Past researches have (Talib *et al.,* 2001; Unisa *et al.,* 2009; West *et al.,* 2005) shown that there is a tendency to report age into certain preferred age or certain preferred number ending with certain digits 0, 5 or 2. All these types of errors distort the age-sex data hugely.

The studies on age reporting error need special attention since the errors in the age distribution particularly in censuses are examined more intensively than any other information Shryock and Siegel (1973). Therefore, error in age reporting has been a perennial cause of concern for many social scientists including demographers, actuaries, sociologist, economist etc., especially in developing countries like India. Since the socio-economic characteristics vary so much with age, that why population cannot be meaningfully compared with respect to these characteristics unless age has been controlled. Accordingly, age is considered to be the variable of highest priority in demographic analysis Srinivasan (1998).

1.1 Age-sex structure of population of Northeast India

Age- sex distribution of population plays an important role in any demographic analysis. The most important demographic parameters e.g. fertility, mortality, marriage and their rates are derived from Census data considering the age distribution of the population. If major proportion of the population is not reported their ages accurately then it will be very difficult to get the reliable estimates of those parameters (Ambanavar and Visaria, 1975; Mukhopadhyay and Majumdar,

2009; Unisa *et al.,* 2009). Accurate age distribution of the population is also very important for the policy and programme point of views. Therefore a quick review of age- sex distribution of population of Northeast India derived from the national census shown in the Figure 1.
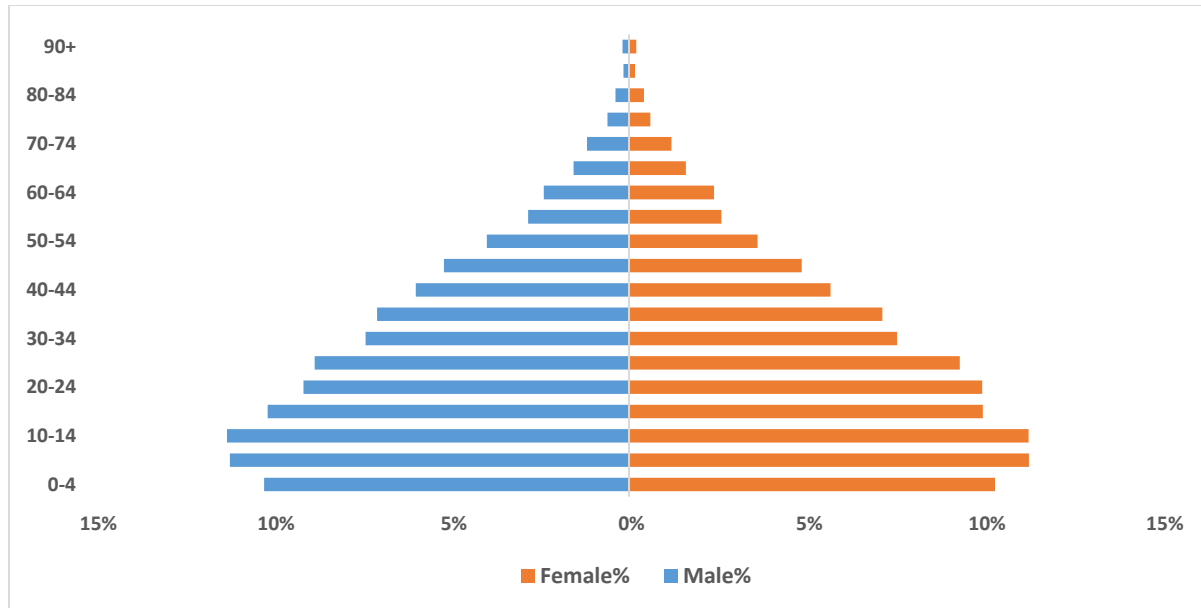

Figure 1: Population Pyramid giving percentage distribution of population in 5 years age group, Northeast India, 2011

The population of Northeast India shown through the age sex pyramid can tell us a lot and can gain many crucial information by examining the shape of it like – (i) Growing young dependent population (0-14), (ii) A huge economically active population (15-64) and (iii) Lower elderly dependent or aging population (65$^+$).

The Northeast India population pyramid almost looks like Expansive shape pyramid with a wide base (larger percentage of people in younger age groups) which indicates indicating high birth rates and high fertility rates) and narrow top which indicates high death rate and lower life expectancies. It suggests a growing population.

The present study attempts to address the issues regarding the quality of age data reported by the North-Eastern states of India. The specific objectives are as follows:

- To analyse the digit preference/avoidances in age reporting across the North-eastern states of India, and

- To analyse the potential factors associated with digit preferences/avoidances in age reporting to all the North-eastern states of India.

2. DATA SOURCE AND METHODOLOGY

2.1. Data source

The Indian Census is the largest single source of a variety of statistical information on different characteristics of the people of India from 1872. It is the most credible source of information on Demographic data like (Population Characteristics) Economic Activity, Literacy and Education, Housing and Household Amenities, Urbanization, Fertility and Mortality, Scheduled Castes and Scheduled Tribes Language, religion, Migration, Disability and many other socio-cultural. In India, Population Census is conducted every ten year which provides valuable information for planning and policy developments for central as well as for the state Government. This study used Census of India, 2011 data and only eight North-eastern states like Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura are considered for the final analysis and excluding the other states plus Union territories. The responsibility of conducting the decennial Census rests with the Office of the Registrar General and Census Commissioner, India under Ministry of Home Affairs, Government of India. The rich diversity of the people of India is truly brought out by the decennial census which has become one of the tools to understand and study the people of India.

2.2 Methodology

The data used in the study are secondary in nature which are taken from the Census of India official website [9]. In this study graphical method and Mayer's index are used to evaluate age data.

Graphical Method of Evaluating Age Data:

When data are in single years, a quick way to draw a line graph plotting age in single years on the X-axis and the percentage of population at various ages on the Y-axis. If the population under

study has not experienced any sharp change in fertility and mortality or sudden changes in migration, then the population age distribution is expected to follow a smooth graph.

Mayer's Blended Index:

To meet the objective the Myer's Index (Myers, 1940) is calculated for the overall population as well as male and female, rural and urban, literate and illiterate separately. The Myer's index is used when data are in single year. The computation of Myer's Index is as below.

Myer's index reflects preferences or avoidances for each of the ten digits from 0 to 9. The method is based on the assumption that the number of person by age varies linearly. This method is applied to the populations in the age range 10 to 99 assuming that the tendency of preference or avoidance in reporting ages ending in certain digits negligible below 10 years and above 99 years ages. The steps in the calculation of Myer's blended index are as follows:

1. Sum of population ending in each digit over the whole range starting with the lower limit of the range (e.g. 10, 20, 30, 40…80; 11,21,31,…,89).

2. Ascertain sum excluding the first population combined in step 1 (e.g. 20,30, 40,…,90; 21, 31, 41,…,99).

3. Weight the sums in steps 1 and 2 and add the results to obtain a blended population (e.g., weight 0.1, and 0.9 for 0 digit, weights 0.2 and 0.8 for 1, etc.)

4. Convert distribution in step 3 into percentages.

5. Take the deviation of each percentage in step 4 from 10.0 which is the expected value for each percentage.

6. A summary index of preference for all terminal digits is derived as one half of the sum of the deviation from 10.0%, each without regard to signs.

The Mayer's Index varies from 0 to 180 and for individual digit, it varies between 0 to 90. The smaller the index, the higher the accuracy of age reporting. The deviation of 10 percent indicates preference or avoidance of age ending on a particular digit.

The complete analysis was done by using Microsoft Excel software (version-2016).

3. RESULTS

Census of India provide age data in single years, therefore age data can be evaluate by graphical method also. Figure 2 shows that the line graph constructed by combining the age data of North-Eastern states to study the overall distortion in age data. From the figure, it can observed that the age distribution is not smooth completely, especially peaks at ages ending with certain digits like 0 and 5 which reflects preference for ages ending with digits 0 and 5. From this figure, it is also evident that digit preference in age reporting clearly exist among adult age group (20-70 years). Thus, it can be conclude that in the middle age range the most significant error in age reporting due to digit preference and it hampers age distribution to a great extent.

Since Myer's index measure the preferences and avoidances for each of the ten digits from 0 to 9. From Figure 3, it is clear that all the states of North-East India highly preferred digit 0 and 5. There exist a little preference for digit 8 except for the states of Assam and Tripura. Digit 0 is highly preferred by Assam followed by Arunachal Pradesh, Manipur and Tripura and less preferred by Mizoram and Sikkim. Similarly digit 5 is highly preferred by Assam, Arunachal Pradesh and Tripura and less preferred by Mizoram and Sikkim. Rest of the digits like 1, 2, 3, 4, 6, 7 and 9 are mostly avoided by all states of Northeast India. Table 1 presents the values Myer's index for all the states together with overall North-East and India's index. From table No. 1 it is confirmed that on an average the reporting pattern of age data is far satisfactory in North-Eastern region of India (Myer's Index = 12.9) in compare to national average (Myer's Index = 26.0).
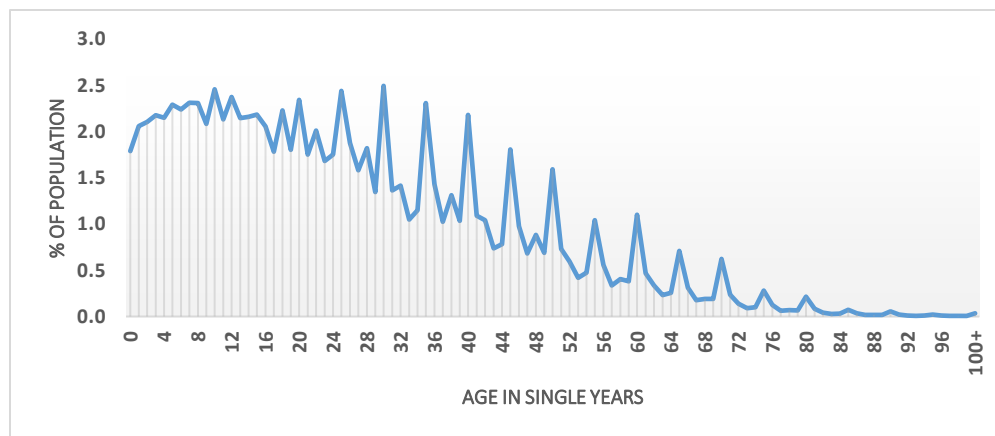

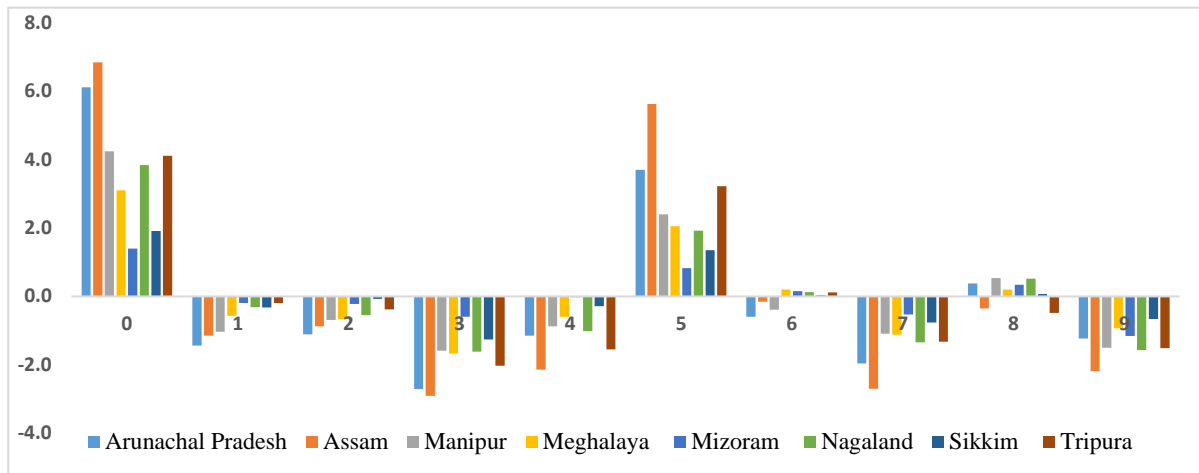**Figure 2:** Age reported by single years in Northeast India, 2011

Figure 3: Bar diagram shows the preferred and avoided terminal digits for Northeaster states of India based on Census 2011 data.

Overall age reporting is batter in Mizoram (5.4) among all of the North-eastern states of India followed by Sikkim (6.7) and Meghalaya and worst age reporting performed by Assam (24.9).

Table 2 presents the potential demographic factors associated with digit preferences and avoidances in age reporting for all the North-eastern states of India. In compared to male population female has highly prefer to certain ages ending with digits 0 and 5 in age reporting across all the state of Northeast India. But in Mizoram, this preference gap between male and female are almost negligible i.e. male and female both are equally prefer digit 0 and 5 while reporting age data and this difference is highest in Tripura. Illiterate people of Northeast India, highly prefer terminal digit 0and 5 respectively and avoid rest of the digits where strongly avoided few terminal digits like 3 and 7 respectively while reporting age data. Even educated people also use to prefer terminal digits 0 and 5 while reporting age. In Assam, digit preference gap between literate and illiterate people is highest compared to other North-eastern states where this gape is comparatively better (low) in Meghalaya. The tendency of digit preference is high in rural areas than urban areas in age reporting except in Sikkim. In Sikkim urban people highly prefer digits in age reporting in competition to rural people. Figure 4 presented the map of North-East India

showing the level of digit preference and literacy rate. From this map, it is evident that error in age reporting is comparatively less in states with higher literacy than the states with low literacy rate.
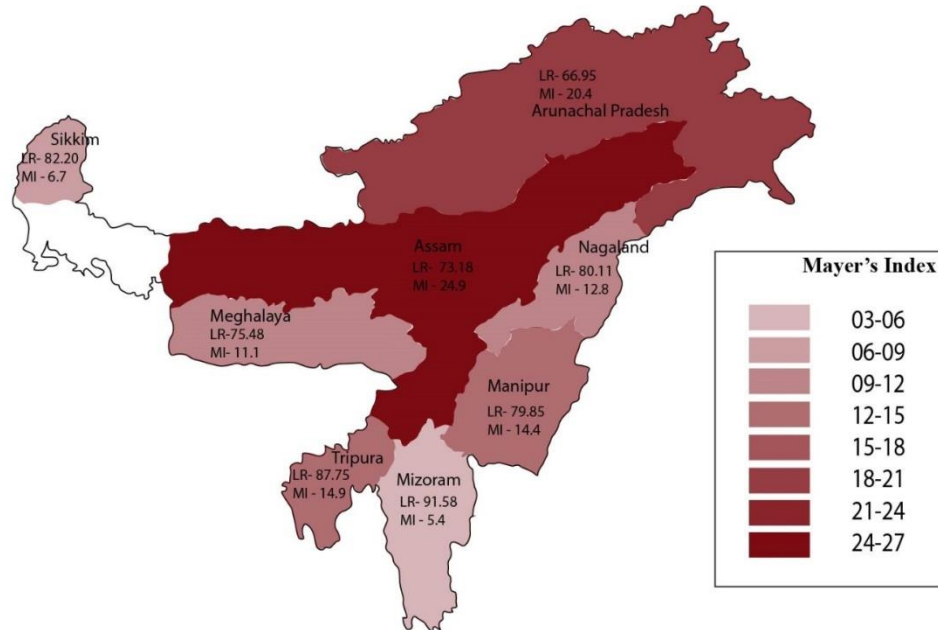
Figure 4: Mayer's Index (MI) value along with Literacy Rate (LR) for

the North Eastern states of India, 2011 Census

(Data source: censusindia.gov.in)

Table 1: Mayer's Blended index shows the preference and avoidance of terminal digits for overall population of North-eastern states of India, census 2011.

States	Terminal Digit										Mayer's Index
	0	1	2	3	4	5	6	7	8	9	
Arunachal Pradesh	6.1	-1.4	-1.1	-2.7	-1.2	3.7	-0.6	-2.0	0.4	-1.2	20.4
Assam	6.9	-1.2	-0.9	-2.9	-2.1	5.6	-0.2	-2.7	-0.4	-2.2	24.9
Manipur	4.2	-1.0	-0.7	-1.6	-0.9	2.4	-0.4	-1.1	0.5	-1.5	14.4
Meghalaya	3.1	-0.6	-0.7	-1.7	-0.6	2.1	0.2	-1.1	0.2	-0.9	11.1
Mizoram	1.4	-0.2	-0.2	-0.6	0.0	0.8	0.2	-0.5	0.3	-1.2	5.4
Nagaland	3.8	-0.3	-0.6	-1.6	-1.0	1.9	0.1	-1.3	0.5	-1.6	12.8
Sikkim	1.9	-0.3	-0.1	-1.3	-0.3	1.4	0.0	-0.8	0.1	-0.7	6.7
Tripura	4.1	-0.2	-0.4	-2.0	-1.6	3.2	0.1	-1.3	-0.5	-1.5	14.9
Northeast India	3.9	-0.5	-0.5	-1.8	-1.1	2.6	0.0	-1.2	0.0	-1.4	21.0
India	7.2	-1.9	-0.6	-2.6	-1.9	5.4	-0.8	-2.7	0.4	-2.4	26.0

Table 2: To analyse the potential factors associated with digit preferences/avoidances in age reporting for North-eastern states of India, Census 2011.

States	Terminal digits →	0	1	2	3	4	5	6	7	8	9	Mayer's Index
Arunachal Pradesh	Male	6.0	-1.3	-1.0	-2.7	-1.2	3.7	-0.6	-1.9	0.2	-1.3	19.8
	Female	6.2	-1.6	-1.2	-2.8	-1.1	3.7	-0.6	-2.0	0.5	-1.1	21.0
	Literate	3.7	-1.4	-0.7	-2.0	-0.6	2.8	0.01	-1.3	0.8	-1.4	14.7
	Illiterate	10.1	-1.5	-1.9	-4.0	-2.1	5.2	-1.6	-3.0	-0.3	-0.9	30.7
	Rural	6.7	-1.5	-1.2	-2.8	-1.2	3.8	-0.8	-2.1	0.3	-1.2	21.7
	Urban	4.2	-1.4	-0.9	-2.3	-0.9	3.5	-0.1	-1.4	0.5	-1.3	16.4
Assam	Male	6.3	-1.1	-0.7	-2.8	-2.1	5.3	-0.04	-2.4	-0.1	-2.2	23.2
	Female	7.4	-1.2	-1.0	-3.1	-2.2	6.0	-0.3	-3.0	-0.6	-2.2	26.8
	Literate	4.7	-1.3	-0.3	-2.2	-1.5	4.3	0.0	-2.1	0.2	-1.9	18.5
	Illiterate	11.6	-0.9	-2.1	-4.6	-3.5	8.6	-0.4	-4.2	-1.6	-2.9	40.4
	Rural	7.2	-1.2	-1.0	-3.0	-2.3	5.9	-0.1	-2.8	-0.4	-2.3	26.2
	Urban	5.1	-1.0	-0.5	-2.3	-1.5	4.2	-0.4	-2.0	-0.1	-1.5	18.5
Manipur	Male	4.1	-0.8	-0.6	-1.5	-0.9	2.3	-0.3	-1.0	0.4	-1.5	13.6
	Female	4.4	-1.2	-0.7	-1.6	-0.8	2.5	-0.5	-1.2	0.7	-1.5	15.2
	Literate	3.2	-1.0	-0.5	-1.2	-0.6	2.0	-0.2	-0.8	0.7	-1.5	11.8
	Illiterate	7.6	-1.2	-1.3	-2.7	-1.6	3.6	-1.0	-1.9	0.1	-1.6	22.6
	Rural	4.5	-1.0	-0.8	-1.7	-1.0	2.4	-0.4	-1.2	0.7	-1.5	15.13
	Urban	3.5	-1.1	-0.5	-1.3	-0.6	2.5	-0.3	-0.9	0.3	-1.6	12.6
Meghalaya	Male	3.2	-0.5	-0.7	-1.8	-0.7	2.1	0.3	-1.1	0.2	-0.9	11.5
	Female	3.0	-0.6	-0.6	-1.6	-0.5	2.0	0.2	-1.2	0.3	-0.9	10.8
	Literate	2.1	-0.6	-0.4	-1.3	-0.3	1.6	0.4	-0.8	0.4	-1.0	9.1
	Illiterate	5.8	-0.5	-1.3	-2.8	-1.3	3.3	-0.3	-1.9	-0.4	-0.7	18.1
	Rural	3.4	-0.7	-0.7	-1.8	-0.6	2.3	0.1	-1.2	0.1	-1.0	11.9
	Urban	2.0	-0.2	-0.5	-1.3	-0.6	1.3	0.4	-0.8	0.5	-0.8	8.5
Mizoram	Male	1.4	-0.2	-0.3	-0.6	-0.1	0.9	0.1	-0.5	0.3	-1.0	5.4
	Female	1.4	-0.2	-0.2	-0.6	0.1	0.8	0.2	-0.6	0.4	-1.3	5.6
	Literate	1.1	-0.2	-0.1	-0.5	0.1	0.7	0.2	-0.4	0.4	-1.2	4.9
	Illiterate	5.1	0.1	-0.9	-1.8	-1.0	2.5	-0.5	-1.7	-0.4	-1.3	15.4
	Rural	1.9	-0.2	-0.3	-0.7	-0.1	1.1	0.1	-0.7	0.2	-1.2	6.4
	Urban	1.0	-0.2	-0.1	-0.5	0.1	0.6	0.2	-0.4	0.5	-1.1	4.8
Nagaland	Male	3.8	-0.2	-0.5	-1.6	-1.0	1.8	0.1	-1.3	0.4	-1.5	12.2
	Female	3.9	-0.5	-0.6	-1.7	-1.0	2.0	0.1	-1.3	0.7	-1.7	13.5
	Literate	2.8	-0.3	-0.4	-1.3	-0.7	1.6	0.3	-1.1	0.6	-1.5	10.7
	Illiterate	7.8	-0.2	-1.1	-2.7	-2.3	3.3	-0.7	-2.4	0.1	-1.7	22.3
	Rural	4.5	-0.3	-0.6	-1.8	-1.3	1.9	0.02	-1.5	0.6	-1.6	14.1
	Urban	2.3	-0.3	-0.4	-1.3	-0.5	1.9	0.4	-0.9	0.3	-1.5	9.8
Sikkim	Male	1.8	-0.2	0.1	-1.2	-0.2	1.3	-0.1	-0.7	0.1	-0.7	6.5
	Female	2.0	-0.4	-0.2	-1.3	-0.3	1.4	0.2	-0.8	0.1	-0.6	7.3
	Literate	1.1	-0.2	0.1	-1.0	-0.2	0.9	0.1	-0.5	0.3	-0.7	5.1
	Illiterate	5.1	-0.8	-0.8	-2.3	-0.8	3.0	-0.4	-1.7	-0.9	-0.4	16.2
	Rural	2.0	-0.3	-0.1	-1.2	-0.2	1.3	-0.005	-0.8	0.0	-0.6	6.5
	Urban	1.6	-0.4	0.1	-1.4	-0.4	1.6	0.1	-0.8	0.4	-0.8	7.5

States	Terminal digits →	0	1	2	3	4	5	6	7	8	9	Mayer's Index
Tripura	Male	3.6	-0.2	-0.2	-1.8	-1.4	2.8	0.1	-1.0	-0.4	-1.5	13.1
	Female	4.7	-0.2	-0.5	-2.3	-1.7	3.7	0.1	-1.6	-0.6	-1.5	16.8
	Literate	3.2	-0.2	-0.2	-1.7	-1.3	2.6	0.2	-1.1	-0.3	-1.3	12.9
	Illiterate	9.4	0.1	-1.4	-4.2	-3.3	6.8	-0.1	-2.9	-1.8	-2.6	32.7
	Rural	4.5	-0.2	-0.4	-2.2	-1.6	3.5	0.1	-1.5	-0.5	-1.7	16.4
	Urban	3.1	-0.1	-0.2	-1.6	-1.3	2.4	0.1	-0.8	-0.4	-1.1	11.2
Northeast India	Male	5.5	-0.9	-0.7	-2.4	-1.7	4.3	-0.03	-2.0	-0.1	-1.9	19.6
	Female	6.3	-1.0	-0.9	-2.7	-1.8	4.9	-0.2	-2.5	-0.3	-1.9	22.4
	Literate	4.0	-1.0	-0.3	-1.9	-1.2	3.5	0.02	-1.7	0.2	-1.7	15.6
	Illiterate	10.7	-0.9	-1.9	-4.2	-3.2	7.5	-0.5	-3.7	-1.3	-2.5	36.4
	Rural	6.4	-1.0	-0.9	-2.7	-1.9	5.0	-0.1	-2.4	-0.3	-2.0	22.7
	Urban	3.9	-0.8	-0.4	-1.8	-1.1	3.2	-0.2	-1.4	0.1	-1.4	14.3

4. DISCUSSION AND CONCLUSION

Generally, the quality of age reporting in a country reflected the overall level of educational attainment. In addition to education, there may be other factors which might influence the quality of age data in India. It is fact that the mind of an educated person is trained, and that he is likely to appreciate the importance of the age data better. In India, majority of the population is illiterate and therefore the age data from census suffer from a number of problems such as ignorance of age, negligence in reckoning the correct age, deliberately misreported, and misunderstanding of the question [1, 4]. Also, among illiterates it is very difficult to get the exact age of the population because of the ignorance of the respondents. But while talking about the errors at specified age we have to consider all kinds of errors whether coverage or content error simultaneously [6].

Overall, in case of age reporting the North-eastern region of India (Mayer's Index=21.0) performed better than the national average (26.0) to some extent. There also interstate difference exists. The people of the North-eastern region highly preferred digit 0 and 5, a little digit 8 also. Rest of the digits like 1, 2, 3, 4, 6, 7, 8 and 9 are mostly avoided by all states of Northeast India while reporting age data. Overall age reporting is better in Mizoram (5.4) among all of the North-eastern states of India followed by Sikkim (6.7) and Meghalaya and worst age reporting performed by Assam (24.9). Sex of the respondent, place of residence and educational status of people has found highly potential factors associated with digit preferences/avoidances in age reporting to all the North-eastern states of India. To same extent, in Assam, the illiterate people were highly

preferring digit 0 and 5. Among the three potential demographic factors educational level of respondents has played strongest role in digit preference or avoidance while reported age data.

REFERENCES

[1] Ambanavar, J. P. and Visaria, P. (1975). Influence of literacy and education on the quality of age returns. Demography India, 4(1), 11–15.

[2] Mukhopadhyay, B. K. and Majumdar, P. K. (2009). A Multivariate Statistical Analysis of Reporting Error in Age Data of India. *Journal of Social Sciences*. 19(1), 57–61.

[3] Shryock, H. S. and Siegel, J. S. (1973). The Methods and Materials of Demography, Vol. No. 1, USA: U.S. Department of Commerce.

[4] Siegel, J. S. and Swanson, D. A. (2004). The methods and materials of demography. 2nd edition, Academic Press, 125–73.

[5] Srinivasan, K. (1998). Basic Demographic Techniques and Applications, Sage Publications, New Delhi.

[6] Talib, A. L., Ali, M. S., Hamid, M. S. and Zin, K. M. (2001). Age Reporting Behaviour: A Case Study of 1991 and 2000 Population and Housing Censuses, *Malaysia. J Dep stat.*, 1, 61–84.

[7] Unisa, S., Dwivedi, L. K., Reshmi, R. S. and Kumar, K. (2009). Age reporting in Indian census: An insight. 26th IUSSP International Population Conference. Morocco, 27 Sept-Oct.

[8] West, K. K., Robinson, J. G. and Bentley, M. (2005). Did Proxy Respondents Cause Age Heaping in the Census 2000? ASA Section on Survey Research Methods, 3658–3665.

[9] censusindia.gov.in