www.gi.sanu.ac.rs, www.doiserbia.nb.rs,
J. Geogr. Inst. Cvijic. 68(3) (345-361)

Original scientific paper UDC: 911.3:312(540)
DOI: https://doi.org/10.2298/IJGI180218006A

QUALITY OF AGE STATISTICS IN INDIA: AN INSIGHT OF CHANGING COURSE OF RELIABILITY

Rabiul Ansary[1*], *Mohammad Arif*[2]
[1] Jawaharlal Nehru University, School of Social Sciences, Centre for the Study of Regional Development, New Delhi, India; email: rabiulansary.ansary786@gmail.com
[2] Visva-Bharati University, Institute of Social Sciences, Department of Geography, Santiniketan, India; email: arifaligs@gmail.com

Received: February 18, 2018; Reviewed: September 4, 2018; Accepted: September 12, 2018

Abstract: The second phase of census operation in India — population enumeration — collects data on individual's characteristics from every household. Out of 30 questions that are asked the pivotal query is that of the age and the sex. Both these data have certain crucial demographic, an economic and social angle which helps to build future policies and alleviating any concern. However, the available age information procured through census operation shows age heaping around the digits of "0" and "5", which seems to be declining with age. Besides, the data also have a gendered perspective on the question of age heaping. Owing to such misleading information future policy prescriptions stand questioned. This descriptive study is a step forward towards resolving the lacunae by estimating the magnitude of age heaping in every state of India, secured with the help of Whipple's Index. In addition, the study relates the accuracy of age reporting to characteristics of literacy, urban population, and level of birth registration. And concludes that with raising the level of these predictors reporting of correct age can be effectively secured.

Keywords: age error; biases; Whipple's Index; literacy rate; birth registration

Introduction

The accurate information of basic demographic characteristics (e.g. age, birth interval, duration of the marriage, age at marriage, working status and income of the households) are inherent with varieties of discrepancies in the developing countries. It is due to *errors,* i.e. *unconscious misreporting of information* or *Biases,* or so-called *purposive misreporting of information* or the combination of both. An accurate age-sex data is the most important demographic variable for the policy recommendation, implication and market economy. The age facilitates in estimating the changing pattern of age-sex structure, population projection and calculating population growth, mortality, and morbidity. It is further assessed for age-sex specific health conditions, aging of population and economics of population. Similarly, we cannot deny the role of marriage market,

[*] Correspondence to: rabiulansary.ansary786@gmail.com

which highly shapes the attitude of the people to report analogous ages in particular society.

The first population census was conducted in Indian provinces in 1872. But the major concern of accurate age distribution data in India goes back to 1991 census, when tables on age were generated on sample tabulation. For the first time in 2011 census, data on both dates of birth and age has been recorded. A direct question on age was asked from 2001 census and the tables on age were published on the full count basis. Being a developing country, the quality of age reporting is flooded with a number of discrepancies in India.

Literature Review

While it is noted above that age misreporting have several negative consequences, the heaping is highest for “0” and “5” digits, with 5 coming next to 0 in a sequence of preference (Stockwell, 1966). Reporting of age is very susceptible to errors with both the nature and quality of data varying across space and time (Moultrie, Sayi, & Timæus, 2012; González, Attanasio, & Trang Ha, 2014). The survey of existing literature clearly shows that in developed countries misreporting and biases are approximately negligible while it is intense in developing countries, for instance, quite emphatically, Siegel and Swanson (2004) found that there was no digit preference in the 1991 census of the USA. Whereas, number of attempts to evaluate age statistics of developing countries have revealed enormous distortions (Caldwell, 1966; Caldwell, & Igun, 1971; Carrier, & Hobcraft, 1971; Byerlee, & Terera, 1981; Ewbank, 1981).

Some irregularities in age data from African and Asian demographic surveys have been noted by previous studies (Caldwell, 1966; Caldwell, & Igun, 1971; Nagi, Stockwell, & Snavley, 1973; Byerlee & Terera, 1981; Ewbank, 1981; Jowett, & Li, 1992; Denic, Khatib, & Saadi, 2004; Palamuleni, 2012). But off late, the quality of census data with regard to age reporting has improved pronouncedly in Asia, but lagging behind in African countries (Cleveland, 1996). The study based on census of selected countries of Arab region by United Nations Statistics Division (2013) indicated that although conducting census dates back to mid-nineteenth century, in this region still age misreporting continues to be a problem in most of the studied countries, which affects derivations of population characteristics of other age groups. Similarly, the pushing of particular age into another age group leads to gaining or losing of characteristics of that age group.

Very few studies in India are conducted so far to evaluate the age statistics provided by the Censuses of India. The micro level study of Pardeshi (2010) in the Yavatmal District (the Indian state of Maharashtra) found very poor quality of age data and age heaping at ages with terminal digits “0” and “5”.When viewed structurally, a number of studies have reasoned on the following factors — level of education of the respondents, nature of birth registration, culture, caste and religion (Mukherjee, & Mukhopadhyay, 1988). The study of Scott and Sabagh (1970) has indicated knowledge of one’s age a culturally controlled phenomenon. The heaping of ages with terminal digits “0” or “5” or at other digits is also common due to cultural preference or avoidance of certain digits (Nagi, Stockwell, & Snavley, 1973). In some cultures, certain numbers may be specifically avoided e.g., 13 in the West and 4 in East Asia (Siegel, & Swanson, 2004).

The present study attempts to address the issues regarding the quality of age data reported across the states in India over the decades. The specific objectives are as follows:

– to show the state level trend and pattern of digit preference in India in the last three census records — 1991, 2001 and 2011;
– to analyse the gender differential of age preference across the states of India; and
– to analyse the factors associated with digit preferences/biases in age reporting.

Data Bases and Methodology

The investigation is carried out both at the national and state level based on the census data covering the period from 1991 to 2011 census. The information of single year and five-year age-sex group data is taken from socio-cultural tables (Series-C) and Whipple’s Index at the national and states level is calculated. The data on literacy rate and urban population is collected from Primary Census Abstract 2011, and the level of birth registration from the Vital statistics of India based on the Civil Registration System of 2012 for understanding the determinants of quality age data.

Apart from the calculation of Whipple’s Index and its illustration with the appropriate chart the results of the gender differences in age reporting are presented with the help of a suitable cartographic technique. The age distribution data for Uttarakhand, Jharkhand, and Chhattisgarh are combined with their mother states, Uttar Pradesh, Bihar and Madhya Pradesh, respectively for the year 2001 and 2011 for comparing with that of 1991. Similarly, since the census operation was not conducted in Jammu and Kashmir in 1991, the corresponding data for Jammu and Kashmir are deducted from the all India figure during the computation of 2001 and 2011 statistics. The study of Union Territories is excluded in the study. The range of age groups for the present study is 23 to 62 for two pragmatic reasons. The study excluded the old age and childhood because they are more strongly affected by other types of errors of reporting than by preference for specific terminal digits. Whipple's Index is prepared by the sum of the population reported age with terminal digits (here "0" and "5") in the age groups 23 to 62 and divided by the total population into the age groups 23 to 62. Then the result is multiplied by 100. The index value lies between 500 indicating that only terminal digits "0" and "5" were reported and 100 representing no preference for "0" or "5," during age reporting. The analysis of census data of India since 1991 to 2011 census revealed some interesting facts with terminal digits "0" and "5". Whipple's Index for ascertaining age heaping around "0" and "5" terminal digits.

$$WI = \frac{\sum (P_{25} + P_{30} + P_{35} + P_{40} + ... + P_{60})}{0.2\sum (P_{23} + P_{24} + P_{25} + ... + P_{62})} 100 \tag{1}$$

In addition, factors that drive any anomaly in age reporting are analysed for finding any significant association of them with the index values. The predictors used are — literacy rate, percentage of urban population, and level of birth registration. And these associations are realised with the help of a multiple linear regression model. The formula may be written as:

$$Y = \alpha + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + U \tag{2}$$

Where Y is the Whipple's Index, β indicates regression coefficients, X indicates predictors, α is the intercept and U indicates the error (or unexplained part, residual). The error term is assumed to have constant variance.

It is emphatically assumed that data quality in the developing economies is poor relative to those generated in the developed realm. As a result, 5% difference (WI = 105) from an error-free age reporting (WI = 100) is taken as a considerable error in a developed country's data but such an excess and much more than that is within a tolerable limit in a developing county's data.

Several of earlier studies have classified the Whipple's Index values into a number of hierarchical classes. The United Nation recommends a standard for measuring the age heaping using Whipple's Index which is used in the present study.

Table 1: Whipple's Index score

Whipple's Index	Quality of Data	Deviation from Perfect
< 105	Highly accurate	< 5%
105–110	Fairly accurate	5–9.99%
110–125	Approximate	10–24.99%
125–175	Rough	25–74.99%
> 175	Very Rough	≥ 75%

Source: United Nations Workshop on Census Data Evaluation (Using Whipple's Index) in Kampala (Uganda), 2012

Results and Discussion of Digit Preferences

Age heaping occurs when records show age with preponderance or absence of certain terminal digits in large-scale sample survey or census. The available age data provided by the censuses of India (1991; 2001; 2011) over the last three decades indicate declining trend of digit preference, yet the data quality of age reporting is "rough" (*WI*-171) in 2011 census. The Figure 1 shows the general trend of digit preference at the time of age reporting with terminal digits "0" and "5" at the time of the census enumeration in India for 1991 and 2011 censuses. The 1991 census reported very higher percentage of the population in the age group 25–50 preferring age ending with either zero or five.

The early ages i.e. childhood and adolescent (age group 0–20) show an oscillating graph in 1991 census, while the same graph gets smoothened in 2011 (Srinivasan & Shastri, 2001). It indicates the inconsistent pattern of misreporting of age with terminal digit zero and five in recent times in early age group bracket. From the 25-year age onward till age 50 registered a very high digit preference both in the 1991 and 2011 censuses, but it does show a steady decline in 2011 census (Mukhopadhyay, 1983). Still, in 1991 and 2011 censuses, the age group 25 to 50 formed a high plateau. And from age groups, 55-year age onwards propensity of age heaping declined. The census year 2011 came up with more or less fairly accurate age data of the population as compared to the previous censuses probably because for the first time in 2001 census question on date of birth along with age was asked. Although the population in the age group 25–50 still reported more preference with the terminal digit zero or five but registered huge decline as compared to the 1991 census. Thus, the height of plateau (gap between 1991 and 2011 censuses) of the same age groups also declined over the census period with more or less smooth curvature in 2011 census. The childhood and adolescent age population also indicated more or less smoothing of age reporting as compared to the 1991 census. The smoothening of early or childhood age is the reflecting of the effects of compulsory birth registration since the 1990s. The birth certificate becomes mandatory at the time of school admission or other important administrative works.

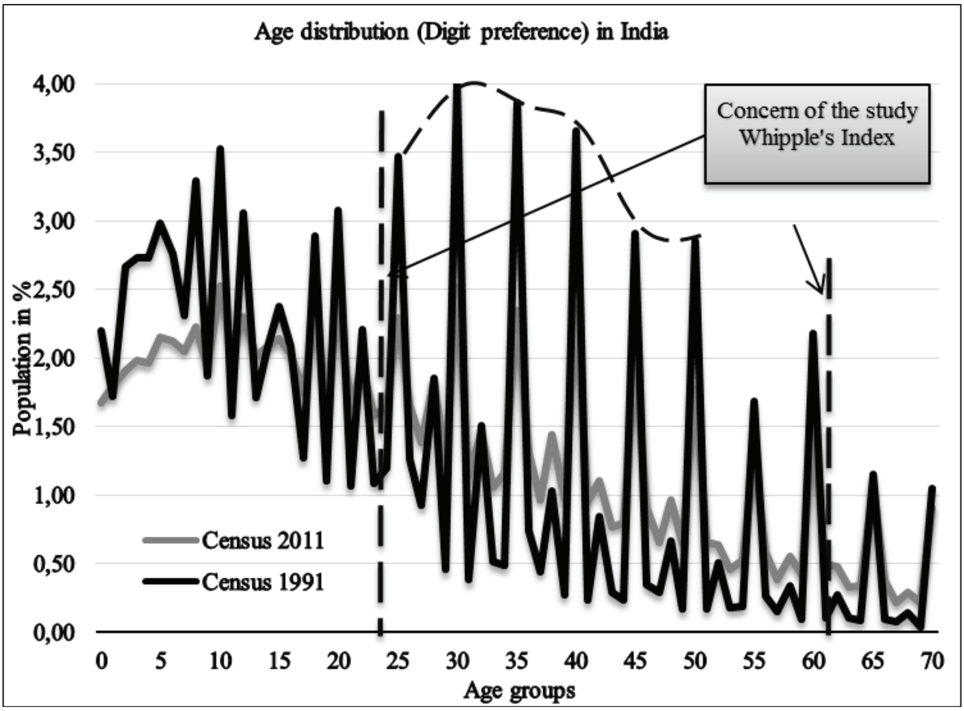

Figure 1: Trend and pattern of digit preferences over the censuses in India (Source: Census of India, 1991; 2011)

The major thrust of this study is to assess the data quality collected by the Census of India over the three census decades using the Whipple's Index (methodology discussed in the earlier section of the paper), which will be explained in detail in this section. Although the history of conducting a census in India goes back to the 19th century, still information of age distribution is very rough. The age distribution of the 1991 census indicated "very rough" age data (*WI* value is 290) — an indication of the very high preference of age with the terminal digit zero or five. In 2001 census, the index value fell to 230 with 60 points decline over the intercensal period but nevertheless, the data quality still stands as very rough. Further in 2011 census, data quality improved with index value 171. At the state level, not a single state reported accurate age data in 2011 census. But it definitely showed the sign of improvement in its quality than the previous censuses. The state-level analysis of the age distribution data in the last three consecutive censuses (1991–2011) shows a declining value of Whipple's

Index (Mukherjee, 1976). In 1991 census, all the states reported a very rough data quality except the state of Kerala. Kerala instead reported a "rough quality" of age distribution data with WI index value — 168. The highest index value was recorded in the state of Bihar, followed by Rajasthan (Table 3). The seven major states which accounted for 45 percent of India population reported index value higher than the national average in 1991 (Unisa, Dwivedi, Reshmi, & Kumar, 2015). The 2001 census showed sign of improvement of data quality. The all India index value dropped to 230 in 2001 from 290 in 1991.

Table 2: Whipple's Index value classification

Symbol	Index value
No states of India	< 105 (Highly accurate)
No states of India	105–110 (Fairly accurate)
	110–125 (Approximate)
	125–175 (Rough)
	More than 175 (Very Rough)
	National Average

Source: Based on United Nations Workshop on Census Data evaluation (Using Whipple's Index) in Kampala (Uganda), 2012

Table 3: Whipple's Index score across the states of India

Sl. No.	States	1991	States	2001	States	2011
1	Kerala	168.91	Kerala	147.59	Mizoram	112.07
2	Mizoram	176.20	Mizoram	155.40	Kerala	119.39
3	Sikkim	176.48	Sikkim	157.19	Sikkim	120.02
4	Manipur	215.78	Haryana	175.35	Goa	126.29
5	Himachal Pradesh	220.33	Himachal Pradesh	179.88	Himachal Pradesh	130.08
6	Nagaland	221.15	Nagaland	189.62	Meghalaya	130.61
7	Goa	225.21	Manipur	193.50	Nagaland	134.30
8	Meghalaya	242.67	Goa	196.95	Manipur	136.53
9	West Bengal	254.49	Gujarat	203.82	Tripura	138.87
10	Gujarat	256.70	Meghalaya	206.67	Gujarat	147.10
11	Tamil Nadu	257.67	Punjab	215.25	Tamil Nadu	154.87
12	Arunachal Pradesh	265.89	Rajasthan	216.52	Punjab	155.05

Table 3: Whipple's Index score across the states of India

Sl. No.	States	1991	States	2001	States	2011
13	Punjab	271.89	Tamil Nadu	220.02	Maharashtra	157.38
14	Tripura	273.19	West Bengal	220.35	Haryana	157.93
15	Odisha	274.88	Maharashtra	222.86	Arunachal Pradesh	159.72
16	Assam	275.57	Tripura	223.52	West Bengal	163.23
17	Haryana	277.80	Madhya Pradesh	223.63	Madhya Pradesh	168.77
18	**India**	**290.30**	Arunachal Pradesh	228.70	Odisha	170.01
19	Maharashtra	292.50	**India**	**229.99**	Jammu & Kashmir	170.17
20	Madhya Pradesh	293.50	Odisha	239.31	**India**	**171.04**
21	Karnataka	301.27	Jammu & Kashmir	242.57	Assam	173.34
22	Uttar Pradesh	319.08	Assam	247.05	Rajasthan	179.43
23	Andhra	325.44	Uttar Pradesh	249.58	Karnataka	180.57
24	Rajasthan	325.55	Andhra	253.10	Uttar Pradesh	186.08
25	Bihar	343.60	Karnataka	253.54	Bihar	189.36
26	Jammu & Kashmir	NA	Bihar	263.75	Andhra Pradesh	191.23

Note: In 2001 the state Jharkhand was bifurcated from Bihar, Uttaranchal was bifurcated from Uttar Pradesh and Chhattisgarh curved out from Madhya Pradesh. Thus, the figure of 2001 and 2011 censuses of Uttar Pradesh, Bihar and Madhya Pradesh is the adjusted figure (Uttar Pradesh and Uttaranchal), (Bihar and Jharkhand) and (Madhya Pradesh with Chhattisgarh) to make the 1991 census figure comparable with the 2001 and 2011 censuses. Source: Census of India, 1991, 2001 and 2011. NA: Census in the state of Jammu and Kashmir was not conducted due to internal disturbances in 1991.

Analysis at the state level for 2001 census showed data for four states as rough, reflecting a nuanced improvement across these states. Along with Kerala two North Eastern states of Mizoram and Sikkim and Haryana reported rough quality of data. The state of Kerala stays at the top position (*WI*-148) while Bihar lies at the bottom (*WI*-264) in the index value chart of the states in 2001 census (Mukherjee, 1976; Mukhopadhyay, 1983). Thus, consistency of index value and its implicative issues got inherited by these two extreme states from 1991. The state of Rajasthan recorded highest decline in Index value (109 points) between 1991 and 2001 census, followed by the state of Haryana with 102 points decline. The improvement in the data quality at the national level from "very rough" (1991 and 2001) to "rough" (2011) census indicates improvement in age reporting (Yusuf, 1967) and is quite encouraging. The all India index value registered 120 points decline from 1991 to 2011 census (Prakasam, 1984; Saxena, Verma, & Sharma, 1986). At the states level three small states (Mizoram, Sikkim and Kerala), in terms of size of population and geographical area reported index value ranging between 110 and 125 ("Approximate" quality). The lowest index value is calculated in the state of Mizoram (*WI*-112), followed by Kerala and Sikkim in 2011 census (Table 1). The bottom two positions are acquired by the state of Andhra Pradesh (*WI*-191) and Bihar (WI-189). During the inter-census period (from 2001 to 2011 census), the states of Tripura (85 point), Meghalaya (76 point), and Bihar (74 point) recorded highest percentage point decline in the Whipple's Index value. The lowest percentage points declined in the case of Haryana (17 point). Major contribution towards such a feat is due to the inclusion of a question on date of birth since census 2001. Moreover, the implementation of Right to Education Act, where every child has the right to include himself or herself into formal education system.

Gender Differences of Digit Preferences

This section tries to address the second objective of the study. The gender difference of digit preference at the time of age reporting also indicates interesting trend and pattern across the states over the census period in India. In 1991 census, the all India computed index value for female (*WI*-293) exceeded that of male (*WI*-288) by 5 percentage points. Similar trend is observed in the latter censuses but has shown signs of declining index value. The computed Whipple's Index value for female was 293 in 1991 census, while it comes down to 174 in 2011 census recording a 119-percentage point change. Whereas, male recorded 120 percentage points change in the same time interval. The state level analyses of the index also indicate wide regional variation. The northern and central states i.e. Bihar, Madhya Pradesh, Punjab, Haryana, Rajasthan and Gujarat reported very rough data quality for female relative to that of male (Figure 2a/b). The calculated index value of the north eastern states i.e. Meghalaya, Mizoram and Manipur also have higher female index value than male in the three consecutive censuses

It reveals the story of biasness towards female age reporting by concentrating them within the terminal digit 0 or 5 (Suong, 1995). Basically, it is the respondent during the enumeration that plays a significant role in reporting all the socio-economic and demographic information to the enumerator. In all the southern states, along with Jammu and Kashmir and Himachal Pradesh, the value for males is higher than the females along the same temporal scale. In 1991 census, computed index for male is observed to be the highest in Bihar, followed by Andhra Pradesh, while the lowest was observed in Kerala followed by Mizoram. The highest male index value in 2001 census was recorded in Karnataka (*WI*-259) followed by Andhra Pradesh (*WI*-251) and Assam (*WI*-250) indicating that age reporting for male population in these states are more biased towards the terminal digit 0 or 5 than rest of the states. In the same time period female population age in Uttar Pradesh and Bihar is more biased toward the terminal digit 0 or 5 than rest of the states. In 2011 census, Andhra Pradesh got escalated to the top position in term of male computed Whipple's Index value of 194 followed by Karnataka and Bihar. While in case of female the highest computed value is observed in Uttar Pradesh followed by the states of Bihar and Rajasthan in the same time period.

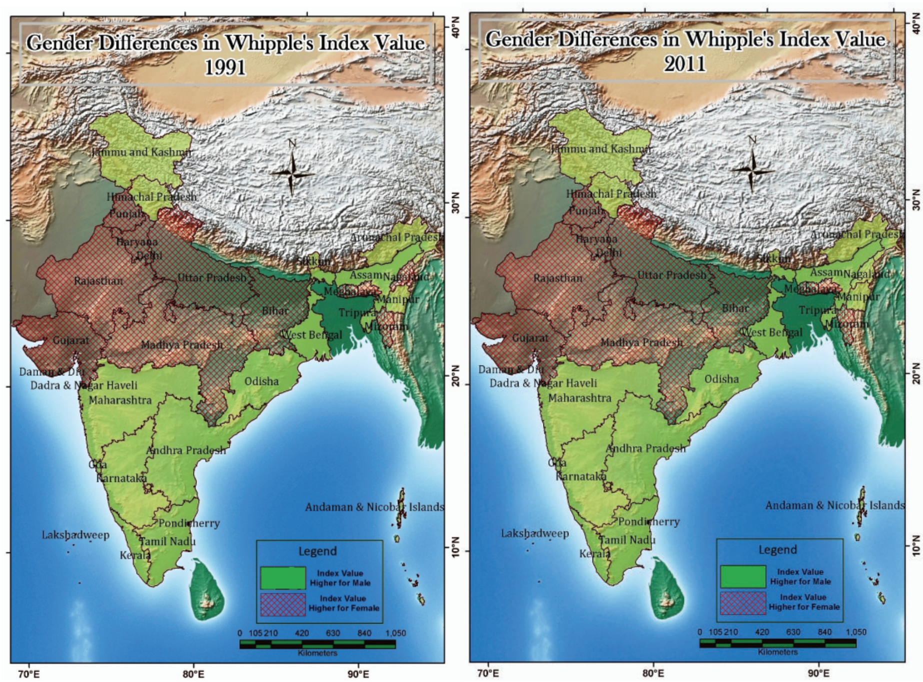

Figure 2a/b: Gender differences in Whipple's Index values for 1991 and 2011 censuses of India (Source: Prepared by the Authors, 2018)

Determinants of Age Statistics in India

A separate linear statistical model is prepared for evaluating the influencing factors of digit preference. In each model Whipple's Index is regressed for each year of age heaping and is weighted by the estimated population of the area. The regression model is used to see the relationship between this Index and development in the present study. The literacy rate, percentage of urban population and level of birth registration are taken as the proxy indicator of development.

Literacy among the indicators is the most powerful catalyst towards error free age reporting. According to one study the information of age, negligence in reckoning the exact age, deliberate misstatement and misunderstanding of the questions are responsible for the error of ages (Mukhopadhyay, 1983). In India, where still 28% of the population is illiterate (Census of India, 2011), the age data suffer from a number of problems such as ignorance of exact age, negligence in reckoning the correct age, deliberate misstatement, and misunderstanding of the question (Natarajan, 1972; Ambkanavar & Visaria, 1975; Suong, 1995). There is a good probability that the estimates made by a literate person will be closer to the actual age of an individual than that estimated by an illiterate person. Also, an educated person easily remembers his age (Ambkanavar & Visaria, 1975). Academia and researchers argue that the long run solution to the problems of age misreporting can be overcome with the socio-economic development and improvement of the level of education (i.e., literacy level) (Yusuf, 1967; Jain, 1980). But the study of Ambkanvar and Visaria (1975) found that the age data quality in Indian censuses since 1951 has deteriorated in spite of the rapid growth of literacy and education in the post-independence period. Another study has emphatically argued against any improvement in age reporting despite the improvement in educational levels between 1971 and 1991. The studies by Edmonston and Bairagi (1981) and Mukhopadhyay (1983) have not observed any significant improvement in data quality with the betterment of literacy level. Choudhary (2006) also found no such positive changes associated with the increased levels of literacy and the quality of age reporting from 1961 to 1991. However, a current study done by Agrawal and Khanduja (2015) found that the states with greater literacy rate reported higher quality age data in Census 2011.

The correlation and regression models are highly fitted in the present study. The basic assumption of the model for establishing a linear relationship is the absence of significant outliers, presence of homoscedasticity, and the normally distributed residuals (errors) of the regression line. The summary of the model presented below Model 1 (Table 4) shows Correlation Coefficient (R) that reflects on the intensity of the relationship, and Coefficient of Determination (R^2), which throws light on the degree of variation in dependent variable accounted by the independent variable. The R value in the three consecutive censuses indicates higher correlation between Whipple's Index and literacy rate and percentage of urban population. The correlation value of WI and literacy rate and percentage of urban population in 1991 census is 0.847. The R^2 value indicates 71.8% of the total variation in the dependent variable (Whipple's Index value) can be explained by the independent variables in 1991 census, which is exceptionally good. Even in 2001 census, the value of the correlation is slightly lower than the previous census but still highly correlated and the total 61.3 percent variation in the dependent variable is explained by the independent variables. The summary of 2011 census indicates highest correlation in the study time period. The correlation value is a bit more than 0.85 indicating very high correlation among the studied variables. Similarly, total variation in the dependent variable as explained by the independent variables is also very high (73.5%). The standard errors (residual) of the estimates over the years have also decreased.

Table 4: Model 1 – Possible determinates for quality of age data (Whipple's Index)

	(1991)	(2001)	(2011)
Literacy (%)	**-3.642 (-7.39)*** **	**-2.903 (5.90)*** **	**-2.737 (-7.21)*** **
Level of Birth Registration	**NA**	**-1.801 (4.02)** **	**-1.382 (-3.80)** **
Urban Population (%)	**-1.817 (-2.93)* **	**-.875 (2.04)* **	**.379 (-1.55)* **
Constant	**425.09(17.41)*** **	**387.53 (13.19)*** **	**351.94 (13.92)*** **
R (Correlation)	0.847	0.783	0.857
R Square (Co-Efficient)	**0.718**	**0.613**	**0.735**
F- Statistics	27.95***	18.18***	31.92***
Number of Observations	**24**@	**25**	**25**

Note: Dependent variable Whipple's Index, Predictors: (Constant), urban population, Literacy rate, T-Statistics are presented in the parentheses. *** Result is significant at 1% level and * result is significant at 10% level. @= Number of Observation in 1991 is 24 (No census enumeration took place in Jammu and Kashmir in 1991) while for rest of the censuses it was conducted and thus the number of observations is 25. Source: Computed by authors, 2018.

The result of the ANOVA (regression equation fits for the data or statistically significance of the model) along the same time period has also validated the regression model. The statistical significance of the regression model in the present study indicates that the regression model predicts the dependent variable significantly for the study. Here the results of 1991, 2001 and 2011 censuses are significant at 1% level. Thus, overall the regression model run is a good fit for the data and it significantly predicts the outcome of Whipple's Index. The coefficients table generated in the regression model provides us with the necessary information to predict a value of the Index from the literacy rate and percentage of urban population apart from determining whether predictors contribute significantly to the model. The Model 1 presented the predicted value of Whipple's index of the three censuses. The model 1991 predicted that one unit increase in literacy rate one an average will decrease the index value by 3.64 units which are statistically significant at 1 percent level (Saxena et al., 1986; Agrawal & Khanduja, 2015). Similarly, with the percentage of urban population going up we can see the value of Whipple's Index decrease (Model 1). The model predicted that one-unit change in urban population leads to a decline of 1.81 units of the Index value and the result is significant at 10% level. Although over the census years the unstandardized coefficients (Beta value) of literacy rate are showing a declining trend but still in 2011 census, it is the most determining variable to enhance the data quality of age in India. Similarly, the role of a percentage of urban population has also registered a declining trend. The 2011 census indicates that still one-unit change (increase) of literacy may decrease the Whipple's Index value by 2.73 units and the result is statistically significant at 1% level. It re-concretizes the assumption that the rising level of education creates awareness among the people which ultimately leads to fewer digit preferences during the reporting (Yusuf, 1967; Ambkanavar & Visaria, 1975; Jain, 1980; Saxena, et.al, 1986). At the same time, period one-unit change in the urban population may only decrease *WI* value of 0.37 unit which is lower than the previous census (1991) and significant at 10% level. This indicates that over the decades the determining effect of urban population has reduced.

It is expected that there would be fewer chances of misreporting of age (errors) with the maintenance of birth registration. The information of such "Vital Statistics" (birth) in India is based on the civil registration system, which is collected and maintained by the Registrar General of India. The level of birth registration is calculated by the method of number of registered births during the year divided by the number of estimated births for the year. It can be said that one-unit improvement of level of birth registration will lead to decrease in 1.80 units of Whipple's Index value and the result is statistically significant at 5% level. A similar result is observed in 2011 census. Thus, it can be believed that improvement in birth registration will lead to fewer errors inaccurate age reporting. This momentum may in future provide more reliable data on age distribution and less digit preferences. The level of birth registration is very satisfactory in most of the North-Eastern states (Manipur, Mizoram, Nagaland, Meghalaya, and Sikkim) along with the three Southern states (Kerala, Goa and Tamil Nadu). These are states which also reported the very low level of Whipple's Index with a sign of improvement over the last three decades. However, the diverse cultural inclinations in India because of the diversified population may also have contributed towards avoidance or preference for certain digits at the time of age reporting.

Conclusion

The evaluation of reliability and quality of age data with reference to preference or avoidance of certain terminal digits (here 0 and 5) in India and across the states revealed that the data quality of India is highly biased towards age ending with terminal digits 0 and 5 but is moving gradually towards a more reliable data. Few small states emerged as faring well in such betterment. Besides, the gender biases of the digit preferences in India indicate a spatial regional pattern. Age reporting errors or biases are more severe for women than men in most of the major states in India. Even the age distribution data quality is rough to very rough in the study time period but despite showing signs of improvement since 1991 census in India.

The trend analysis of the Whipple's Index score across the states also indicates rising trends and pattern of the reliability of information pertaining to age reporting. In 2011 census, Kerala, Sikkim and Mizoram emerged as the more reliable states for age distribution data quality as compared to rest of the states of India. Similarly, declining trend in digit preferences in the childhood age bracket indicates rising consciousness among the parents. The age groups 25 to 50 reported highest preferences for terminal digit ending with "0" or "5" digit at the time of age reporting. A distinct regional pattern of digit preference has emerged from the study. While studying the said pattern in three time periods literacy emerged as the sole significant factor influencing age reporting. The score of Whipple's Index is negatively related to literacy rate across the states. To the considerable extent, the time constraints (20 days enumeration window) play a role, if not a major one, for such biases or errors in the data for India. Similarly, low skilled manpower at the local level and lack of wiliness of the enumerators and respondents has also degraded the data quality in India. Sometimes the engagement of any local influential persons or Gram Panchayat member in the discussion also shapes the census data. Problems of such biases can be mitigated by well-kept civil registers which is available in India but lacks proper maintenance. It is also necessary to improve the methods of data collection and their feeding into digital formats.

Also, instead of giving additional charges to the local administrative officer's government should come up with a separate administrative infrastructure to facilitate quality census operation. Finally, keeping in view the intensity of erroneous age reporting it is suggested that any planning strategy should try to evaluate the quality of age data before using it for the purpose.

Acknowledgments: Authors are grateful to the anonymous reviewers whose precious comments and views helped improve this article. Further we are also thankful to Office of the Registrar General & Census Commissioner (ORGI), Government of India for their kindness to access huge data services form ORGI site and census data centre in Jawaharlal Nehru University. Both the authors also acknowledged that fellowship granted by University Grants Commission.

References

Agrawal, G., & Khanduja, P. (2015). Influence of Literacy on India's Tendency for Age Misreporting: Evidence from Census 2011. *Journal of Population and Social Studies*, *23*(1), 47–56. doi: https://doi.org/10.14456/jpss.2015.1

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

Byerlee, D., & Terera, G. (1981). Factors affecting reliability in age estimation in rural West Africa: A statistical analysis. *Population Studies, 35*(5), 455–465. doi: https://doi.org/10.1080/00324728.1981.11878517

Caldwell, J. C. (1966). A Study of Age Misstatement among Young Children in Ghana. *Demography, 3*(2), 477–490. doi: https://doi.org/10.2307/2060173.

Caldwell, J. C., & Igun, A. A. (1971). An Experiment with Census-type Age Enumeration in Nigeria. *Population Studies*, *25*(2), 287–302. doi: https://doi.org/10.1080/00324728.1971.10405804

Carrier, N., & Hobcraft, J. (1971). *Demographic estimation for developing societies*. London: Population Investigation Committee, London School of Economics.

Cleveland, L. (1996). Cross-Census Assessment of Age Sex Ratios: An application of newly UN assessment guidelines to micro data census samples from the IPUMS International. Retrieved from www.statistics.gov.hk/wsc/CPS103-P11-S.pdf

Choudhary, C. R. (2006). *A Study of Quality of Single Year Age Data in India*. *MPS Seminar Paper*, Mumbai, International Institute for Population Sciences.

Denic, S., Khatib, F., & Saadi, H. (2004). Quality of age data in patients from developing countries. *Journal of Public Health*, *26*(2), 168–171. doi: https://doi.org/10.1093/pubmed/fdh131

Edmonston, B., & Bairagi, R. (1981). Errors in age reporting in Bengali populations. *Rural Demography, 8*(1), 63–87.

Ewbank, D. C. (1981). *Age Misreporting and Age-Selective Under enumeration: Sources, Patterns, and Consequences for Demographic Analysis*. Washington, DC: National Academy Press.

González, J. F., Attanasio, L., & Trang Ha, J. (2014). An Assessment of the Age Reporting in the IPUMS-I Microdata. Paper submitted for presentation at the *2014 Annual Meeting of the Population Association of America*.

Jain, S. P. (1980). Census single year age returns and informant bias. *Demography India, 9*(1/2), 286-296.

Jowett, J. A., & Li, Y.-Q. (1992). Age — Heaping: Contrasting Patterns from China. *GeoJournal, 28*(4), 427–442. doi: https://doi.org/10.1007/BF00273112

Moultrie, T. A., Sayi, T. S., & Timæus, I. M. (2012). Birth intervals, postponement, and fertility decline in Africa: A new type of transition?. *Population Studies, 66*(3), 241–258. doi: https://doi.org/10.1080/00324728.2012.701660

Mukherjee, B. N., & Mukhopadhyay, B. K. (1988). A Study of Digit Preference and Quality of age data in Turkish Census. *Genus, 44*(1/2), 201-227. Retrieved from http://www.jstor.org/stable/29788946

Mukherjee, S. B. (1976). *The Age Distribution of the Indian Population. A Reconstruction for the States and Territories, 1881–1961*. pp. 257, Honolulu, Hawaii, USA: East West Centre, East West Population Institute.

Mukhopadhyay, B. K. (1983). Pattern of change in age reporting during 1961–71, Indian census data. *Demography India, 12*(1), 131–144.

Nagi, M. H., Stockwell, E. G., & Snavley, L. M. (1973). Digit Preference and Avoidance in the Age Statistics of Some Recent African Censuses: Some Patterns and Correlates. *International Statistical Review, 41*(2), 165–174. doi: https://doi.org/10.2307/1402833

Natarajan, D. (1972). *Age and Marital Status, Census of India, 1971.* New Delhi, India: Registrar General.

Office of the Registrar General and Census Commissioner. (1991). *Census of India 1991, Report and Series on Age, Series 1.* New Delhi, India: Registrar General and Census Commissioner.

Office of the Registrar General and Census Commissioner. (2001). *Census of India 2001, Report and Series on Age, Series 1.* New Delhi, India: Registrar General and Census Commissioner.

Office of the Registrar General and Census Commissioner. (2011). *Census of India 2011, Report and Series on Age, Series 1.* New Delhi, India: Registrar General and Census Commissioner.

Palamuleni, M. E. (2012). Age reporting in the North West Province, South Africa, 1996-2007. Proceedings from the *2013 Annual Meeting of the Population Association of America*, New Orleans, April 11–13.

Pardeshi, G. S. (2010). Age Heaping and Accuracy of Age Data Collected During a Community Survey in the Yavatmal District, Maharashtra. *Indian Journal of Community Medicine, 35*(3), 391–395. doi: https://doi.org/10.4103/0970-0218.69256

Prakasam, C. P. (1984). *On quality of age data for population count-1981, in Indian states*. Paper presented at Annual Conference of Indian Association for the Study of Population, Bangalore, 24–27 December 1984.

Saxena, P. C., Verma, K. R., & Sharma, K. A. (1986). Errors in age reporting in India, a socio-cultural and psychological explanation. *Indian Journal of Social Work, 47*(2), 127–135.

Scott, C., & Sabagh, G. (1970). The historical calendar as a method of estimating age: The experience of the Moroccan Multi-purpose Sample Survey of 1961–63. *Population Studies, 24*(1), 93–109. doi: https://doi.org/10.1080/00324728.1970.10406115

Siegel, J. S., & Swanson, D. A. (2004). *The Methods and Materials of Demography*. London: Elsevier Academic Press Inc.

Srinivasan, K., & Shastri, V. D. (2001). *Errors in age reporting of children in the 2001 Census: A preliminary appraisal*. Paper presented at Symposium on Sex Ratio in India, Mumbai, 10–11 January.

Stockwell, E. G. (1966). Patterns of digit preference and avoidance in the age statistics of some recent national censuses: A test of the turner hypothesis. *Eugenics Quarterly, 13*(3), 205–208. doi: https://doi.org/10.1080/19485565.1966.9987668

Suong, Y. (1995). *Quality of Age Data by Sex in Censuses of Some Selected Asian Countries*. DPS Seminar Paper, Mumbai, International Institute for Population Sciences.

Unisa, S., Dwivedi, L. K., Reshmi, R. S., & Kumar, K. (2015). Missing Age of Population in Indian Censuses during the Last 50 Years: Prevalence, Trends and Differentials. *Social Science Spectrum, 1*(1), 8–19. Retrieved from http://www.socialspectrum.in/index.php/sp/article/view/7

United Nations Statistics Division. (2012). *United Nations Regional Workshop on Census Data Evaluation, Final Report of the Workshop*, Department of Economic and Social Affairs. Retrieved from https://unstats.un.org/unsd/demographic/meetings/wshops/Uganda/2012/docs/WS%20Uganda%20final%20report.pdf

United Nations Statistics Division. (2013). *A study of age reporting in selected Arab Censuses of population*, Economic and Social Commission for Western Asia.

Vital statistics of India Based on the Civil Registration System. (2012). Office of the registrar general, India Ministry of home affairs, Vital statistics Division, West Block-I, R. K. Puram, New Delhi- 110066.

Yusuf, F. (1967). On the Extent of Digital Preference in Reporting of Ages in Pakistan. *The Pakistan Development Review, 7*(4), 519–532.