ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 01, pp.45144-45152, January, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

UNDERSTANDING DIGIT PREFERENCES IN INDIA USING MODIFIED WHIPPLE INDEX: AN ANALYSIS OF 640 DISTRICTS OF INDIA

***Manish Singh**

International Institute for Population Sciences, Govandi Station Road, Deonar, Mumbai 400 088, India

ARTICLE INFO

Article History:
Received 29th October, 2016
Received in revised form
14th November, 2016
Accepted 02ed December, 2016
Published online 31st January, 2017

Key words:

Census, Age-misreporting,
Myers' blended index, Whipple index.

ABSTRACT

In the present study, an attempt has been made to review the original Whipple index, and several modifications suggested to it over time. Several researchers throughout the world try to modify the Whipple index by changing the linearity assumptions and single year age range to give the better result. This study also tries to re-modify the modified Whipple index by changing the single year age range and compare it with the other indices. To know the current scenario of age misreporting in India, its states and district this newly developed index is used for calculation. The data required for this study is single year age-sex distribution of the population of India, its states and district for the census years 2001 and 2011.

Citation: Manish Singh, 2017. "Understanding digit preferences in India using modified Whipple index: An analysis of 640 districts of India", *International Journal of Current Research*, 9, (01), 45144-45152.

INTRODUCTION

Information on "Age data" collected from censuses and other demographic sample surveys being affected by several age reporting errors need rigorous accuracy checks before using them for any purpose, especially in developing countries like India, particularly in certain specific population groups (Registrar General of India, 2008). Usually, age data suffers from problems like the 'age under-enumeration' and 'age distortions' due to liking for certain ages e.g. digits like 0 and 5 as preferred more compared to digits like 1 or 9 in societies having low literacy rate. People while reporting age data does not realize 'age' as an important factor as age misreporting in fact, influence to a great extent all types of demographic analysis that consider 'age' as a variable (Registrar General of India, 2008). Such a tendency of digit preference is known as heaping and is usually seen in single-year age returns of censuses and sample surveys. Respondent's unawareness about his/her 'data of birth' is said to be a major cause for incorrect reporting of their age. In Indian census, the family head who is the usual respondent of census question might not be aware of ages of each and every family members and thus there is chance of misreporting of ages (Jain, 1980). Mason and Cope (1987) says that there are four main causes of age misreporting in any censuses or surveys. Ignorance of actual ages, miscommunication between interviewers and informants, distortion of ages, errors in recording or processing.

****Corresponding author: Manish Singh,***
International Institute for Population Sciences, Govandi Station Road, Deonar, Mumbai 400 088, India.

Review of relevant literature

Evidence shows that age misreporting errors in the age data of India, its states, and union territories have been studied by the Registrar General of India over the years, by means of applying various indexes, consecutively, by means of analysing the census age return reports of various censuses. A comparative overview of Registrar General of India's various efforts in analysing the age data can be evidenced from a reading of the Chapter-4 on "Appraisal of Age Data" given in Registrar General of India (2008,Pp.73-98). However, several individual researchers also made several attempts to gauze the misreporting errors in age data of India, its states and union territories by means of various studies (see, Unisa et al., 2009; Prakasam, 1984, Pardeshi, 2010). In essence all the above studies brings out the fact that there is a continuous need to evaluate and adjust the age data before using the same for further analysis, let the data be collected from census or surveys. Thus said, it is to say few attempt will be made further to review the above work in detail as the purpose of the present study is somewhat different.

Objectives

- To check the validity of extended modified Whipple index, the present study compare the four indices namely Whipple's, Myer's, Modified Whipple and Extended Modified Whipple indices in India and it's major states of census year 2001 by sex.

- To assess the quality of single year age data, by sex and residence separately, of India and its major states by means of computing extended modified Whipple index of census year 2001 and 2011.
- To calculate age misreporting in 640 districts of India using extended modified Whipple index of census year 2011.

MATERIALS AND METHODS

The Whipple's Index and its modifications (An appraisal)

The discussion carried out below is heavily based on the original Whipple's index and its modification. The discussion made in the studies by Noumbissi (1992), Spoornberg (2007, 2009)and Nasir and Hinde (2014).

The Original Whipple's Index (WI)

The Whipple's original index of digit preference popularly known as the index of concentration is calculated in two steps using the single year age data of each sex separately of the ages 23 to 62, as follows: Step (1): Sum the number of persons in the age range 23 and 62 inclusive, Step (2): Calculate the ratio of the reported ages ending in 0 or 5 to one-fifth of the total sample. In a formula, WI original index can be shown as:
W = 5 (P25 + P30 + P35 + …+P60)/(P23+P24+…+P61+P62) -----(A)

Where,

The above index is developed assuming that a "continuous and linear decrease in the number of persons of each age within the age range of 23 and 62." It implies that the above linearity assumption cannot be applied to other ages of 0-22 years and 63 and above years. The index thus constructed uses the age range of 23 and 62, both inclusive. It is realized the choice of the age limits fixed here is made very arbitrarily but found to be effective.

The calculated values of WI index thus defined above are observed to fall between a minimum value of 100 and a maximum value of 500. A maximum is expected only when "no returns are recorded with any digits other than the two 0 and 5." The quality of the index sometimes divided into highly accurate (when WI is under 105), fairly accurate (if WI is in between 105-110), Approximate (if WI is in between 110-125), Rough (if WI is in between 125-175), and Very rough (if WI is more than 175) (Registrar General India, 2008, P.74).

The First Modification (WI)

Roger et al. (1981, p.148) gives the following two formulae as first modifications for the original WI:

$W0 = 10\ (P_{30} + P_{40} + P_{50} + P_{60})/(P_{23}+P_{24}+\ldots+P_{61}+P_{62})$ -----(B)
$W5 = 10\ (P_{25} + P_{35} + P_{45} + P_{55})/(P_{23}+P_{24}+\ldots+P_{61}+P_{62})$ -----(C)
From the above two W may be obtained as W = (W0 + W5) /2 ---------(D)

This first modification allows one to distinguish age preferences are in favour of 0 or 5, but realized that they were derived assuming a linearity over a ten-year age range, which is unrealistic (Spoorenberg, 2007).

The Second Modification: Noumbissi (1992)

This modification suggested by Noumbissi (1992) unlike the first, but following the original WI, "is based on the linearity assumption over an age range of 5 years" and introduces the following ten formulae and thus allows one to calculate age heaping for all 10 digits of 0, 1,2,..,9.

Digit specific modified Whipple's indexes (W_i's) for all 10 digits are as given below:

$W_0 = 5\ (P_{30} + P_{40} + P_{50} + P_{60})/({}_5P_{28}+{}_5P_{38}+{}_5P_{48}+{}_5P_{58})$ -----(E)

$W_5 = 5\ (P_{25} + P_{35} + P_{45} + P_{55})/({}_5P_{23}+{}_5P_{33}+{}_5P_{43}+{}_5P_{53})$ -----(F)

$W_1 = 5\ (P_{31} + P_{41} + P_{51} + P_{61})/({}_5P_{29}+{}_5P_{39}+{}_5P_{49}+{}_5P_{59})$ -----(G)

$W_2 = 5\ (P_{32} + P_{42} + P_{52} + P_{62})/({}_5P_{30}+{}_5P_{40}+{}_5P_{50}+{}_5P_{60})$ -----(H)

$W_3 = 5\ (P_{23} + P_{33} + P_{43} + P_{53})/({}_5P_{21}+{}_5P_{31}+{}_5P_{41}+{}_5P_{51})$ -----(I)

$W_4 = 5\ (P_{24} + P_{34} + P_{44} + P_{54})/({}_5P_{22}+{}_5P_{32}+{}_5P_{42}+{}_5P_{52})$ -----(J)

$W_6 = 5\ (P_{26} + P_{36} + P_{46} + P_{56})/({}_5P_{24}+{}_5P_{34}+{}_5P_{44}+{}_5P_{54})$ -----(K)

$W_7 = 5\ (P_{27} + P_{37} + P_{47} + P_{57})/({}_5P_{25}+{}_5P_{35}+{}_5P_{45}+{}_5P_{55})$ -----(L)

$W_8 = 5\ (P_{28} + P_{38} + P_{48} + P_{58})/({}_5P_{26}+{}_5P_{36}+{}_5P_{46}+{}_5P_{56})$ -----(M)

$W_9 = 5\ (P_{29} + P_{39} + P_{49} + P_{59})/({}_5P_{27}+{}_5P_{37}+{}_5P_{47}+{}_5P_{57})$ -----(N)

It is noted that

$W_i = 1$ (indicates there is no digit preference or avoidance)
$W_i < 1$ or $W_i > 1$ (indicates there is a digit preference or avoidance of the digit)

The Third Modification: (Spoorenberg, 2007)

Realizing the fact that the above 10 digit-specific modified Whipple's indexes are very cumbersome to handle when studying the spatial or temporal etc. aspects, Spoorenberg (2007, p.17) suggested the following total modified Whipple's Index (W_{tot}) as a summary measure that summarizes all age performance and avoidance effects:

$Wtot = \sum_{i=0}^{9}(|W_i \quad 1|)$ ---- (o)

Here,

W_i = digit-specific modified Whipple's index for each of the ten digits (0-9) developed by Noumbissi.
$W_{tot} = 0$, when there is no preference
$W_{tot} = 16$, when $W_0 = W_5 = 5$ and all other $W_i = 0$

Validity of the summary index W_{tot} has been verified by Spoorenberg by applying it to various single-year age data sets of various regions and time periods in the world (Spoonberg, 2009). It has also been compared with that of Myers' blended index and the original Whipple's index. The results of the analysis show that W_{tot} performs well when compared to original Whipple's index and 'it produces practically the same results as Myer's blended index (Spoorenberg, 2007, P.737)."Favouring W_{tot}, and dismissing the original WI, (Spoorenberg, 2009, p.5) in his conclusion states that " If one

wants to assess with more precision the quality of age reporting and its change through time the original Whipple's index is not a completely fair and reliable measure." Thus this section may be concluded with the end note that W_{tot} may be used as a fitting alternative to the traditionally used original Whipple's index and the Myers' blended index, in all future analysis of the age sex data. It is also realized that W_{tot} is easy to calculate and easy to understand. In addition can be shown graphically in comparing time trends of different units of *analyses* like states in India.

The Fourth Modification: (Nasir and Hinde, 2014)

Nasir and Hinde in 2014 states that to reduce the margin of error in developing countries it is better to use three year linearity assumption rather than ten of five. They further developed the total modified Whipple's Index by changing linearity assumptions over three year age range. They proposed the following expression to measure the age misreporting on terminal digit 0 as follows.

$$W0 = 3\ (P_{20} + P_{30} + P_{40} + P_{50})/({}_3P_{20}+{}_3P_{30}+{}_3P_{40}+{}_3P_{50}). \text{ -----(P)}$$

The general expression for all the terminal digit is as follows:

$$FMW_i = 3\ (P_{1i} + P_{2i} + P_{3i} + P_{4i} + P_{5i})/({}_3P_{1i}+{}_3P_{2i}+{}_3P_{3i}+{}_3P_{4i}+{}_3P_{5i}) \text{ ----(Q)}$$

Where i =0, 1, 2…9.

Also, $FMW^{*}_{tot} = \sum_{i=0}^{9}(|FMW_i \quad 1|)$ ---- (R)

The Fifth Modification

The total modified Whipple's Index and Myers' blended method both calculate the age misreporting on all the terminal digit. Data required for calculation is the single year age distribution of population but the total modified Whipple's Index use the range of 21 to 64 while on the other hand Myers' blended method use the range of 10 to 99. Keeping in mind all the suggestions and modifications over time the present study propose a new index which is nothing but the extension of the range from (10-91) and apply in India and its major states of census year 2001 and 2011.

The general expression for the proposed study are as follows:

$$EMW_i = 3(P_{1i} + P_{2i} + P_{3i} + P_{4i} + P_{5i} + P_{6i} + P_{7i} + P_{8i})/({}_3P_{1i}+{}_3P_{2i}+{}_3P_{3i}+{}_3P_{4i}+{}_3P_{5i}+{}_3P_{6i}+{}_3P_{7i}+{}_3P_{8i}) \quad \text{(S)}$$

Where i =0, 1, 2…9.

Also, $EMW_{tot} = \sum_{i=0}^{9}(|EMW_i \quad 1|)$ ---- (T)

If there is no age heaping, then

$EMW_0 = EMW_1 = EMW_2 = \ldots\ldots = EMW_9 = 1$
and $EMW_{tot} = \sum_{i=0}^{9}(|1 \quad 1|) = 0$

If all reported ages end in 0 or 5, then $EMW_0 = EMW_5 = 3$ and all other $EMW_i = 0$.
then $EMW_{tot} = 8 \text{ X } (|0 \quad 1|) + 2 \text{ X } (|3 \quad 1|) = 12$

Here,
EMW_i = digit-specific extended modified Whipple's index for each of the ten digits (0-9).
$EMW_{tot} = 0$, when there is no age heaping

$EMW_{tot} = 12$, when $EMW_0 = EMW_5 = 3$ and all other $EMW_i = 0$.

Change of Origin and Scale

To fit the above index according to United Nations standard, I have to change the origin and scale of the index by using the equation:

(0+A)/h=100..... (U)

(12+A)/h=500..... (V)

By solving equations (U) and (V), I obtain the value of A and h as follows

A= 3 and h=0.03

The United Nation standard for measuring the age-misreporting

Whipple's Index	Quality of Data	Deviation from Perfect
< 105	very accurate	< 5%
105–110	relatively accurate	5–9.99%
110–125	ok	10–24.99%
125–175	bad	25–74.99%
> 175	very bad	≥ 75%

RESULTS AND DISCUSSION

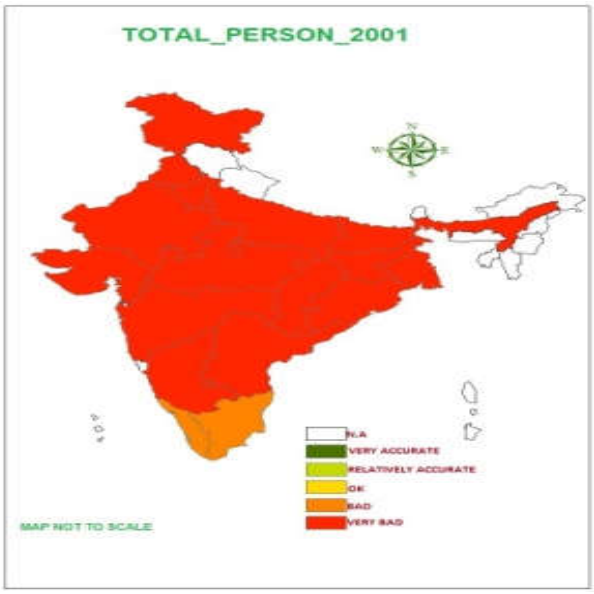
An analysis of result given in Table 1 & 2 show the comparison of four indices namely original Whipple index, Myer's index, modified Whipple index and Extended modified Whipple index of India and its major states by sex in 2001. In comparison of all the four indices it observes that the values obtained from Myer's index and Extended modified Whipple index are closer as compare to other indices. As various studies suggest that age misreporting and literacy level is highly correlated with the population which means that those population which has low literacy level experiencing high age misreporting and vice-versa. According to census 2001, the highest literacy rate was observed in the state of Kerala (male: 94.2%, female: 87.9%) and lowest literacy rate was observed in the state of Bihar (male: 60.3%, female: 33.7%). In table1 all the four indices shows that the lowest age misreporting was observed in Kerala and highest in Bihar while on the other hand, in case of females all the four indices are in favour of Kerala as lowest age misreporting but only Extended modified Whipple index is in favour of Bihar as highest age misreporting shows the accuracy of this index. A geographic information system (GIS) is used to analyze 19 major states of India by sex and residence of the census year 2001 and 2011. Figure 1 shows the map of 19 major states of India of total persons, total males and total females of the census year 2001 and 2011. These map shows the geographical variation between state to state i.e. the qualities of data varies from state to state and also regional variation exist. The Northern region of India experiences poor quality of data as compared to Central and Southern region. Figure 2 & 3 shows the 19 major states of India by sex and residence of census year 2001 and 2011. From these figures, it was observed that the quality of data improves over time in case of both males and females and also from rural and urban areas. The quality of data is poor in rural areas than urban areas because of low literacy rate in rural areas. The state of Kerala and Tamil Nadu is very exceptional in terms of less digit preference in case of both males and females and also from rural to urban areas.

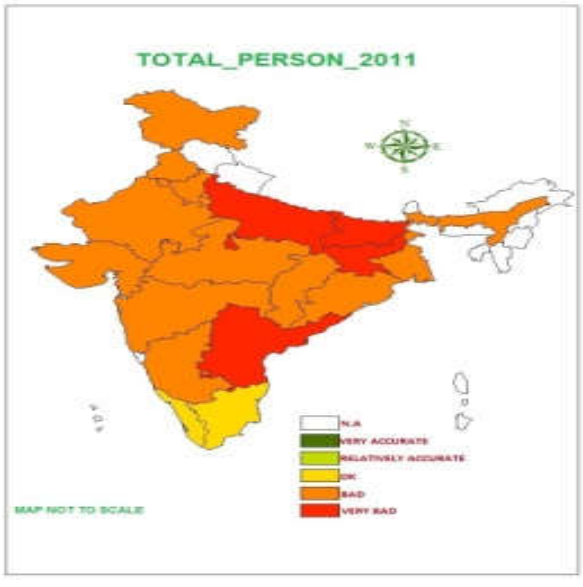
Table 1. Rankings of the major states in the ascending order of the values of the Whipple's, Myer's, modified Whipple's and extended modified Whipple's indices-2001

	Males							
Rank	India/ States	Whipple's Index	India/ States	Myer's Index	India/ States	Modified WI	India/ States	EModified WI
	India	241	India	50.8	India	5.71	India	3.81
1	Kerala	144	Kerala	20.6	Kerala	2.20	Kerala	1.52
2	Haryana	185	Haryana	34.3	Chhattisgarh	3.68	Haryana	2.57
3	Delhi	212	Tamil Nadu	38.8	Haryana	4.03	Tamil Nadu	3.03
4	Tamil Nadu	212	Maharashtra	41.9	Tamil Nadu	4.46	Delhi	3.10
5	West Bengal	215	Delhi	42.2	Maharashtra	4.88	Maharashtra	3.11
6	Maharashtra	219	Chhattisgarh	45.5	Delhi	4.94	Chhattisgarh	3.33
7	Gujarat	221	Gujarat	46.6	West Bengal	5.32	Gujarat	3.43
8	Chhattisgarh	222	West Bengal	46.9	Gujarat	5.35	West Bengal	3.46
9	Punjab	234	Punjab	49.6	Punjab	5.61	Punjab	3.66
10	Rajasthan	235	Rajasthan	50	Rajasthan	5.67	Rajasthan	3.69
11	Orissa	238	Orissa	52.2	Orissa	5.76	Karnataka	3.82
12	Assam	244	Karnataka	52.5	Assam	6.01	Assam	3.94
13	Karnataka	248	Assam	53.2	Karnataka	6.12	Orissa	3.96
14	Madhya Pradesh	251	Andhra Pradesh	53.4	Madhya Pradesh	6.21	Andhra Pradesh	3.98
15	Jammu & Kashmir	254	Jammu & Kashmir	53.5	Jammu & Kashmir	6.21	Jammu & Kashmir	3.99
16	Andhra Pradesh	255	Madhya Pradesh	55.1	Andhra Pradesh	6.24	Madhya Pradesh	4.13
17	Jharkhand	260	Jharkhand	56.9	Jharkhand	6.44	Jharkhand	4.40
18	Uttar Pradesh	294	Uttar Pradesh	66.5	Uttar Pradesh	7.82	Uttar Pradesh	5.03
19	Bihar	302	Bihar	68.2	Bihar	8.11	Bihar	5.36

Table 2. Rankings of the major states in the ascending order of the values of the Whipple's, Myer's and modified Whipple's and extended modified Whipple's indices-2001

	Females							
Rank	India/ States	Whipple's Index	India/ States	Myer's Index	India/ States	Modified WI	India/ States	EModified WI
	India	218	India	48.4	India	5.31	India	3.60
1	Kerala	151	Kerala	23.7	Kerala	2.42	Kerala	1.78
2	Haryana	166	Haryana	32.9	Chhattisgarh	3.50	Haryana	2.48
3	Gujarat	186	Delhi	41.3	Haryana	3.70	Delhi	3.00
4	Punjab	195	Gujarat	41.6	Gujarat	4.51	Gujarat	3.12
5	Rajasthan	197	Chhattisgarh	43.2	Delhi	4.78	Chhattisgarh	3.18
6	Chhattisgarh	198	Rajasthan	44.9	Rajasthan	4.84	Punjab	3.40
7	Delhi	202	Maharashtra	45.4	Punjab	5.00	Rajasthan	3.40
8	Madhya Pradesh	206	Punjab	45.4	Maharashtra	5.13	Maharashtra	3.41
9	Uttar Pradesh	207	Tamil Nadu	45.9	Tamil Nadu	5.17	Tamil Nadu	3.44
10	West Bengal	226	Madhya Pradesh	49.1	Madhya Pradesh	5.23	West Bengal	3.68
11	Maharashtra	227	West Bengal	49.1	West Bengal	5.49	Jammu & Kashmir	3.70
12	Tamil Nadu	228	Jammu & Kashmir	50.9	Jammu & Kashmir	5.81	Madhya Pradesh	3.71
13	Jammu & Kashmir	229	Orissa	52.7	Orissa	5.86	Orissa	3.99
14	Bihar	235	Jharkhand	53.4	Jharkhand	5.89	Jharkhand	4.02
15	Jharkhand	237	Assam	54.1	Uttar Pradesh	5.90	Assam	4.03
16	Orissa	241	Uttar Pradesh	54.3	Assam	6.22	Andhra Pradesh	4.14
17	Assam	250	Andhra Pradesh	56.5	Bihar	6.29	Karnataka	4.20
18	Andhra Pradesh	251	Bihar	58.2	Andhra Pradesh	6.30	Uttar Pradesh	4.27
19	Karnataka	259	Karnataka	58.6	Karnataka	6.67	Bihar	4.58





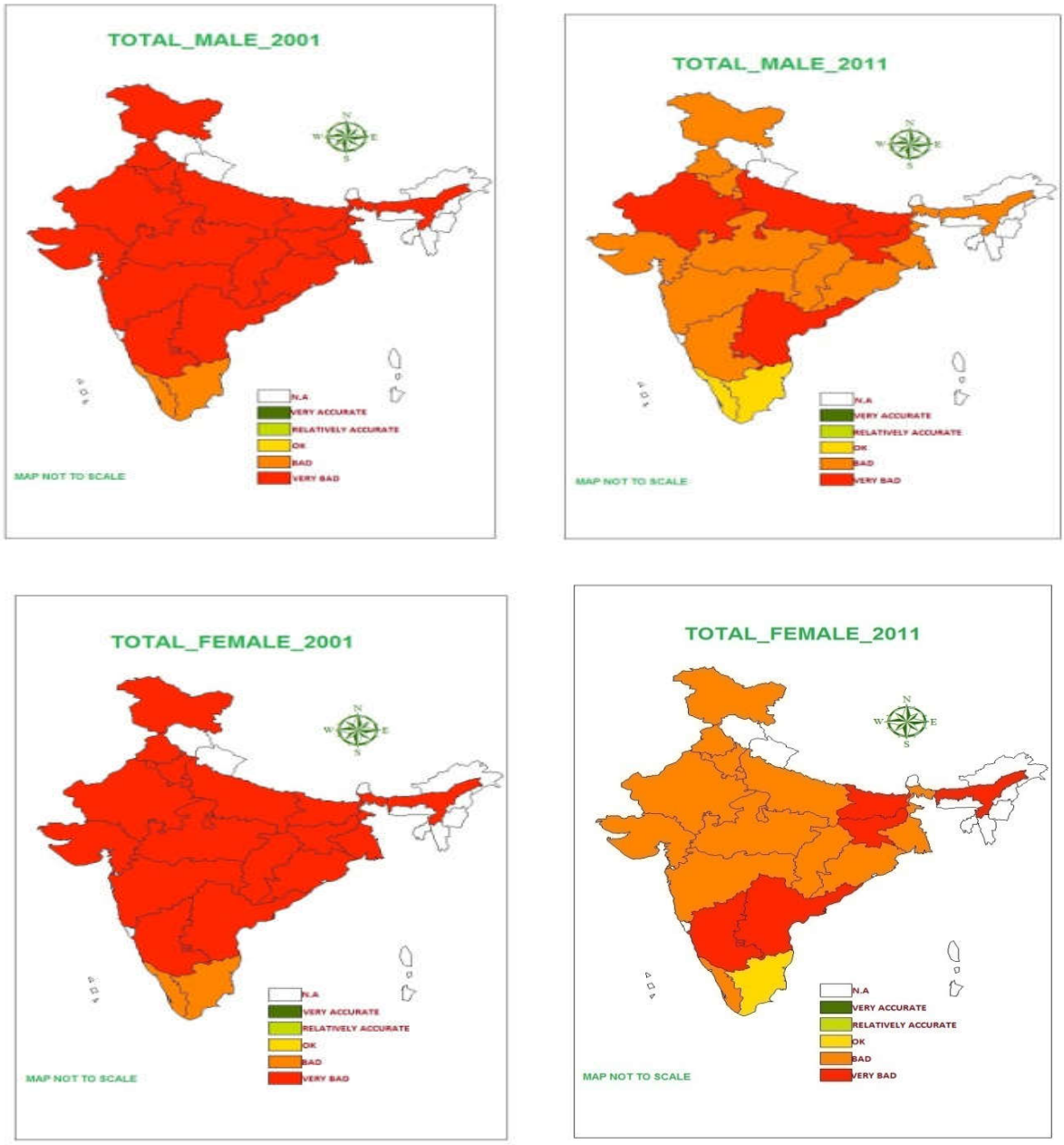
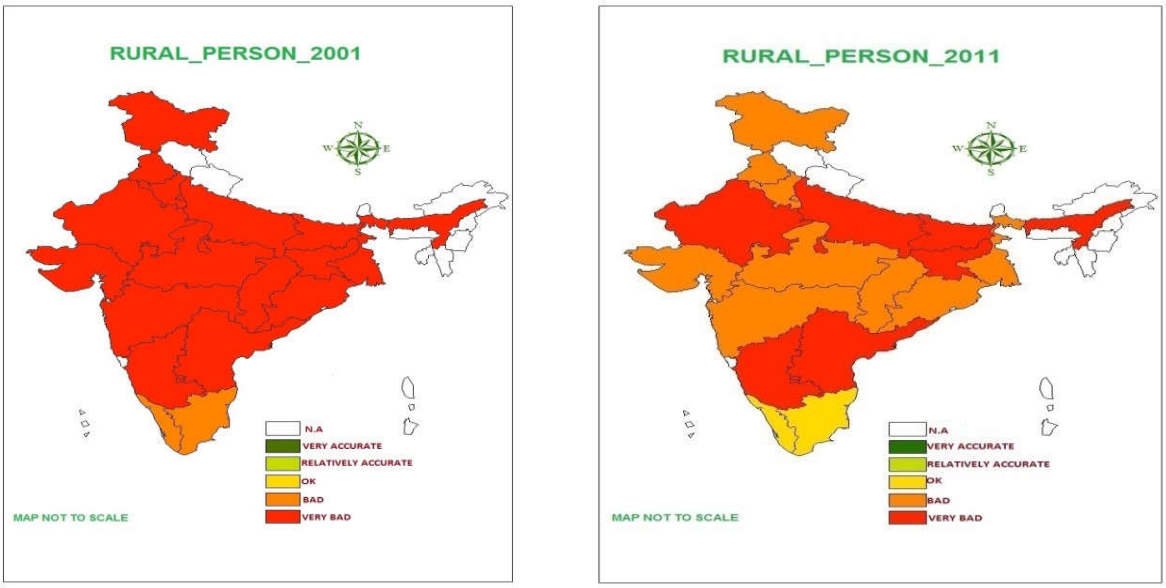


Figure 1. Map shows the 19 major states of India of Total Persons, Total Males and Total Females of Census Year 2001 and 2011



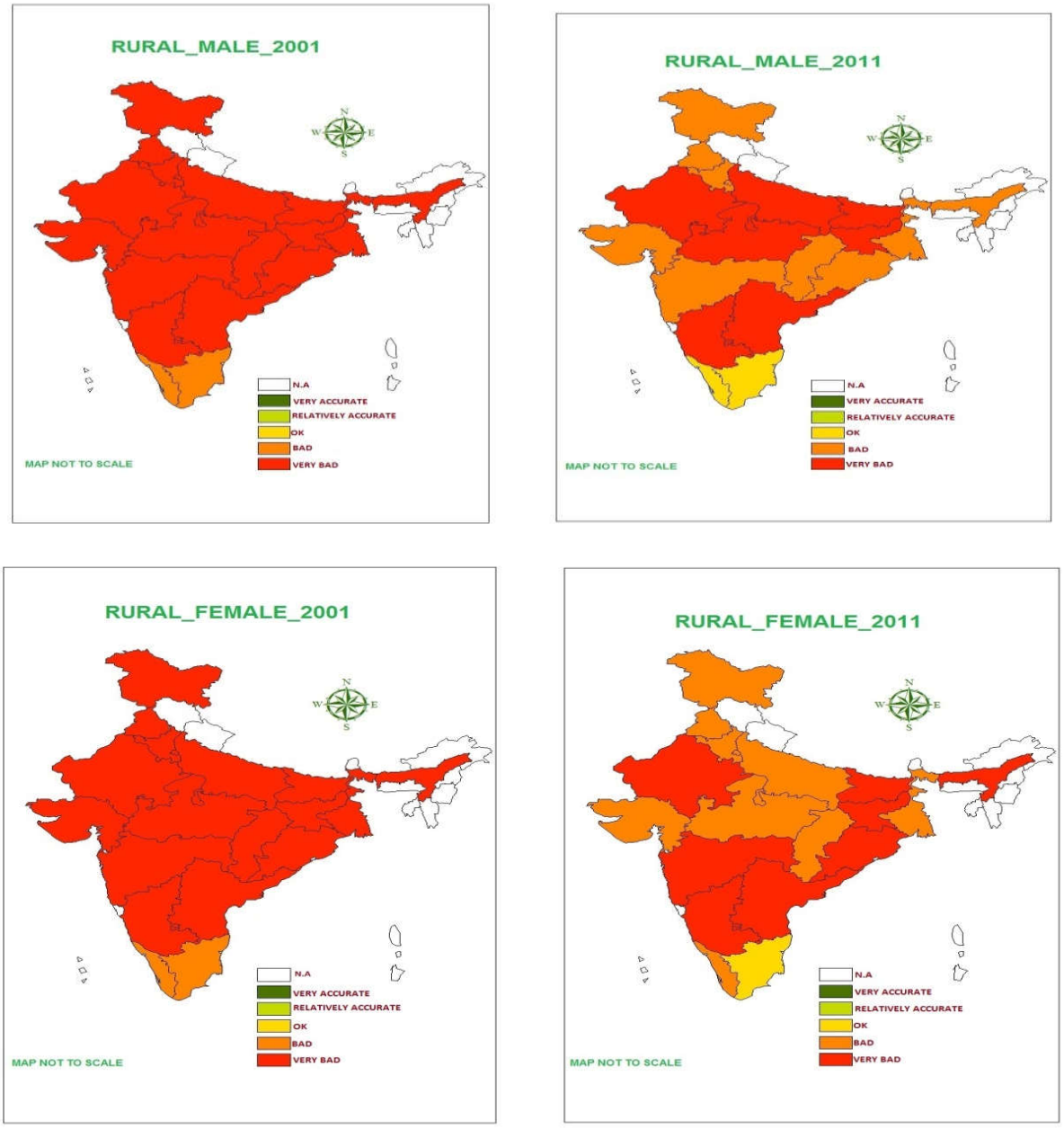
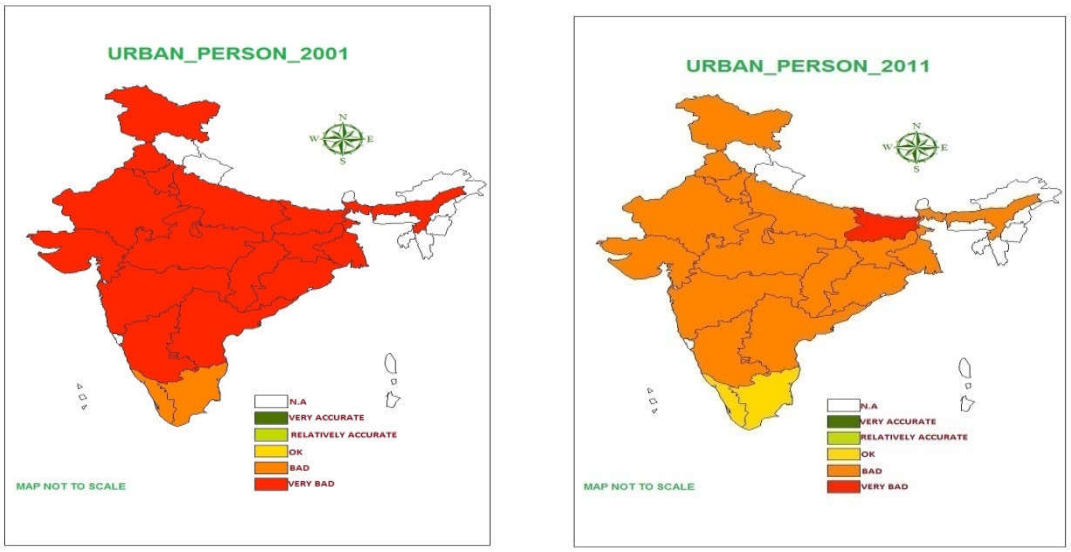


Figure 2. Map shows the 19 major states of India of Rural Persons, Rural Males and Rural Females of Census Year 2001 and 2011



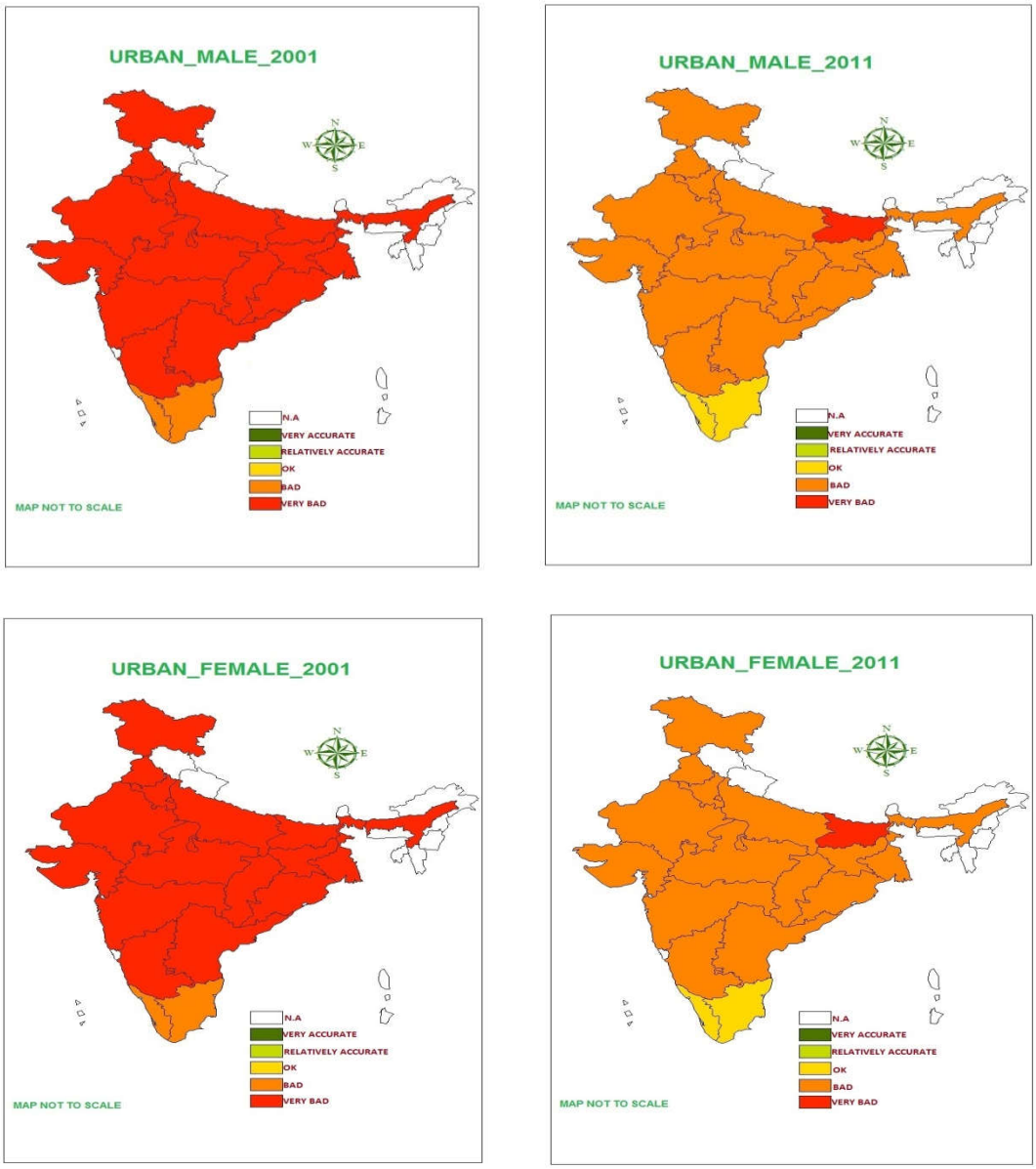


Figure 3. Map shows the 19 major states of India of Urban Persons, Urban Males and Urban Females of Census Year 2001 and 2011

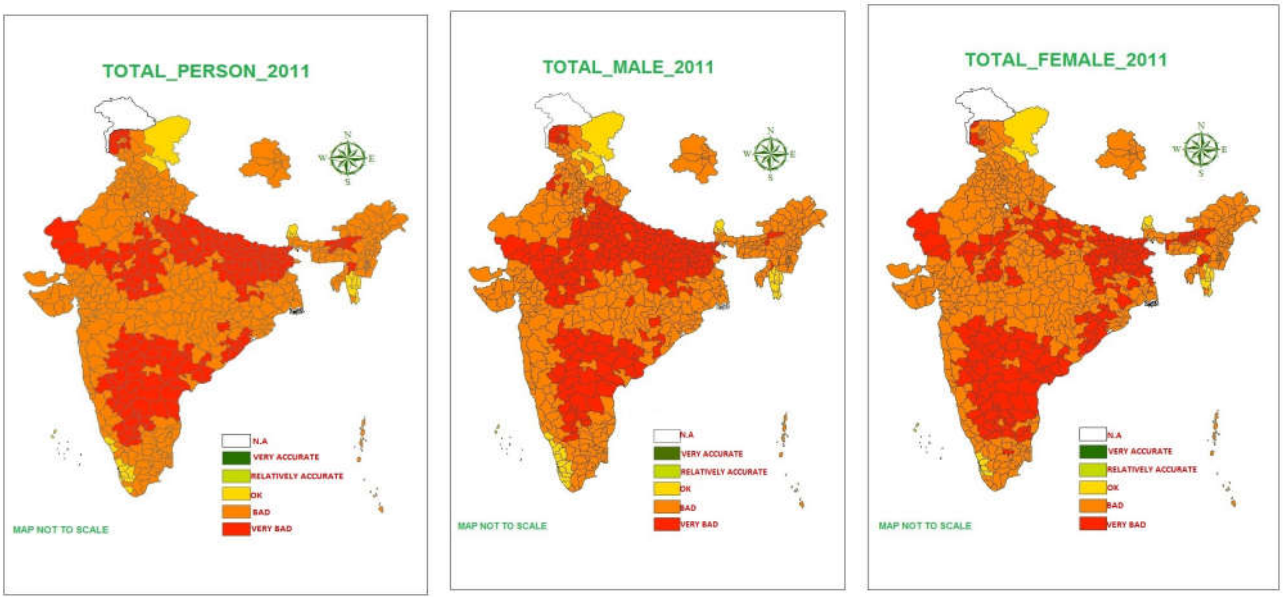


Figure 4. Map shows the 640 district of India of Total Persons, Total Males and Total Females of Census Year 2011

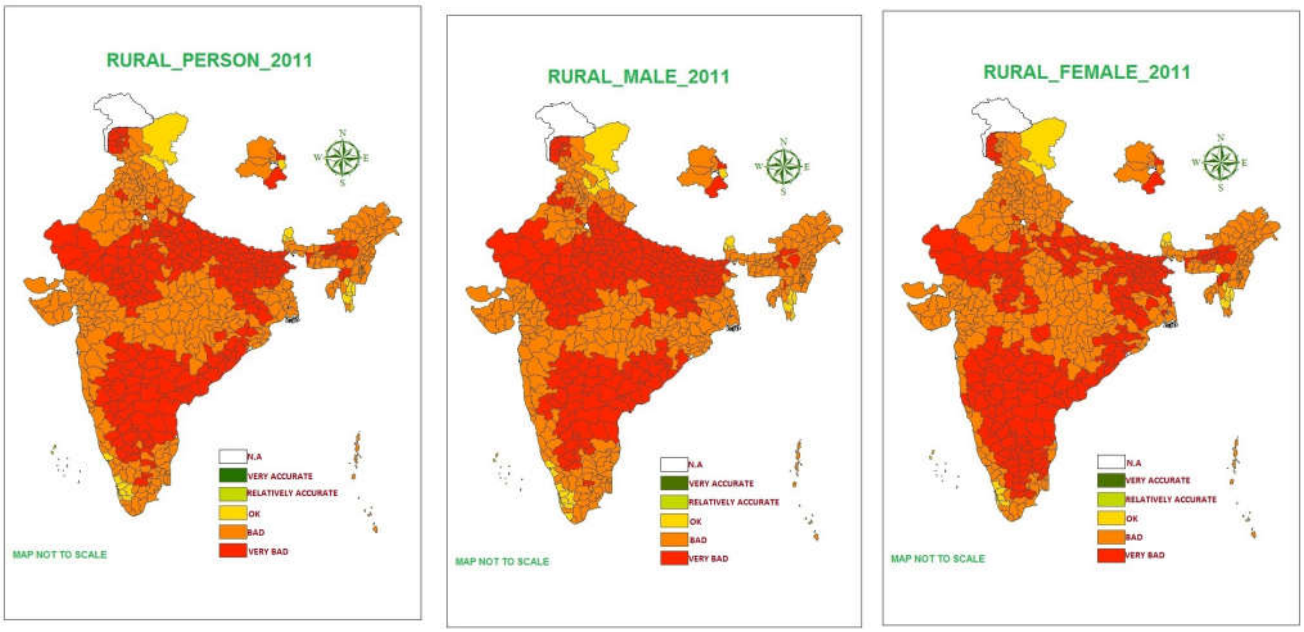


Figure 5. Map shows the 640 district of India of Rural Persons, Rural Males and Rural Females of Census Year 2011

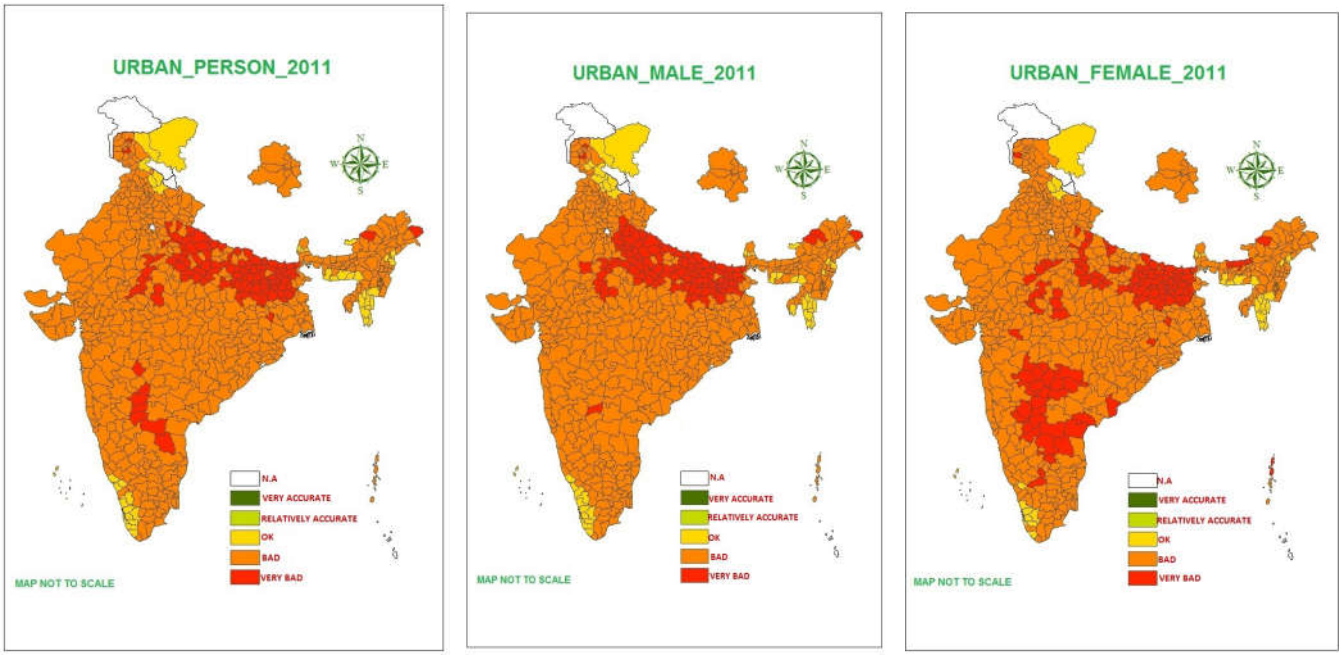


Figure 6. Map shows the 640 district of India of Urban Persons, Urban Males and Urban Females of Census Year 2011

An analysis of 640 districts of India by sex and residence shows that no districts of India is under very accurate (< 105) and relatively accurate (105–110) category. In totality, only 4% is ok (110–125), 62% is bad (125–175) and 34% is very bad (> 175). In case of males, only 6% is ok (110–125), 58% is bad (125–175) and 37% is very bad (> 175) and in case of females, only 3% is ok (110–125), 65% is bad(125–175) and 32% is very bad (> 175). Urban areas have a better quality of data as compared to rural areas because of high literacy rate in urban areas.

Conclusion

The results of the analysis indicate that extended modified Whipple's index gives more accurate results as compared to original Whipple index, Myer's index and modified Whipple index. There seems an improvement in the age data of 2011 census compared to 2001 census. No states and districts of India are under very accurate and relatively accurate category and more than 50% of the states and districts of India is under the bad category.

REFERENCES

General, R. Census Commissioner India, 1997. Census of India 1991, series 1, India, Part VIA-C Series, Social-Cultural Tables, Volume 2, Tables C-3 Part A and B, C-4,C-5 and C-6, India, states and Union Territories, New Delhi.

General, R. Census Commissioner India, 2005. Census of India, 2001,C series: Social and Cultural Tables, Table C-13: Single Year Age Returns by Residence and sex, New Delhi, Available at: http://www.censusindia.net/results/C_Series/c13_India.pdf(accessed 27 February 2007).

General, R. Census Commissioner India, 2013. Census of India, 2011,C series: Social and Cultural Tables, Table C-13: Single Year Age Returns by Residence and sex, New Delhi, Available at: www.censusindia.gov.in/2011census/ Age_level_Data/India/Age_data.xls

Jain, S.P. 1980. 'Census Single Year Returns and informant Bias', Demography India, vol.9, nos.1&2, pp.286-296.

Mason, K. O. & Lisa, G. Cope 1987. 'Sources of Age and Date-of-Birth Misreporting in the 1900 US Census'. Demography, 24(4), 563-573.

Myers, R. J. 1940. Errors and bias in the reporting of ages in census data.

Nasir, J. A. & Hinde, A. 2014. Pak. J. Statist. 2014 Vol. 30 (2), 265-272 An extension of modified whipple index–further modified whipple index. *Pak. J. Statist*., 30(2), 265-272.

Noumbissi, A. 1992. "L'indice de Whipple modifie: une application aux donnees du Cameroun, De la Suede et de la Belgique>>, Population, 47(4), pp. 1038-1041.

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: official publication of Indian Association of Preventive & Social Medicine*, 35(3), 391.

Prakasam, C.P. 1984. 'On quality of age data for population count-1981, in Indian states', Paper submitted to the Annual Conference of Indian Association for the Study of the population, held at Indian institute for Management, 24th December to 27th December, 1984, Bangalore.

Roger, G., Waltisperger, D. & Corbille-Guitton, C. 1981. Les structures par sexe et âge en Afrique.

Spoorenberg, T. 2007. Quality of age reporting: Extension and application of the modified Whipple's index. Population-E62 (4): 729-742.

Spoorenberg, T. 2009. Assessing the quality of age reporting at a time of general data quality improvement: going beyond the original Whipple's index. XXVI IUSSP International Population Conference, Morocco 27 September, 2009, Session P-5. Morocco.

Unisa S, Dwivedi LK, Reshmi RS. & Kumar K 2009. Age Reporting in Indian Census: An Insight. Paper presented in the XXVI IUSSP International Population Conference, Morocco.

****