Title: Contextualizing gender and regional differences in quality of age reporting in India: Application of post data collection quality indicators

Authors:

First Author and Corresponding Author:
Dr. Md Illias Kanchan Sk
Assistant Professor
International Institute for Population Sciences (I.I.P.S.)
Mumbai - 400088
Contact No. (+91) 8097831594
Email: illiaskanchan@gmail.com
ORCID: https://orcid.org/0000-0002-9598-9119

Second Author:
Babul Hossain
Research Scholar
International Institute for Population Sciences (I.I.P.S.)
Mumbai - 400088
Contact No. (+91) 8967106293
Email: bhossain399@gmail.com
ORCID: https://orcid.org/0000-0001-5355-1429

ABSTRACT

With the contrasting variation in the development level between the north and southern states in India, the study argues the accuracy of different approaches to estimate age misreporting and the gender difference in the age heaping process between the north and south states in India. The finding suggests that the modified total Whipple's index (W_{tot}) has been proven a more accurate and most straightforward index measuring the quality of age data. Our finding indicates no substantial difference among males and females in age misreporting, but preference for age reporting with terminal digit 0 and 5 is much higher in Northern states than Southern states.

KEYWORDS: Age misreporting, regional difference, gender, modified total Whipple's index

Introduction

Indian Census is the single most massive demographic exercise across the globe, covering a current population of over 121 million Indian residents (RGI 2011). "It is static as well as dynamic in nature. Static in the sense that it gives a broad picture of a country at a particular time point and dynamic within the meaning that every new census data can be compared to previous censuses. The dynamic nature of census helps in assessing the magnitude and direction of the demographic indicators, and also it helps in evaluating implemented policies and paves the further course of action for the policymakers (Yadav, Vishwakarma & Chauhan 2020).

It is conducted in two separate rounds: House-listing and Housing Census and Population Enumeration. The respondents' characteristics are enquired in the second phase, i.e. population enumeration. Question on age has been asked ever since the initiation of census operations. However, the kind of question has always been modified. In 1941 census, age was recorded in years and months. However, in the 1961 census, age as of last birthday was recorded. For the first time, in 2011 census, date of birth was enquired.(RGI 2011)

Irrespective of the type of question, age has continuously been of primary importance. Its age-sex pyramid and age represent the demographic profile of any country is the basic input for the same. The shape of the pyramid reflects the stage of the demographic transition of the country. Age is the single most crucial variable in any survey. Any demographic estimate or measurement needs age data as a primary input (Vaupel 1998; Sun et al. 2017).

The data on age not only presents a comprehensive snapshot of the country but also plays a pivotal role in population projections. Census data is widely used in research and policymaking and decisions. For instance, estimates of population in the age group of 5-14 or the school-going population will help decide the requirement of schools or population above 18, which represent the estimates of voters. Similarly, the estimates of population in the age groups above 60 are required for old-age social security policy and planning (Diamond 1977; Caswell et al. 2018).

However, the census data is not as accurate as of the developed countries and the maximum error in found in the age data (Cleland 1996). The event of incorrect reporting of age by the respondent or incorrect recording by the interviewer is called age misreporting. While reporting, there is a natural preference for terminal digits 0 and 5, followed by 2,4,6 and 8. Digits like 1,3, 7 and 9 are generally ignored (Jowett & Li 1992). For instance, a person aged 13 might prefer reporting his age to the closest round figure like 10 or 15 or even 12 or 14. This results in over-concentration of population data in the age groups ending with 0, 2, 4,5 6 and 8, more prominently in 0 and 5. This event is called digit preference or age heaping (Shryock, Siegel & Larmon 1980; Newell 1990). Age heaping is a misreporting of age due to habit of rounding the numbers irrespective of socio-economic background (RGI 2011). Hence, it goes without saying that inaccurate data results in faulty policymaking and decisions for the nation. In clinical studies, age is one of the most commonly assessed variables, and other parameters are often analyzed and results interpreted in relation to age. When study subjects are divided into different age groups, those who provide incorrect age could be put into a wrong age subgroup, and this may affect results (Denic, Saadi & Khatib 2004). Study has suggested that age misreporting is a hinderernce iin the proper policy management and planning process (Yadav, Vishwakarma & Chauhan 2020).

This paper aims to compare the prevalence and magnitude of age heaping and age accuracy in selected major northern and southern states of India. While the states of Punjab, Rajasthan, Uttar Pradesh and Bihar represents the Northern part of India; Andhra Pradesh, Karnataka, Kerala and Tamil Nadu represents the south. The southern states of India, especially Kerala and Tamil Nadu, have the highest rate of literacy and lowest fertility rates. The state of Kerala has already set a global standard for the Human Development Index (Franke & Chasin 1994; Véron 2001). The northern states, on the other hand, present a stark contrast. The states of Bihar and Uttar Pradesh have the lowest literacy and highest fertility rates (Rasul & Sharma 2014; Sharma 2015). Hence, the paper attempts to provide a comparative analysis assuming that there would exist a significant difference in the prevalence of age heaping in the northern and southern states. Indices such as Whipple's Index, Modified Whipple's Index, Myer's Index and Age Ratio have been employed for highlighting on the existing statistical discrepancy in the empirical data taken from Census (2011).

Data and methodology

Data Source

The data on single age as well as in age groups of 5 are taken from the Census of 2011. As mentioned earlier, the socio-econonmic and demographic indicators shows a differential between the northern sand southern states (Paul & Sridhar 2015). Thus the study covers Punjab, Rajasthan, Uttar Pradesh and Bihar from northern part of India and Andhra Pradesh, Karnataka, Kerala and Tamil Nadu representing the south India. Comparison has been made between the Northern and Southern states to see how difference in age heaping has affected the level of development in these two parts of India. To assess the quality of age reporting in the northern and southern parts of India the following indices are calculated –

1. Age Ratio (AR)
2. Whipple's Index (W_i)
3. Modified Whipple's Index (W_{tot})
4. Myers blended Index (M_i)

Age Ratio (AR)

It is the ratio of the population in a given age group to half the sum of the population in the preceding and following groups, multiplied by 100. The age ratio is calculated for a five-year age interval by the following formula-

$$_5AR_x = \frac{_5P_x}{\frac{1}{2}(_5P_{x-5} + {_5P_{x+5}})} \times 100$$

Where,

$_5AR_x$ = Age ratio of age group x to x+5

$_5P_x$ = Population in age group x to x+5

$_5P_{x-5}$ = Population in five-year age group preceding the age group x to x+5

$_5P_{x+5}$ = Population in five-year age group following the age group x to x+5

Age ratios should be similar throughout the all age groups and it is expected that the age ratios for all age groups should be about equal to the value of 100. The deviation from 100 reflects the presence of errors in repoting of the age data.

Whipples Index (W_i)

The Whipple's index is essentially designed to detect the concentration or heaping in terminal digits of 0 and 5. It is applicable for data for single ages. The index is also designed over the range of 23-62 inclusive. But however, it is based on the assumption of rectangularity that is all the ages are evenly distributed. Employing the assumption of rectangularity in a 10 year range we may measure the preference over the terminal digit '0' in the range 23 to 62 very roughly by comparing the sum of the population in the ages ending with zero and in this range with one tenth of the total population in the range. Similarly based on the assumption of rectangularity or of linearity in a 5 year range, we may measure heaping on multiples of 5 (terminal digits of 0 and 5 combined) in the range 23 to 62 by comparing the sum of the populations at the ages in this range ending with '0' and '5' and one fifth of the total population.

$$\text{Whipples Index } (W_i) = P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60} / 1/5 * \Sigma(P_{23} - P_{62})$$

Modified Whipples Index (W_{tot})

One of the short comings of the Whipples index is that it only measured the avoidant or preference of only those ages ending with 0 or 5 without any distinction. Thus two changes are made to the formula, the first helping to distinguish the avoidance or preference for ages ending with 0 and that of 5. But however it is based on the unrealistic assumption of linearity over a range of 10 years. The first modification was done to differentiate between preference for reporting age with terminal digit 0 and 5 as follows (Roger et al., 1981, p. 148):

$$W_0 = 10(P_{30}+P_{40}+P_{50}+P_{60}) / (P_{23}+P_{24}+\ldots +P_{61}+P_{62})$$

$$W_5 = 10(P_{25}+P_{35}+P_{45}+P_{55}) / (P_{23}+P_{24}+\ldots +P_{61}+P_{62})$$

If we sum up the two i.e. W_0 and W_5 and divide the sum by 2, the original whipple's index will obtained again

$$W = (W_0+W_5)/2$$

The second change which was put forward by Noumdissi in the year 1992 is based on much more realistic assumption of linearity over a age range of 5 years instead of 10. It is also based on the principles and assumption of the original Whipple index. The original Whipples Index is calculated by the following formula:

$$W = 5 (P_{25} + P_{30} + P_{35} + \ldots\ldots + P_{60}) / (P_{23} + P_{24} + \ldots\ldots + P_{61} + P_{62})$$

P_x denotes the population of age x in completed years.

Age heaping can thus also be calculated for the entire digit starting from 0 to 9 individually.

$$W1 = 5(P_{31}+P_{41}+P_{51}+P_{61}) / ({}_5P_{29}+{}_5P_{39}+{}_5P_{49}+{}_5P_{59})$$

$$W2 = 5(P_{32}+P_{42}+P_{52}+P_{62}) / ({}_5P_{30}+{}_5P_{40}+{}_5P_{50}+{}_5P_{60})$$

$$W3 = 5(P_{23}+P_{33}+P_{43}+P_{53}) / ({}_5P_{21}+{}_5P_{31}+{}_5P_{41}+{}_5P_{51})$$

$$W4 = 5(P_{24}+P_{34}+P_{44}+P_{54}) / ({}_5P_{22}+{}_5P_{32}+{}_5P_{42}+{}_5P_{52})$$

$$W6 = 5(P_{26}+P_{36}+P_{46}+P_{56}) / ({}_5P_{24}+{}_5P_{34}+{}_5P_{44}+{}_5P_{54})$$

$$W7 = 5(P_{27}+P_{37}+P_{47}+P_{57}) / ({}_5P_{25}+{}_5P_{35}+{}_5P_{45}+{}_5P_{55})$$

$$W8 = 5(P_{28}+P_{38}+P_{48}+P_{58}) / ({}_5P_{26}+{}_5P_{36}+{}_5P_{46}+{}_5P_{56})$$

$$W9 = 5(P_{29}+P_{39}+P_{49}+P_{59}) / ({}_5P_{27}+{}_5P_{37}+{}_5P_{47}+{}_5P_{57})$$

If the value of this digit specific modified Whipple's index is equal to 1, it reflects that there is no digit preference or avoidance for that particular digit. In spite of overcoming the drawbacks of the original Whipple's index by extending the principle to the entire 10 digits the modification proposed by Noumdissi (1992) is not practical for any spatial, temporal and clash or Spatio-temporal comparison. Thus this brings out the need for a summary index which will help in establishing the overall spatial and temporal variability of age reporting. This overall index that summarizes all the age preference and avoidance can be calculated by summing up the absolute differences between the digit specific Whipple's index and 1. Thus the total modified Whipple's index can be written as the following 0 indicating no preference or avoidance.

$$W_{tot}=\sum_{i=0}^{9}(|W\text{i} - 1|)$$

If there is no preference for any digit then $W_0=W_1=W_2=............=W_9=1$ and thus

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

On the other hand if all the ages reported ends with either 0 or 5, then W_0 and W_5 will be equal to 5 and all other that is W_i will be 0 ,thus making tbe value of W_{tot} reach its maximum value 16.

Myers index

The Myre's blended index is designed for evaluating single year age data, it can give the extent of digit preferences for all the digits 0,1, 2………...9. It can be used to report errors for all ages 10-99 ages. It is also based on the assumption of rectangularity. If the sum at any given digit exceeds 10% of the total blended population, it indicates over the selection of ages ending with that digit that is the existence of digit preference for that particular digit. Similarly, a negative deviation or sum that is less than 10% of the total blended population indicates an under the selection of the ages ending in that digit that is digit avoidance for that particular digit. Here, we have calculated Myers' blended index for the age range from 23 to 62 years as the same age group applied on Whipple's index (original and W_{tot}) for the comparison purpose of measuring general realiability of age distribution.

Result

State-wise age ratio by sex

Figure 1 depcits the male-female comparison of the age rtios between north and south India for the 5-year age group. The curve is not showing the even distraibution of population, expecting of age ratios of every 5-year age-group close to the value of 100 in both the regions. The similar pattern of age reporting is observed owing to gender affiliation in North as well as in South India. Irrespective of sex, the age ratio curve is more uneven and showing sharp jumps and clustering throughout all ages, while Southern India experiences disorted picture after the middle youth (above 38-42 year age groroup). The more uneven and ceratin jumps in age reporting is the indication of huge diffrences between number of populations in adjacent groups. A peak positive devation is observed in the age group of 58-62 for both the ragions but the North Indian states are experiencing more sharp dip peak compared to Southern states.

The results show that Uttar Pradesh and Bihar form North India, and Karnataka and Andhra Pradesh from South having highest of age ratio value deviated from 100 indicating quite higher misreporting of age both for male and female.

State-wise digit-specific modified Whipple's index by sex

Figure 2 and Figure 3 shows the state-wise digit-specific modified Whipple's index (W_i) and total modified Whipple's index value (W_{tot}) for each sex. India is experiencing a classic pattern of age reporting, and irrespective of the sex, the strong preference for ages ending with digit 0 and 5 is noted for both in northern states and southern states. W_i depicts that the prefernce for age reporting with terminal digit 0 and 5 is much higher in Northern states with minimal diffrences between sex compared to Southern states. Furthermore, Bihar and Rajasthan from Northern India, and Andhra Pradesh and Karnataka from Southern India have higher digit-specific modified Whipple's index value for digit 0 and 5 comapred to other states of India. Moreover, Kerala shows the most promising situation with lowest digit-specific modified Whipple's index value for all the digits.

The total modified Whipple's index (W_{tot}) suggests that, for both male and female values for Kerala are systematically close to zero, whereas Bihar with W_{tot} value of 4.9 for male and 4.0 for female had the highest values followed by Uttar Pradesh, Rajashtan and and Andhra Pradesh (Figure 3). However, there was a distinctive sex pattern of total modified Whipple's index among the states. It is also noted that total modified Whipple's index (W_{tot}) value for male is higher than the female in northern India while in southern India, W_{tot} value for femle had a much higher value of total modified Whipple's index than male indicating disorted age reporting quality.

Comparison between Myers' blended method, the original Whipple's index and modified Whipple's Index

Myers' blended index, the original Whipple's index and the total modified Whipple's index (W_{tot}) are compared to investigate the quality of age reporting in Indian states (Table 1). The index values derived from the Myers' method are shown in the first column of Table 1. The Myers' method is used to acces the tendency preferences or avoidance for each of the ten possible digits, from 0 to 9 and proposes a blended index. It follows the assumpmption that, without any age heaping, the aggregate population of each age of the ten digits (0 to 9) are anticipated to be nearly 10% of the total population. For the comaprion purpose, we have calculated Myers' blended index for the age range from 23 to 62 years in the present study as the same age group applied on Whipple's index (original and W_{tot}). The Myers' blended index for the study population reflects almost uniform values (23.9% to 26.9%) with the minimal diffrences between sexes.

Discussion and conclusion

It is well documented that demographic variables such as age and sex are very significant for socio-economic planning, political, administrative, research and several other purposes to make informed decisions at all levels (Maxwell et al. 1972; Shryock, Siegel & Larmon 1980; Newell 1991, Moultrie et al. 2013). Census is the source of massive data on the size, distribution and characteristics of the population, and it also provides information on the country's housing units (Crayen 2009; Hayes 2014). The accuracy of data on the age-sex structure of the population was evaluated using available techniques. Since the data from census is the basis for any study about the Indian population, their quality should be given due consideration (James & Rajan 2004; Rajan & James 2004; Gerland 2014).

In this study, an attempt was made to measure the variation on the errors in age reporting in the context of gender and region. The indices of digit preference i.e. Myers' blended index, Whipple's index (W_i), and the total modified Whipple's index (W_{tot}) are utilized to investigate the quality of age reporting in Indian states. This study revealed that India is witnessing a classic pattern of age reporting. The most preferred terminal digits while reporting age were '0' and '5' both in northern and southern states of India (Borkotoky & Unisa 2014; Pardeshi 2014). As far as region was concerned, the study found a notable difference in digit preference, while the study found minimal differences owing to gender affiliation in the country. It is depicted that the liking for ages ending with digit 0 and 5 is much higher in northern states with minimal differences between sex compared to southern states (Borkotoky & Unisa 2014).

In comparison to South India, North Indians are poor with the quality of age reporting. A number of determinants such as literacy, household size, degree of interaction with administration, use of calendars, astrology, etc. could contribute to this substantial variation (Bose 1996, 2000; Paul & Sridhar 2015). The strong factors leading to age heaping is literacy. Studies conducted in the past agreed that educated people report more accurate age than their counterparts (A'Hearn, Baten &

Crayen 2009; Pal, Mukhopadhyay & Tewari 2015). Studies are enough which provide conclusive evidences of having strong relationship between high age misreporting and low-literacy and other low-socio-economic status (A'Hearn, Baten & Crayen 2009). This fact may explain of comparatively high digit preference in age reporting in northern states than the southern states. As we know that age data is extensively used to estimate several demographic, economic, and health indicators, it is highly recommended that age of the respondent to be reported more accurately. Studies conducted in India and other Asian and Latin American countries have documented that along with the educational status, rural residents and low socio-economic status are also responsible for the age heaping in the country (Denic, Saadi & Khatib 2004; Baten, Manzel & Stolz 2009; Friesen, Baten & Prayon 2012). It can be manifested that people living in rural areas with low socio-economic status are generally lacking in educational attainment resulting in unavailability of relevant documents such as birth certificates and school certificates for reporting actual age. The recent National Family Health Survey (NFHS-5) fact sheet published in 2021 show that more than two-third of births are not being registered in northern states like Bihar, and hence they do not have the birth certificates (IIPS, 2021). So, the amalgamation of factors like educational attainment, rural residency and low socio-economic status contribute to high age heaping (Fulford 2014; Agrawal & Khanduja 2015).

The study also focuses on the methodological part for understanding the best measure of examining the quality of age data. In comparison of Myers' blended index, the original Whipple's index (W_i) and the total modified Whipple's index (W_{tot}), it can be observed that the age reporting indicated by the original Whipple's index is only partial, since only preference for terminal digit 0 and 5 is taken into account. Yet Myers' blended index and the W_{tot} index, which also measure digit liking for other ages, present that reliability of age reporting quality calculated in this way is much clear than that estimated by the original Whipple's index (W_i). Overall, the original Whipple's index (W_i) only accesses the tendency preferences for terminal digits 0 and 5, whereas the modified total Whipple's index (W_{tot}) considers the preferences and avoidance of all ten digits using Wi indices. Moreover, similar observations can be obtained from W_{tot} as Myers' blended index produces.

In conclusion, it can be said that the study did not find any substantial difference among male and female in age misreporting but preference for age reporting with terminal digit 0 and 5 is much higher in Northern states as compared to Southern states. The modified total Whipple's index (W_{tot}) has been proven a powerful and easiest technique in the study for examining the quality of age data in any country. Our study provides a sense of the difference in the extent of these errors and advocates employing different age data assessment indicator prior to data analysis.

Acknowledgment

None

Disclosure statement

All the authors declare no potential conflict of interest

Funding

None

Reference

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 [JPSS]*, *23*(1), 47-56.

A'Hearn, B., Baten, J., & Crayen, D. (2009). Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, 783-808.

Baten, J., Manzel, K., & Stolz, Y. (2009). *Convergence and Divergence of Numeracy: The Development of Age Heaping in Latin America, 17th to 20th Century*. University of Tuebingen Working Paper.

Borkotoky, K., & Unisa, S. (2014). Indicators to examine quality of large scale survey data: an example through District Level Household and Facility Survey. *PLoS One*, *9*(3), e90113.

Bose, A. (1996). Demographic transition and demographic imbalance in India. *Health Transition Review*, 89-99.

Bose, A. (2000). North-South divide in India's demographic scene. *Economic and Political Weekly*, 1698-1700.

Caswell, H., de Vries, C., Hartemink, N., Roth, G., & van Daalen, S. F. (2018). Age× stage-classified demographic analysis: a comprehensive approach. *Ecological monographs*, *88*(4), 560-584.

Cleland, J. (1996). Demographic data collection in less developed countries 1946–1996. *Population Studies*, *50*(3), 433-450.

Crayen, D. (2009). Quantitative analyses of long-run human capital development: age heaping as an indicator for numeracy.

Denic, S., Saadi, H., & Khatib, F. (2004). Quality of age data in patients from developing countries. *Journal of Public Health*, *26*(2), 168-171.

Diamond, P. A. (1977). A framework for social security analysis. *Journal of public economics*, *8*(3), 275-298.

Franke, R. W., & Chasin, B. H. (1994). *Kerala: Radical reform as development in an Indian state*. Food First Books, Subterranean Company, Box 160, 265 South 5th St., Monroe, OR 97456.

Friesen, J., Baten, J., & Prayon, V. (2012). *Women Count: Gender (in-) equalities in the human capital development in Asia, 1900-60* (No. 29). University of Tübingen working papers in economics and finance.

Fulford, S. (2014). Returns to education in India. *World Development*, *59*, 434-450.

Gerland, P. (2014). UN Population Division's methodology in preparing base population for projections: case study for India. *Asian Population Studies*, *10*(3), 274-303.

Hayes, K. (2014). The Importance of a Census. In *Business Journalism* (pp. 187-192). Apress, Berkeley, CA.

Internaional insitiute for population sciences (IIPS) (2021) Key Indicators - 22 States/UTs from Phase - I : National Family Health Survey (NFHS-5) - 2019-20; http://rchiips.org/NFHS/NFHS-5_FCTS/NFHS-5%20State%20Factsheet%20Compendium_Phase-I.pdf

James, K. S., & Rajan, S. I. (2004). Respondents and quality of survey data. *Economic and Political Weekly*, 659-663.

Jowett, A. J., & Li, Y. Q. (1992). Age—heaping: contrasting patterns from china. *GeoJournal*, *28*(4), 427-442.

Maxwell, M. H., Bleifer, K. H., Franklin, S. S., & Varady, P. D. (1972). Demographic analysis of the study. *Jama*, *220*(9), 1195-1204.

Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K., Timæus, I. M., & Zaba, B. (2013). *Tools for demographic estimation*. International Union for the Scientific Study of Population.

Noumbissi A., 1992, "L'indice de Whipple modifié : une application aux données du Cameroun, de la Suède et de la Belgique», *Population*, 47 (4), pp. 1038-1041.

Newell, C. (1990). *Methods and models in demography*. Guilford Press.

Pal, J. K., Mukhopadhyay, B. K., & Tewari, H. R. (2015). Age reporting error and effect of education: a village study. *American Journal of Social Science Research*, *1*(3), 158-162.

Paul, S., & Sridhar, K. S. (2015). *The paradox of India's north–south divide: Lessons from the states and regions*. SAGE Publications India.

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.

Paul, S., & Sridhar, K. S. (2015). *The paradox of India's north–south divide: Lessons from the states and regions*. SAGE Publications India.

Rajan, S. I., & James, K. S. (2004). Second national family health survey: emerging issues. *Economic and Political Weekly*, 647-651.

Rasul, G., & Sharma, E. (2014). Understanding the poor economic performance of Bihar and Uttar Pradesh, India: a macro-perspective. *Regional Studies, Regional Science*, *1*(1), 221-239.

Register general of India. (2011). Census of India. *Provisional Population Totals. New Delhi: Government of India*.

Roger, G., Waltisperger, D., & Corbille-Guitton, C. (1981). *Les structures par sexe et âge en Afrique*. Groupe de démographie africaine.

Sharma, V. (2015). Are BIMARU states still Bimaru?. *Economic and political weekly*, 58-63.

Shryock, H. S., Siegel, J. S., & Larmon, E. A. (1980). *The methods and materials of demography* (Vol. 2). US Department of Commerce, Bureau of the Census.

Sun, Y., Zhang, M., Sun, Z., & Tan, T. (2017). Demographic analysis from biometric data: Achievements, challenges, and new frontiers. *IEEE transactions on pattern analysis and machine intelligence*, *40*(2), 332-351.

Vaupel, J. W. (1998). Demographic analysis of aging and longevity. *The American Economic Review*, *88*(2), 242-247.

Véron, R. (2001). The “new” Kerala model: Lessons for sustainable development. *World development*, *29*(4), 601-617.

Yadav, A., Vishwakarma, M., & Chauhan, S. (2020). The quality of age data: Comparison between two recent Indian censuses 2001–2011. *Clinical Epidemiology and Global Health*, *8*(2), 371-376.

Figure 1. Age ratio by sex in India, 2011

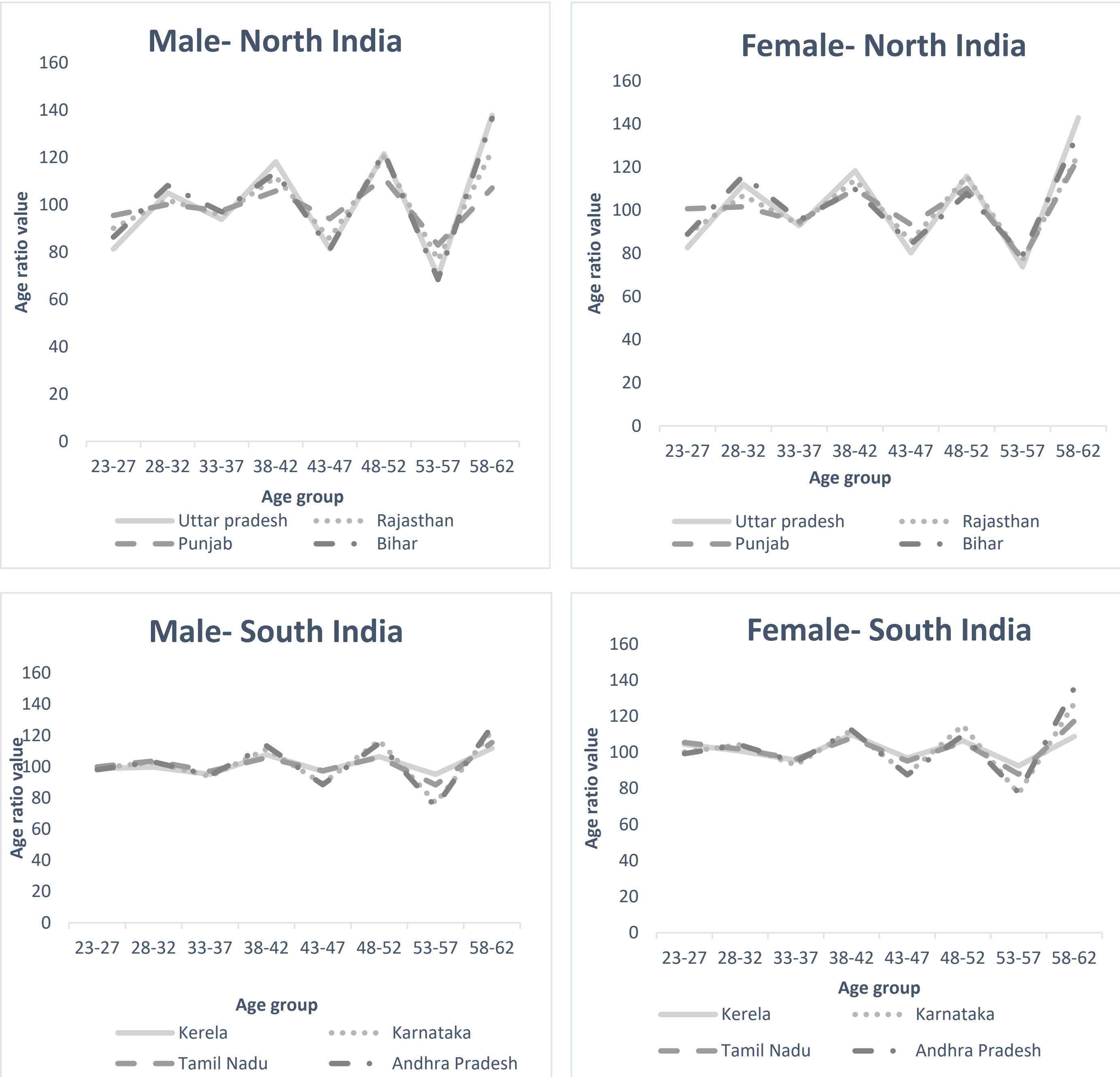


Figure 2. Digit-specific modified Whipple's index by sex in India, 2011

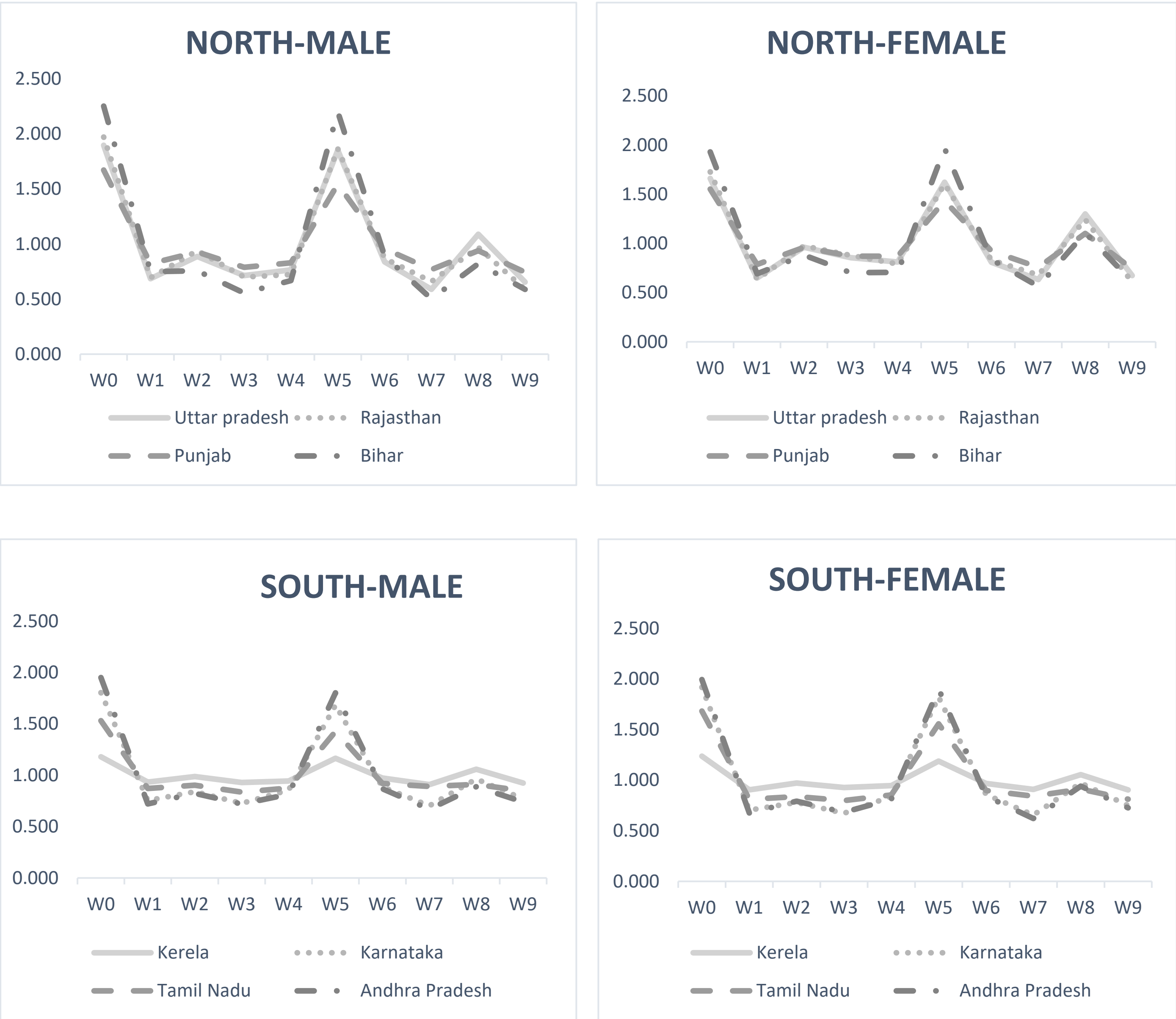
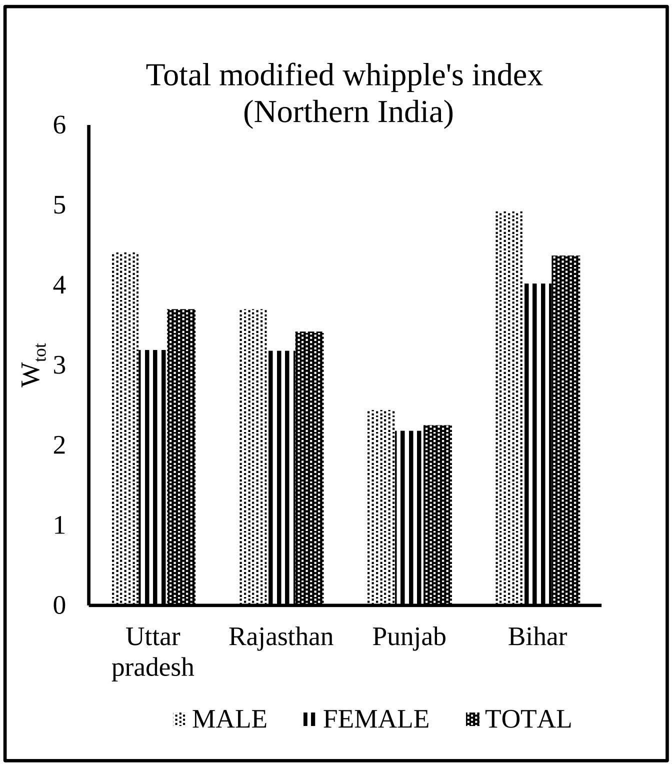


Figure 3. Total modified Whipple's index by sex in India, 2011



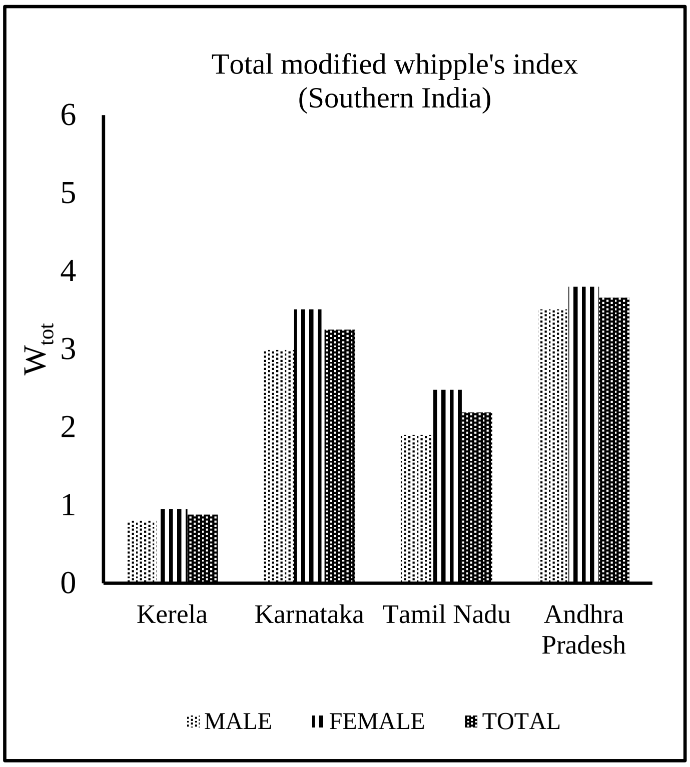


Table 1. Comparison between Myers' blended method, original Whipple's index and modified Whipple's Index, 2011

States		Myers' blended Index (M_i)		Original Whipple's Index (W_i)		Modified Whipple's Index (W_{tot})	
		Male	Female	Male	Female	Male	Female
Northern	Punjab	23.97	26.02	160.61	149.217	2.44	2.18
	Rajasthan	24.51	25.56	191.74	166.53	3.70	3.18
	Uttar Pradesh	25.61	25.69	209.69	164.18	4.41	3.19
	Bihar	26.87	25.69	222.64	194.75	4.92	4.02
Southern	Andhra Pradesh	25.09	24.91	187.60	194.79	3.51	3.80
	Karnataka	24.78	25.22	174.25	187.00	2.99	3.51
	Kerala	26.73	23.27	117.24	121.28	0.80	0.95
	Tamil Nadu	25.28	24.73	147.77	161.81	1.90	2.47
Male/ Female Ratio of values							
Northern	Punjab	0.92		1.08		1.12	
	Rajasthan	0.96		1.15		1.16	
	Uttar Pradesh	1.00		1.28		1.38	
	Bihar	1.05		1.14		1.23	
Southern	Andhra Pradesh	1.01		0.96		0.92	
	Karnataka	0.98		0.93		0.85	
	Kerala	1.15		0.97		0.85	
	Tamil Nadu	1.02		0.91		0.77	