ASSESSMENT OF AGE REPORTING DATA : A PAKISTANI PERSPECTIVE

Muhammad Kashif[1], Khushnoor Khan[2]
[1]Department of Mathematics and Statistics, University of Agriculture, Faisalabad, Pakistan
[2]Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia
Email: mkashif@uaf.edu.pk

ABSTRACT*: *This paper focuses on unearthing the age heaping phenomenon commonly prevalent in our society. The data was taken from two population census (1981, 1998) to study the related phenomenon and the technique applied for the current study was Modified Whipple's Index (MWI) method. The results thus obtained are compared with other summary measures of age reporting, namely- Whipple's Index (WI) and Myer's Blended Index (MBI). It is exhibited from the results that quality of age reporting for 1998 census improved when compared to the 1981 census data. It was witnessed that age heaping is more rampant in women than in men and likewise in rural areas, it is more widespread than the city areas. The results also demonstrated the overall efficiency of (MWI) concerning the quality of age reporting; moreover, it also takes into account the inclination and evading of all ten digits. Hence, it is a just and secure method and comes up with the similar results as MBI. The principal benefit of Modified Whipple's Index stems from the fact that it is simple to calculate and easy to interpret in practical situations.

Keywords: Age heaping, Census data, Modified Whipple's Index, Myer's Blended Index

INTRODUCTION

The population census is no doubt an intricate, extensive process that more often than not is put into operation once in every decade. Even though suitable scientific methods are adhered to for data collection, yet, the responses may turn out to be incorrect. The data contained in the census serves diversified purposes, particularly in carving national policy and conducting demographic research studies. Therefore, the extent of research in these areas and validity of their findings hinges upon the precision of census data. Among the high volume of data collected in Census, Age-sex structure is a fundamental component and plays a significant role in the demographic analysis [1]. Age reporting is a crucial prerequisite for accurate estimates of the fundamental component of population change (i.e. fertility, mortality and migration measures). So misreporting of age severely affects the quality of data [2]. That's why, age reporting error has been a continuing cause of concern for many social scientists, demographers, sociologist and economists in developing countries [3]. It is, therefore, important to examine where in the age distribution and in which population subgroups the misreporting occurs, and with what magnitude.

Data regarding age in countries like Pakistan are prone to errors. A standard procedure in reporting age is the tendency towards rounding the ages to the closest figure ending in '0' or '5' or, for that matter, converting it in an even number.

Several techniques have been developed (mostly by demographers and epidemiologists) to measure the magnitude of errors emanating from digit preference as well as methods of correcting or adjusting the information. Some of the techniques are robust while others are not. Whipple index is a customary measurement in order to evaluate the age heaping. This measure has been widely used to evaluate the worth of age reporting in census and survey data. This index measures only the attraction for age ending in 0 and 5. By modifications to the original Whipple's Index, Spoorenberg [4] proposed a new synthetic index that accounts the attraction/repulsion of all ten digits. This technique is known as the Total Modified Whipple's Index and is founded on the same assumption of the first WI.

The extent of the current study is to observe the trend of digit preference in two censuses of Pakistan (1981, 1998) by applying the methodology as proposed by [4]. To assess this advance method, comparison of the MWI with the other measures will be carried out.

MATERIAL AND METHODS

We studied the single year age data from two censuses reported by Pakistan Census Organization. These censuses were carried out in 1981, and 1998 and the population composition was described according to age, sex and residence status (urban and rural). Three standard indices are used to detect the digit preference, namely, the WI, total modified Whipple's Index (MWI), and MI.

WI is simple to determine and extensively applied [5]. It detects age heaping on terminal digit '0' and '5' within 23 to 62 both years inclusive. The WI is calculated as

$$WI = \frac{\left(\sum_{j=0}^{7} \gamma_{25+j\times 5}\right)\times 100\times 5}{\sum_{i=23}^{62}\gamma_i}, \quad \ldots\ldots\ldots(1)$$

Where γ_x is the population of age x at completed years.

The WI varies from 100 to 500. When avoiding '0' and '5' n WI =100 is used. But when ages end in '0' and '5' are being reported then WI=500 is employed [6].

Fundamental principle behind the calculation of WI is that it considers a continuous and linear decrease in the number of persons within the age range 23 and 62 both inclusive. This linearity assumption is not plausible between other small (0-22 year) and high (63 years and above) age groups that are why they are excluded from the calculation. This calculation of WI gives the preferences for avoidance of ages ending in both '0' and '5' indistinctively. However, opposing effects of digit preferences and avoidances can cancel out each other and affect the original Whipple's Index. To remove this constraint, two modifications have been made to the original formulation of Whipple's index given by [7-8].

Roger [7] proposed first change in original formulation of WI and calculated the two measures of index as

$$W_0 = \frac{\left(\sum_{j=0}^{3} \gamma_{30+j\times 10}\right)\times 10}{\sum_{i=23}^{62} \gamma_i}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

$$W_5 = \frac{\left(\sum_{j=0}^{3} \gamma_{25+j\times 10}\right)\times 10}{\sum_{i=23}^{62} \gamma_i}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(3)$$

By taking an average of (2) and (3), we refer back to the original WI.

This first variation assumes a linear distribution of ages over a ten year age range and presents a mean to differentiate among preferences for ages ending in '0' and those terminating at '5'. The second modification, proposed by [8], states a credible assumption of linearity over an age range of 5 years rather than 10 years.

The following modification is advocated by [8] for the calculation of the two indices.

$$W_0 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{30+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{28+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(4)$$

$$W_5 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{25+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{23+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

The above modifications are based on the same assumptions as of the original Whipple's Index (linearity and rectangularity) and allow to measure age heaping for all terminal digits. The degree of preference or avoidance for each digit, can be established as follows:

$$W_1 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{31+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{29+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

$$W_2 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{32+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{30+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

$$W_3 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{23+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{21+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots(8)$$

$$W_4 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{24+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{22+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

$$W_6 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{26+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{24+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

$$W_7 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{27+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{25+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (11)$$

$$W_8 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{28+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{26+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

$$W_9 = \frac{5\times\left(\sum_{j=0}^{3} \gamma_{29+j\times 10}\right)}{\sum_{i=0}^{3} {}_5\gamma_{27+i\times 10}}, \quad \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

Where γ_x is the population of completed age x and ${}_5\gamma_x$ the population of the age range (x, x+4).

The problem with the extension proposed by [8] is that it is not convenient to compare changes through time and across countries. There is still need for a summary index that computed the variability of age reporting. To overcome this [4] proposed a new synthetic index called the modified total Whipple's Index (W_{tot}). It is estimated as the sum of the absolute difference between the digit particular MWI (W_i) and 1 and summarizes all age preference and avoidance effects. The total MWI (W_{tot}) is written as below:

$$W_{tot} = \sum_{i=0}^{9} \left(|W_i - 1|\right), \quad \ldots\ldots\ldots\ldots\ldots\ldots(14)$$

Where W_i is the digit specific WI for all of the ten digits (0-9) developed by [8]. If no preference observed, then, W_{tot} = 0. The maximum value attained by W_{tot} is 16. Therefore this

index can be used as a universal measure for reporting the quality of age and is a more precise and sensitive measure by and large. Moreover, W_{tot} is also based on the same assumptions as original Whipple's Index.

The other standard index used in this study is MBI. Myer's [9] index measures preference for all terminal digits '0' to '9'. For the calculation of MBI select the age range for which the digital preference has to be measured. It is based on single year age data from 10 to 89 year. Using this range, take the sum of a number of people whose ages end at a particular digit from the population aging in the range 10 and over, and then followed by the population age range 20 and over. Applying weights to each series and then the results are summed to get a blended population. A summary index is then calculated by adding the absolute deviations between the theoretical and aggregate distribution (10%). Range of MBI is from '0' to '90'. Where '0' representing no heaping, and an index of 90 representing heaping of all reported ages at a single digit [10].

DISCUSSION

The extent of digit preference for total, urban and rural population for two censuses separated for males and females are presented in Table 1. From the examination of Table 1, it is noted that two censuses have quite a different pattern in age reporting. Based on United Nations Standard [2], age reporting in 1981 is very poor, and there seems to be a strong inclination of preferring age ending in 0 and 5. The situation in 1998 is less extreme as compared to 1981. This indicates that digit preference declined over time, and overall quality of age reporting has enhanced, which suggests that poor reporting of age in 1981 census was noticed by the technical people of the census organization, and their enumerators were went training to acquire better information while collecting data on age in the next census.

MBI and the total WMI (W_{tot}) both are calculated for the age range 23-62 years for each sex and reported in 3rd, 4th, 5th, and 6th column respectively. They differ almost in the same way; to within a few details confirming the pertinence and validity of the W_{tot} index and this difference is attributed to the assumptions on which the two indices stand. These findings are similar to the above conclusions based on original WI. The three indices reveal the same pattern in the quality of age reporting despite of the fact that they are based on a different set of assumptions

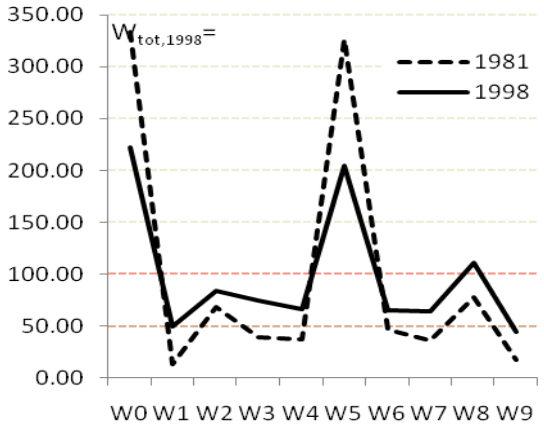
. But the results reported by the original WI are partial because the only digit preference for age ending in '0' and '5' is considered. On the other hand, MBI and total WMI both take into account digit preferences of all other digits. As well as betterment in age reporting quality calculated in this way is evidenced than measured by the original WI. The quality of age reporting improves from 1981. The variation from one census to the next is reported in the middle portion of Table 1, gives a clearer picture.

To evaluate the performance of total WMI, we use it for comparisons between two censuses. The results indicate that male and females have a propensity to misreport their age. In 1981, males had a higher propensity of age heaping than women in urban areas, whereas, the reverse was observed in rural areas. The results of 1998 census indicate that overall females have a higher propensity of age misreporting. But, the variation in W_{tot} values over the period of time shows that overall the quality of age reporting had enhanced both gender and residence-wise.

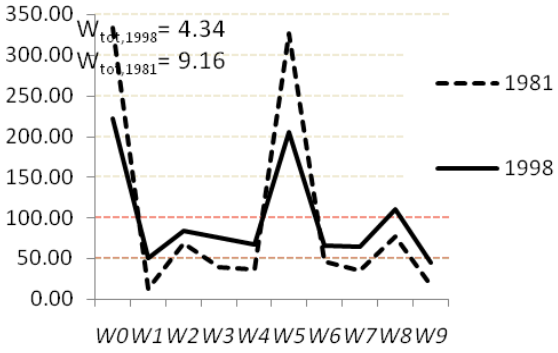
To conclude the discussion we can say that, while the original WI only takes into account preference for ages ending in digits '0' and '5', the MWI (W_{tot}) caters for all ten digits thus utilizing all the information obtained via the specific W_i indices. Additionally, it also gives the same results as MI. WMI main advantage is its simplicity in calculation and its comparability with the original WI. Therefore, by considering all ten digits, W_{tot} index provides a crucial complement to the specific W_i indices and also a correct measure of global age reporting quality.

Figure 1 below provides the preference for each digit as calculated by the digit-specific MWI (W_i) for both censuses. Based on the results presented in figure 1, one can clearly identify that attraction for ages ending with 0 and 5 are the main causes of a low quality of age reporting in 1981. At the same time, because the importance of the attraction on 0 and 5 age digit reduces in 1998, the age reported on the other digit gains significance, explaining the better quality of age reporting.

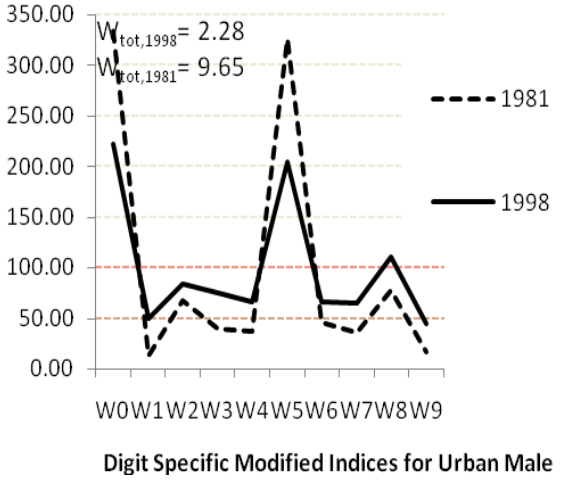
Figure 1. The quality of age reporting by sex: digit –specific modified indices and total modified Whipple's Index in 1981 and 1998 Census in Pakistan.



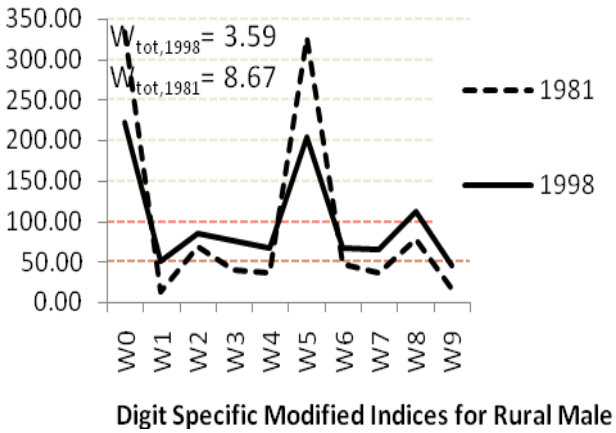
Digit Specific Modified Indices for Male Total



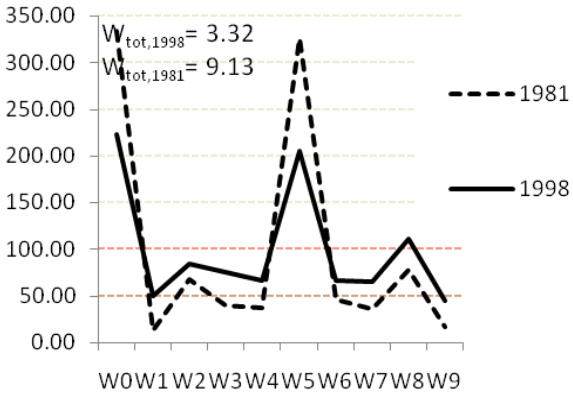
Digit Specific Modified Indices for Female Total



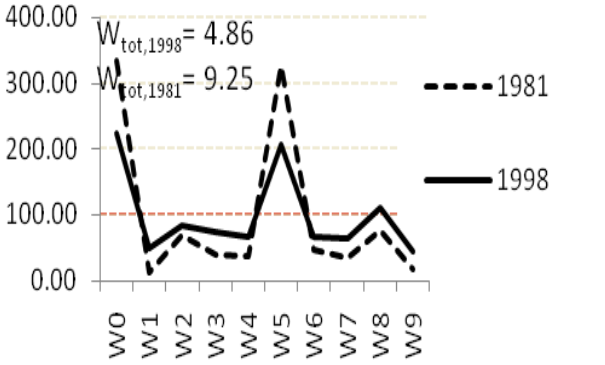
Digit Specific Modified Indices for Urban Male



Digit Specific Modified Indices for Rural Male



Digit Specific Modified Indices for Urban Female



Digit Specific Modified Indices for Rural Female

Table 1: Comparison of the original Whipple's Index, total modified Whipple's Index and Myers' blended index for 1981and 1998 Census of Pakistan

		Original Whipple's Index		Total Modified Whipple's Index		Myer's blended Index	
		Males	Female	Males	Female	Males	Female
1981	Total	3.32	3.27	9.35	9.16	73.54	74.74
	Urban	3.40	3.27	9.65	9.13	77.15	76.48
	Rural	3.14	3.29	8.67	9.25	65.66	70.43
1998	Total	1.72	2.01	3.09	4.34	29.82	38.86
	Urban	1.51	1.76	2.28	3.32	22.22	29.30
	Rural	1.85	2.14	3.59	4.86	34.14	43.77
Variation with respect to Previous Census (%)							
1998	Total	-48.19	-38.53	-66.95	-52.62	-59.45	-48.01
	Urban	-55.59	-46.18	-76.37	-63.64	-71.20	-61.69
	Rural	-41.08	-34.95	-58.59	-47.46	-48.00	-37.85
Male/ Female Ratios of Values							
1981	Total	1.02		1.02		0.98	
	Urban	1.04		1.06		1.01	
	Rural	0.95		0.94		0.93	
1998	Total	0.86		0.71		0.77	
	Urban	0.86		0.69		0.76	
	Rural	0.86		0.74		0.78	

Based on the above findings, it is a concluded that age misreporting is a common problem in Pakistan. The people do not know the importance of their date of birth and population data. To overcome this problem, it is recommended that the age-based question should be improved and repeat with some alteration for cross checking. Better and tactful training are arranged for enumerators, and the importance of census data should be publicized by the governments before the start of the census. The government should motivate and improve the confidence of the people so that they can give correct information during population census.

CONCLUSIONS

Modifications proposed by [4] in the original WI, gives a general measure of age reporting quality. To test its applicability and validity viz-e-viz Pakistani perspective, the modified index is applied to sex-specified reported age data in two censuses of Pakistan conducted in 1981 and 1998. It can be deduced from the analysis that quality of age reporting in Pakistan is poor. The quality of age reporting in 1998 was, however, better than 1981 census data. It is proposed that better-trained enumerators are appointed for interviewing female population. Further, whenever, any data gathering regarding age information takes place; it is recommended to refer to a CNIC and B-form in preference of the person's self-report.

Comparison of results by MBI and the original WI obtained from the same data shows that the total WMI is more sensitive than the original WI. Therefore, we can say that total WMI provides a more accurate measure of age reporting quality and produces results identical to those obtained by MBI. So if one wants to appraise the quality of age reporting with more precision and its changes through time, the total modified WI offers a simple alternative that fully accounts for the changes in the attraction/repulsion of all age digits. The limitation of this study is that we did not have access to the original database from other censuses (conducted in 1951, 1961 and 1972) to calculate the indices in detail and develop a full comparison among all censuses.

REFERENCES

[1] Yazdanparast A, Pourhoseinghole M.A. Abadi A. Digit preference in Iranian age data *Italian Journal of Public Health*.**9** (1):64-68 (2012).

[2] United Nations. Indirect techniques for demographic estimations New York: United Nations Publications. (1995).

[3] Mukhopadhyay, B.K. and P.K. Majundar. A multivariate statistical analysis of reporting error in age data of India *Journal of social sciences* **19**(1):57-61 (2009).

[4] Spoorenberg T. The quality of age reporting: extension and application of the modified Whipple's Index. *Population* (English Edition) **4**(62): 729-41(2007).

[5] United Nations. *Demographic Year Book. Fifty-Sixth* Issue. New York: Department of Economic and Social Affairs, ST/STAT/SER.R35 (2007).

[6] Zeng Y, Vaupel J. Oldest- old mortality in China *Demography Research*. **8**:215-44 (2003).

[7] Roger G., D. Waltisperger; C. Corbille-Guitton. Les Structures par Sexe et Age en Afrique, Paris, Groupe de Demographie Africaine. IDP- INED-INSEE-MINCOOP- ORSTOM, (1981).

[8] Noumbissi, D.C. L'indice de Whipple modifie: une application aux donnees du Cameroun, de la Suede et de la Belgique.*Population* **47**(4):1038- 1041. (1992).

[9] Myers R J. Error and biases in the reporting of ages in census data. *Transactions Actuarial Society America*.**41**:395-415. (1940).

[10] Shryock H, Seigel J. Methods and materials of Demography. Chapter 8. *San Diego: Academic Press.* (1976).