MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2023-043 I November 2023
https://doi.org/10.4054/MPIDR-WP-2023-043

Quality of Reported Ages: A Robust Re-modification in Total Modified Whipple's Index

Afza Rasul I rasul@demogr.mpg.de
Jamal Abdul Nasir
Domantas Jasilionis I jasilionis@demogr.mpg.de
Dmitri A. Jdanov I jdanov@demogr.mpg.de

Quality of Reported Ages: A Robust Re-modification in Total Modified Whipple's Index

Authors:
Afza Rasul[1,2,3], Jamal Abdul Nasir[2], Domantas Jasilionis[3], Dmitri A. Jdanov[3]

[1]Department of Statistics, Lahore College for Women University, Lahore-Pakistan
[2]Department of Statistics, Government College University, Lahore-Pakistan
[3]Laboratory of Demographic Data, Max Planck Institute for Demographic Research, Rostock-Germany

Abstract

In demographic research, th accuracy of the reported ages in surveys and censuses is a persistently important issue. The common indexes developed and used to examine the quality of age data are Whipple, Myers, Bachi's, modified Whipple, and the total modified Whipple's index. The most commonly used and simplest to compute index is the original Whipple's index proposed by George Chandler Whipple. It is a summary measure used to check age heaping on ages ending with digits 0 and 5. The other summary index is the total modified Whipple index (W_{tot}) by Spoorenberg (2007). A re-modification (RW'_{tot}) is proposed for the total modified Whipple index. The new modification, based on the method of the original Whipple index for all digits (0, 1, 2 … 9), is simple, robust, and easy to interpret. The proposed modification is suitable for social, temporal, and spatial comparisons of the quality of self-reported and interviewer-recorded ages at the time of surveys and censuses.

Key Words: Quality of reported ages, Whipple Index, Modified Whipple Index, Total Modified Whipple index

Background

The debate about age misreporting among demographers is not new. A century ago, an American demographer, George Chandler Whipple (1866-1924) gave an index to measure the tendency for human age misreporting. Whipple introduces the index as a ratio of the number of people reporting their age as a multiple of 5 by one-fifth of the total sample/population size in the age range of 23-62 years. It assumes a linear distribution of ages in a 5-year age range. Younger ages (0-22) and older ages (63 and above) are excluded as the linearity assumption is not plausible for these ages. Later other techniques (Bachi, 1951; Myers, 1940) have been developed and used to measure age misreporting in surveys and censuses for age distributions. However, the Whipple index remains the most well-known among all other indices due to its underlying properties and simple calculation. In 1955, the United Nations (UN) developed a criterion based on the Whipple Index which is widely used to check the quality of age data (UnitedNations, 1955) (Table 1).

Demographers have also been trying to introduce new indices to measure accuracy of age reporting. Whipple index follows the assumption of rectangular distribution, which assumes an equal number of persons in each 10-year age group. It identifies the preference (avoidance) of terminal digits of an age ending at ‘0’ and ‘5’. Whipple used the range 23-62 (both inclusive) years of age for the calculation of the quality index for age data. This age range can be changed, but the rectangular distribution and linearity assumptions must be achieved. The attractive benefit of the Whipple index is that it is valid for a single year of age distribution. However, the Whipple index only measures age heaping at two digits, multiple of 5 (ending with 0 or 5), and fails to express age misreporting at other 8 digits.

From Original Whipple to modified versions

Original Whipple’s index is calculated as a ratio of the people reporting their ages as a multiple of 5 (terminal digit 0 or 5) and the persons in the age range 23 and 62.

$$WI = \{5(P_{25} + P_{30} + P_{35} + \ldots + P_{55} + P_{60}) / (P_{23} + P_{24} + P_{25} + \ldots + P_{61} + P_{62})\} * 100 \qquad (1)$$

Where P_x is the population reporting age x in completed years. It assumes a linear distribution of ages in a 5-year age range. In this original version, the Whipple index only checks the preferences (avoidances) at the ages with terminal digits of 0 and 5 without any distinctions. To cover this, the first change was observed in the book (Shryock & Siegel, 1976), which distinguishes the preferences for ages ending with terminal digits 0 and 5.

$$W'_0 = 10(P_{30} + P_{40} + P_{50} + P_{60}) / (P_{23} + P_{24} + P_{25} + \ldots + P_{62}) \qquad (2)$$

$$W'_5 = 10(P_{25} + P_{35} + P_{45} + P_{55}) / (P_{23} + P_{24} + P_{25} + \ldots + P_{62}) \qquad (3)$$

[For ease of presentation multiplier 100 is omitted from equation (2) to onward] By taking the arithmetic mean of equations (2) and (3) we can get equation (1), i.e. original Whipple Index (*WI)*

$$WI = \frac{W'_0 + W'_5}{2} \qquad (5)$$

This modification only provides a meaning to distinguish between the ages ending with terminal digits 0 and 5 and is based on the assumptions of linearity of over 10 years.

The second modification/adjustment proposed/ suggested by Noumbissi (1992), is based on the more sensible assumption of linearity across five years rather than a 10-year age range. Noumbissi feels that the 10-year assumption in the formulation of the Whipple index is crude. Therefore, he suggested a modification in the Whipple index which is based on a more reasonable assumption of linearity taking the age range of 5 years for terminal digits '0' and '5' (Noumbissi, 1992). He suggested the following formulas to calculate heaping at terminal digits '0' and '5':

$$W_0 = 5(P_{30} + P_{40} + P_{50} + P_{60}) / ({}_5P_{28} + {}_5P_{38} + {}_5P_{48} + {}_5P_{58}) \qquad (6)$$

$$W_5 = 5(P_{25} + P_{35} + P_{45} + P_{55}) / ({}_5P_{23} + {}_5P_{33} + {}_5P_{43} + {}_5P_{53}) \qquad (7)$$

Here P_x is the number of individuals reporting their age as x years, and, ${}_5P_x$, is the number of individuals reporting their age in the age range $(x, x+4)$. Similarly, age heaping can be calculated for other terminal digits $(1, 2, \ldots, 9)$ *as:*

$$W_1 = 5(P_{31} + P_{41} + P_{51} + P_{61}) / ({}_5P_{29} + {}_5P_{39} + {}_5P_{49} + {}_5P_{59}) \qquad (8)$$

$$W_2 = 5(P_{32} + P_{42} + P_{52} + P_{62}) / ({}_5P_{30} + {}_5P_{40} + {}_5P_{50} + {}_5P_{60}) \qquad (9)$$

$$W_3 = 5(P_{23} + P_{33} + P_{43} + P_{53}) / ({}_5P_{21} + {}_5P_{31} + {}_5P_{41} + {}_5P_{51}) \qquad (10)$$

$$W_4 = 5(P_{24} + P_{34} + P_{44} + P_{54}) / ({}_5P_{22} + {}_5P_{32} + {}_5P_{42} + {}_5P_{52}) \qquad (11)$$

$$W_6 = 5(P_{26} + P_{36} + P_{46} + P_{56}) / ({}_5P_{24} + {}_5P_{34} + {}_5P_{44} + {}_5P_{54}) \qquad (12)$$

$$W_7 = 5(P_{27} + P_{37} + P_{47} + P_{57}) / ({}_5P_{25} + {}_5P_{35} + {}_5P_{45} + {}_5P_{55}) \qquad (13)$$

$$W_8 = 5(P_{28} + P_{38} + P_{48} + P_{58}) / ({}_5P_{26} + {}_5P_{36} + {}_5P_{46} + {}_5P_{56}) \qquad (14)$$

$$W_9 = 5(P_{29} + P_{39} + P_{49} + P_{59}) / ({}_5P_{27} + {}_5P_{37} + {}_5P_{47} + {}_5P_{57}) \qquad (15)$$

When there is no digit preference/avoidance these digit-specific modified Whipple indices are equal to 1 (Noumbissi, 1992).

The third modification is suggested by Spoorenberg (2007). He felt the extensions proposed by Noumbissi (1992) to all ten digits are not practical for spatial, temporal, and spatiotemporal comparisons. He introduced a summary index based on the modified Whipple index and named it the Total Modified Whipple Index (W_{tot}).

$$W_{tot} = \sum_{i=0}^{9}(|W_i - 1|) \qquad (16)$$

where W_i is the digit specified modified Whipple index for each of the ten digits (0-9) developed by Noumbissi.

If no preference is observed, then

$$W_0 = W_1 = W_2 = W_3 = W_4 = W_5 = W_6 = W_7 = W_8 = W_9 = 1 \text{ and}$$

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

If all reported ages end in 0 and 5, then $W_0 = W_5 = 5$ and all other $W_i = 0$. Hence, W_{tot} reaches the maximum value of 16. This index can be used as a general measure of the quality of age reporting in the complement to Noumbissi's previous development (Spoorenberg & Dutreuilh, 2007).

Re-modification of Total Modified Whipple's Index (RW_{tot}) [Proposed Index]

There are a few shortcomings in the calculation of the equation for individual digits given in equations (6-15) by Noumbissi. If we expand the numerator of the above equations (6-15), the equations will be as follows;

$$W_0 = 5(P_{30} + P_{40} + P_{50} + P_{60}) / \{(P_{28} + P_{29} + P_{30} + P_{31} + P_{32}) + (P_{38} + P_{39} + P_{40} + P_{41} + P_{42}) + (P_{48} + P_{49} + P_{50} + P_{51} + P_{52}) + (P_{58} + P_{59} + P_{60} + P_{61} + P_{62})\} \qquad (6e)$$

$$W_5 = 5(P_{25} + P_{35} + P_{45} + P_{55}) / \{(P_{23} + P_{24} + P_{25} + P_{26} + P_{27}) + (P_{33} + P_{34} + P_{35} + P_{36} + P_{37}) + (P_{43} + P_{44} + P_{45} + P_{46} + P_{47}) + (P_{53} + P_{54} + P_{55} + P_{56} + P_{57})\} \qquad (7e)$$

$$W_1 = 5(P_{31} + P_{41} + P_{51} + P_{61}) / \{(P_{29} + P_{30} + P_{31} + P_{32} + P_{33}) + (P_{39} + P_{40} + P_{41} + P_{42} + P_{43}) + (P_{49} + P_{50} + P_{51} + P_{52} + P_{53}) + (P_{59} + P_{60} + P_{61} + P_{62} + \mathbf{P_{63}})\} \qquad (8e)$$

$$W_2 = 5(P_{32} + P_{42} + P_{52} + P_{62}) / \{(P_{30} + P_{31} + P_{32} + P_{33} + P_{34}) + (P_{40} + P_{41} + P_{42} + P_{43} + P_{44}) + (P_{50} + P_{51} + P_{52} + P_{53} + P_{54}) + (P_{60} + P_{61} + P_{62} \mathbf{+ P_{63} + P_{64}}) \qquad (9e)$$

$$W_3 = 5(P_{23} + P_{33} + P_{43} + P_{53}) / \{(\mathbf{P_{21} + P_{22} +} P_{23} + P_{24} + P_{25}) + (P_{31} + P_{32} + P_{33} + P_{34} + P_{35}) + (P_{41} + P_{42} + P_{43} + P_{44} + P_{45}) + (P_{51} + P_{52} + P_{53} + P_{54} + P_{55})\} \qquad (10e)$$

$$W_4 = 5(P_{24} + P_{34} + P_{44} + P_{54}) / \{(\mathbf{P_{22} +} P_{23} + P_{24} + P_{25} + P_{26}) + (P_{32} + P_{33} + P_{34} + P_{35} + P_{36}) + (P_{42} + P_{43} + P_{44} + P_{45} + P_{46}) + (P_{52} + P_{53} + P_{54} + P_{55} + P_{56})\} \qquad (11e)$$

$$W_6 = 5(P_{26} + P_{36} + P_{46} + P_{56}) / \{(P_{24} + P_{25} + P_{26} + P_{27} + P_{28}) + (P_{34} + P_{35} + P_{36} + P_{37} + P_{38}) + (P_{44} + P_{45} + P_{46} + P_{47} + P_{48}) + (P_{54} + P_{55} + P_{56} + P_{57} + P_{58})\} \quad (12e)$$

$$W_7 = 5(P_{27} + P_{37} + P_{47} + P_{57}) / \{(P_{25} + P_{26} + P_{27} + P_{28} + P_{29}) + (P_{35} + P_{36} + P_{37} + P_{38} + P_{39}) + (P_{45} + P_{46} + P_{47} + P_{48} + P_{49}) + (P_{55} + P_{56} + P_{57} + P_{58} + P_{59})\} \quad (13e)$$

$$W_8 = 5(P_{28} + P_{38} + P_{48} + P_{58}) / \{(P_{26} + P_{27} + P_{28} + P_{29} + P_{30}) + (P_{36} + P_{37} + P_{38} + P_{39} + P_{40}) + (P_{46} + P_{47} + P_{48} + P_{49} + P_{50}) + (P_{56} + P_{57} + P_{58} + P_{59} + P_{60})\} \quad (14e)$$

$$W_9 = 5(P_{29} + P_{39} + P_{49} + P_{59}) / \{(P_{27} + P_{28} + P_{29} + P_{30} + P_{31}) + (P_{37} + P_{38} + P_{39} + P_{40} + P_{41}) + (P_{47} + P_{48} + P_{49} + P_{50} + P_{51}) + (P_{57} + P_{58} + P_{59} + P_{60} + P_{61})\} \quad (15e)$$

It can be observed that all expanded equations (6e-15e) are not based on a complete age range (23-62) years. Moreover, for equations (8e, 9e, 10e, and 11e), few values of the denominator are beyond the age range (23-62). These equations give unrealistic and poor predictions of the indexes for digits (1, 2, 3, and 4). Therefore, there is an obvious shortcoming in the indexes proposed by Noumbissi. The approach to calculating individual digit (0-9) preferences in Noumbissi's version is unrealistic and not based on mathematical formulation. Spoorenberg's total modified Whipple index (W_{tot}) is based on Noumbissi's modification. Therefore, it does not give an appropriate summary value of all terminal digits robustly.

There should be a summary index that should measure age misreporting or heaping at all terminal digits in the data range (23-62) and have no mathematical shortcomings. A new modification in (W_{tot}) is proposed based on the original Whipple index (*WI*). Following the original Whipple index, the Whipple indexes for each terminal digit (0-9) can be calculated as;

$$WI_0 = 10(P_{30} + P_{40} + P_{50} + P_{60}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (16)$$

$$WI_1 = 10(P_{31} + P_{41} + P_{51} + P_{61}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (17)$$

$$WI_2 = 10(P_{32} + P_{42} + P_{52} + P_{62}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (18)$$

$$WI_3 = 10(P_{23} + P_{33} + P_{43} + P_{53}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (19)$$

$$WI_4 = 10(P_{24} + P_{34} + P_{44} + P_{54}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (20)$$

$$WI_5 = 10(P_{25} + P_{35} + P_{45} + P_{55}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (21)$$

$$WI_6 = 10(P_{26} + P_{36} + P_{46} + P_{56}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (22)$$

$$WI_7 = 10(P_{27} + P_{37} + P_{47} + P_{57}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (23)$$

$$WI_8 = 10(P_{28} + P_{38} + P_{48} + P_{58}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (24)$$

$$WI_9 = 10(P_{29} + P_{39} + P_{49} + P_{59}) / (P_{23} + P_{24} + P_{25} + \dots + P_{62}) \quad (26)$$

Based on indexes (WI_0 to WI_9) in the equations (16-26), a modified version of Spoorenberg's total modified Whipple index (W_{tot}) can be calculated as;

$$RW_{tot} = \sum_{i=0}^{9}(|WI_i - 1|) \qquad (27)$$

If there is no heaping at any digits then

$$WI_0 = WI_1 = WI_2 = WI_3 = WI_4 = WI_5 = WI_6 = WI_7 = WI_8 = WI_9 = 1 \text{ and}$$

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

If all reported ages are heaped at a multiple of 10, i.e. ends at digit 0, then $WI_0 = 10$ and all other $WI_i = 0$, and hence, $RW_{tot} = 18$.

A criterion for the quality of age reporting for RW_{tot}

United Nations proposed the following criteria to assess the quality of age reporting based on the Whipple Index (UnitedNations, 1955).

Table 1. UN criteria for Quality of reported ages.

Whipple index Value	Deviation from perfection	Quality of data
<105	<5%	Perfectly Accurate
105-110	5 - 9.99%	fairly Accurate
110-125	10 - 24.99%	Moderate
125-175	25 - 74.99%	Poor/rough
>175	≥ 75%	Very poor/rough

The original Whipple's index is a summary measure of the quality of age data which emphasizes the heaping at digits 0 and 5. It ignores the digit preferences (avoidance) on all other 8 digits i.e. 1, 2, 3, 4, 6, 7 8, and 9. If we go through the parallel to UN criteria for Quality of age reporting, for RW_{tot} the following ranges can be defined as a standard for quality of age reporting (Table 2). To calculate the range for RW_{tot}, the following cases are considered

1. If all ages are correctly reported i.e. $WI_0 = WI_1 = WI_2 = \ldots = WI_9 = 1$, $RW_{tot} = 0$.

2. If all reported ages are heaped at a multiple of 10, i.e. ends at digit 0, $WI_0 = 10$ and all other $WI_i = 0$, $RW_{tot} = 18$. Similarly, if all reported ages ended at any single digit, i.e. any one $WI_i = 10$ and all other $WI_i = 0$, $RW_{tot} = 18$.

3. If any two digits are reported only (say 0 and 5), then $WI_0 = WI_1 = 5$ and all other $W_i = 0$, $RW_{tot} = 16$.

Hence the value of RW_{tot} ranges between 0 and 18. A value of "0" represents the perfection of data. As the value of RW_{tot} departs from 0, it goes towards imperfection of the reported ages which implies that ages are misreported. The following ranges can be used to express the quality of age data based on RW_{tot}.

Table 2. Criteria for Quality of reported ages using RW_{tot}.

RW_{tot} Value	Quality of data	% of People reporting incorrect Age
0.00 - 0.19	Perfectly Accurate	<1%
0.20 - 0.99	fairly Accurate	2 - 4.99%
1.00 – 1.99	Moderate	5 - 9.99%
2.00 - 2.99	Poor/rough	10 – 14.99%
>3	Very poor/rough	≥ 15%

Application to Data

To test the remodified version of the total modified Whipple index (RW_{tot}), we applied it to the Standard Demographic and Health Survey (DHS) data sets series for 3 countries (India, Turkey, and Pakistan). Standards DHS surveys collect primary data using four types of model questionnaires. Among these questionnaires, a household questionnaire is used to collect data on the characteristics of the household for all the usual residents of the household. This is the main questionnaire having all respondents who are used to get information for other types of questionnaires. In this household schedule, information about age, sex, education, relationship to household head, residence, etc. is available. Reported age data from individuals from this household questionnaire is used for this research. Sample size used for this study is given in Table 3(ICF, 1985-2023). More details about the DHS surveys can be found at the DHS website (www.dhsprogram.com).

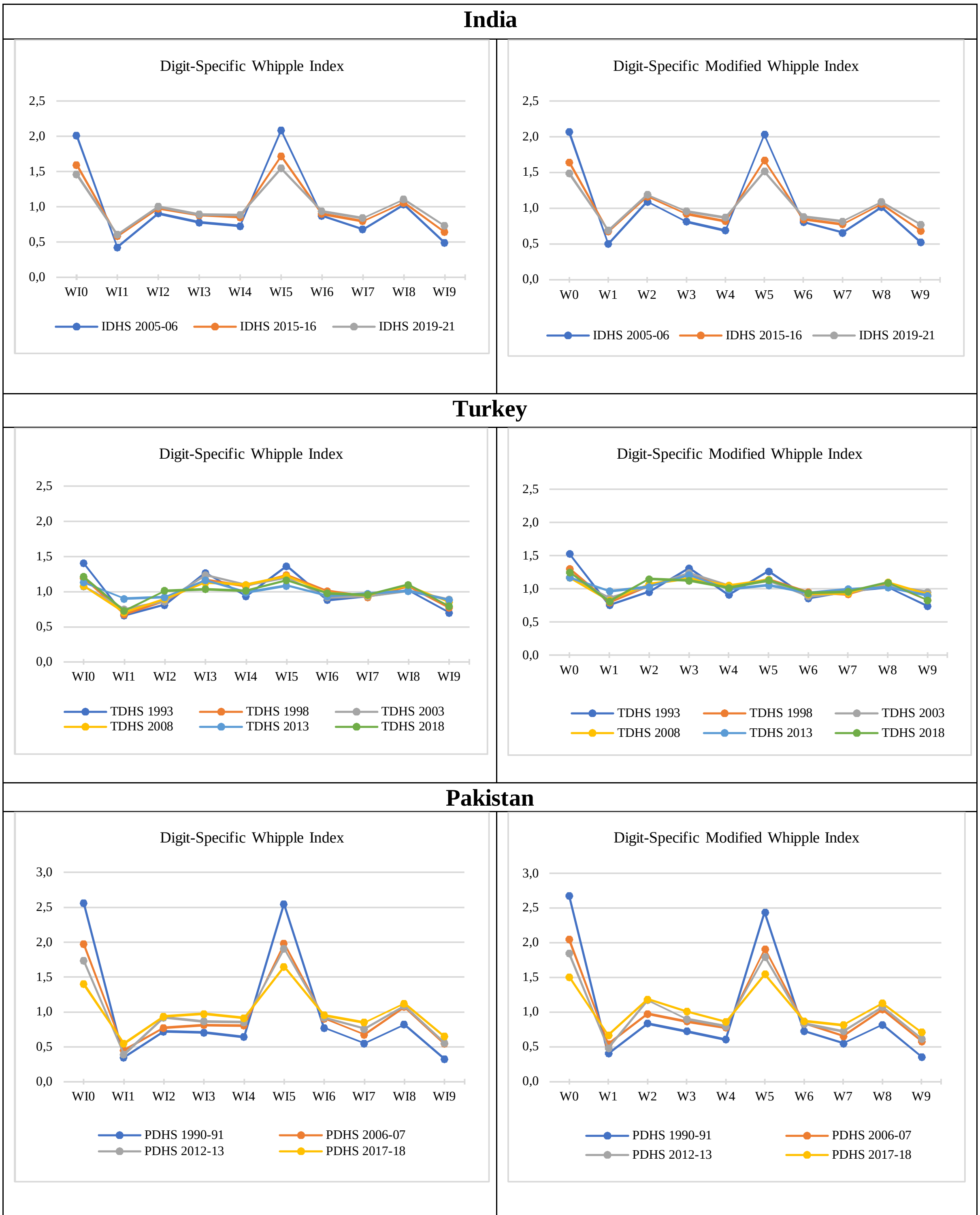
Table 3. Sample size of individual respondent in DHS series of three selected countries

Countries/ years	Total Respondents	Study sample Aged 23-62
India		
2005-06	534161	243520
2015-16	2869043	1380023
2019-21	2843917	1437924
Turkey		
1993	40840	17841
1998	37991	16899
2003	47894	22161
2008	44498	21522
2013	45660	22807
2018	39914	20308
Pakistan		
1990-91	52358	18293
2006-07	727493	264060
2012-13	94169	36647
2017-18	100869	39982
Data Source: Standard DHS of India (2005-2021), Turkey (1993-2018) and Pakistan (1990-2018) Household data files (ICF, 1985-2023)		

Among three selected counties for analysis, quality of age reporting is considered poor in India and Pakistan (Singh et al., 2022), and average in Turkey. Appendix Tables 1A, and 1B give the digit-specific Whipple indexes (*WI_i*) and digit-specific modified Whipple indexes (*W_i*) for India, Turkey, and Pakistan respectively. Figure 1 shows the digit preference (avoidance) using the digit-specific Whipple indexes and modified Whipple indexes. Technically if all digits are correctly reported then the value of all digit-specific indexes must be 1. If the value of an index for any digit is greater than 1, it is the preferred digit by people telling their age at the time of data collection. This may cause heaping at those digits. If the value of the index is less than 1, this shows the avoidance of that digits.

Age reporting in India and Pakistan follows a classical pattern of age misreporting. A strong age heaping at the digits 0 and 5 is reflected in both digits' specific Whipple index as well as in the original Whipple index. Digits 1 and 9 are the most avoided in age reporting in both countries. Preferences for 0 and 5 have declined a little bit over time but the situation of the accurate reporting of age is still looking like a nightmare in both countries. The situation in Turkey is less extreme. Preferences for 0 and 5 declined over time and were very near to 1 in the year 2008 to onwards. In Turkish data sets, digit 3 is as likely to be preferred as 0 or 5. The most observed pattern of digit preference in India is 5, 0, and 8 for digit-specific Whipple's index (*WI_i*) and 5, 0, and 2 for digit-specific Modified Whipple's index (*W_i*). The digit "2" replaces its position from digit "8" due to the overestimation of digit "2" in the digit-specific Modified Whipple's index in all three-survey data sets of India. Similarly, digit "3" also changes its position over digit "4" due to the overestimation of the index for digit 3. Although the digits "1 and 4" have not changed their position these two digits also overestimated the digit in the digit-specific Modified Whipple's index. In the case of Pakistan digit "5" is ranked as the highest preferred digit in digit-specific Whipple indexes, however for the PDHS 2005-06 and PDHS 2012-13 the digit "0" ranked as the highest preferred digit in the digit-specific modified Whipple index. Similarly, digits 2 and 3 ranked at earlier positions due to overestimation in the digit-specific modified Whipple index. In Turkish data sets all overestimated digits "1, 2, 3, and 4" in digit-specific modified Whipple indexes moved to earlier ranks as compared to digit-specific Whipple indexes. In the digit-specific modified Whipple index (Noumbissi's indexes), digit 3 is preferred over digit 5 due to the unstable calculating equation for digit 3. Digits 1 and 9 are the most avoided in all populations (see Figure 1, Appendix A1 and A2).

Figure 1. Quality of age reporting: Digit-specific Original Whipple's index (WI_i) and digit-specific Modified Whipple's index (W_i) for the series of DHS data of India, Turkey, and Pakistan.



Data Source: Standard DHS of India (2005-2021), Turkey (1993-2018) and Pakistan (1990-2018) (ICF, 1985-2023), Figures based on the Author's calculations of Whipple indexes.

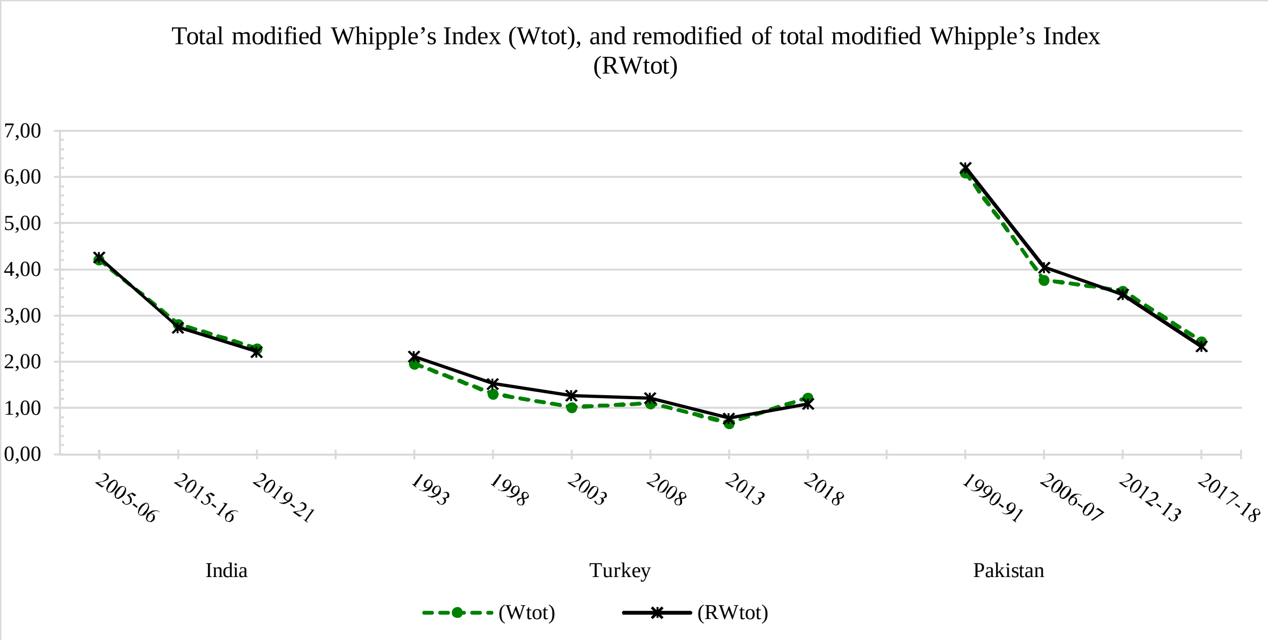
Comparison between Whipple index (*W*), Total modified Whipple's Index (W_{tot}), and remodified total modified Whipple's Index (RW_{tot})

Comparably all three indies give the same patterns over the years and can be used for spatial, temporal, and demographic comparisons within nations and globally. The original Whipple index is a good choice to check the quality of age data when we are interested in checking the heaping at digit 0 or 5 only. The other two indexes will be used to check the overall quality of reported ages. However, digit-specific modified indexes (Noumbissi's indexes) give over-estimated values for digits 1, 2, 3, and 4. Which showed a wrong estimation for preference for the digits 1, 2, 3, and 4. For example; Noumbissi's index for digit 3 in the Turkish data set showed that it is more reported digit than 5. Which is not the actual case. Noumbissi's indexes calculated for digits 1, 2 3, and 4 give the wrong representation of the original reported digit due to inaccuracy of equations (see equations 8-10 or 8e-10e). Total modified Whipple's Index (W_{tot}) based on these impractical equations did not give reliable and accurate information about the quality of age data for individual digits. The new re-modification of the total modified Whipple index is based on mathematical equations (equations 16-26) with the same methodology proposed in the original Whipple index. This modification gives more reliable and robust results based on mathematical modal as compared to the previous proposed by Spoorenberg.

Table 3. Comparison between Whipple index (*W*), Total modified Whipple's Index (W_{tot}) and remodified total modified Whipple's Index (RW'_{tot})

Countries/ years	Original Whipple index (*WI*)	Total modified Whipple's Index (W_{tot})	Remodified of total modified Whipple's Index (Proposed) (RW_{tot})
India			
2005-06	204.92	4.22	4.26
2015-16	165.74	2.82	2.75
2019-21	150.38	2.30	2.22
Turkey			
1993	138.64	1.96	2.12
1998	121.69	1.31	1.53
2003	114.30	1.02	1.27
2008	115.04	1.10	1.22
2013	110.80	0.68	0.78
2018	118.84	1.23	1.09
Pakistan			
1990-91	255.29	6.10	6.21
2006-07	197.58	3.77	4.05
2012-13	181.99	3.54	3.46
2017-18	152.58	2.45	2.34
Data Source: Standard DHS of India (2005-2021), Turkey (1993-2018) and Pakistan (1990-2018) (ICF, 1985-2023)			

Figure 2. Total modified Whipple's Index (W_{tot}), and remodified of total modified Whipple's Index (RW_{tot})



Although the comparison figures of W_{tot} and RW_{tot} both give the same pattern. However, W_{tot} based on Noumbissi's version is impractical for comparison and doesn't give a summary value for any data set. It is also based on unrealistic and non-mathematical formulations and, therefore not robust. RW_{tot} is based on the original Whipple index, and the whole data range thus is preferred over W_{tot}. RW_{tot} gives more precise, robust, and accurate results as compared to W_{tot}.

Simulation study

To evaluate the performance of the proposed modification, we applied it to several simulated data sets for different sizes. Figure 3 gives the digit-specified Whipple and modified Whipple indexes for all ten digits for different series. All series are randomly generated and there is no observed heaping at the digits 0 and 5. For ease of presentation, we compared RW_{tot} with W_{tot} and *WI* respectively. In 1st comparison of RW_{tot} and W_{tot}, consider series S1, where $WI_0 = 1.229$ is most preferred followed by WI_9, WI_8, WI_3, WI_2, WI_1, WI_7, WI_5, WI_6, and $WI_4 = 0.723$ (most avoided digit) while $W_3 = 1.264$ is most preferred followed by W_0, W_2, W_7, W_8, W_9, W_1, W_5, W_6 and $W_4 = 0.843$ (most avoided). Digit 3 ranked up from place 4th to 1st, and digit 2 ranked up from place 5th to 2nd due to over-estimation of digits 2, and 3 in W_{tot} or simply an over-estimation in Noumbissi's indexes. This type of over-estimation in Noumbissi's indexes makes

W_{tot} penniless, and the W_{tot} remains incapable of accessing the quality of age data. Hence, RW_{tot} gives a more accurate and robust estimate for the quality of age distribution.

In all simulated data series, value of the Whipple index (*WI*) is at its perfection level (Table 4). However, RW_{tot} says that data in Series S1 and S2 are moderate, in Series S6 fairly accurate, and in Series S7, S9, and S10 are perfectly accurate. RW_{tot} presents the actual picture of all data series taking all terminal digits. We can observe that in series S1, there is a visible avoidance at digit 4, (WI_4 = 0.723), *WI* ignores it at all. Similarly, in series S2, digit avoidance at digit 3 (WI_3 = 0.741). These preferences and avoidance are accurately covered by RW_{tot} and ignored by *WI*. *WI* emphasizes only heaping at digits 0 and 5 and ignores all other digits. Therefore, if heaping or preferences (avoidance) is present at any(some) other digit(s) (1, 2, 3, 4, 6, 7, 8, 9), *WI* fails to assess the overall quality of age distribution.

Figure 3. Quality of age reporting: Digit-specific Original Whipple's index (WI_i) and digit-specific Modified Whipple's index (W_i) for the simulated data series.

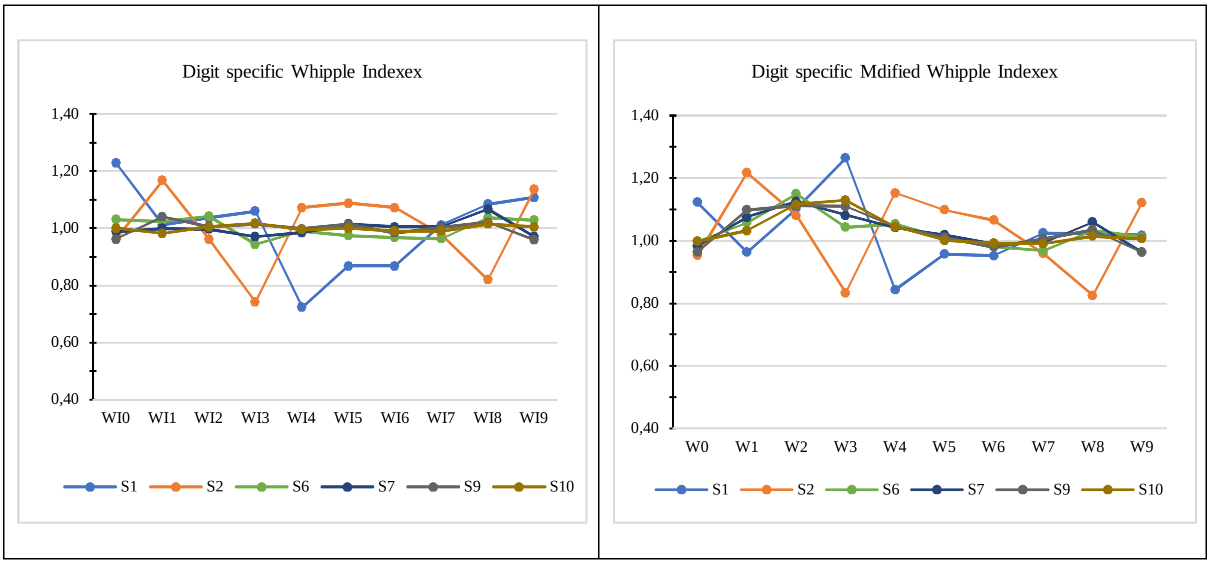


Table 4. Comparison between Whipple index (*W*), Total modified Whipple's Index (W_{tot}) and remodified total modified Whipple's Index (RW'_{tot}) in simulated data series.

Data Series (n)	Original Whipple index (*WI*)	Total modified Whipple's Index (W_{tot})	Remodified of total modified Whipple's Index (Proposed) (RW_{tot})
S1 (415)	104.82	0.839	1.084
S2 (634)	102.52	1.166	1.073
S6 (4144)	100.27	0.408	0.323
S7 (6248)	100.11	0.473	0.185
S9 (21114)	98.78	0.510	0.192
S10 (42030)	100.00	0.365	0.086

Discussion

In the overall summary of all three indexes, the original Whipple index only measures the age heaping for the ages ending at digits 0 and 5. The W_{tot} is based on Noumbissi's indexes, which did not present the actual picture of all digits as indexes for digits (1, 2, 3, and 4) give overestimation. Hence by taking the digit-specific all ten digits from the original Whipple index, the RW_{tot} provides a truer summary measure of the overall quality of age reporting over WI and W_{tot}. The two main qualities of the proposed index RW_{tot} over WI and W_{tot} will make it more useful in all aspects.

1. WI has a limitation, it can capture only heaping at a multiple of five (digits 0 and 5) which is covered by RW_{tot}.

2. RW_{tot} is a more robust method than W_{tot}.

Therefore, to assess the overall quality of age distribution RW_{tot} is a better choice than WI and W_{tot}. The proposed index is completely applicable for spatial, temporal, and sociodemographic comparisons for assessing the quality of age data. Quality of age reporting can also be compared by other demographic characteristics like gender, education, etc. Finally, the proposed index is simple to calculate, fairly constructed on mathematical formulation, and fully based on all age digits. It is practically compatible with the original Whipple's index and total modified Whipple index and provides a more robust measure of the overall quality of age reporting.

The proposed index is not only applicable to age data at the time of interview or data collection, it will be equally applicable to check the digital preference and quality of age in human mortality data. Ongoing demographic and health transitions in Global South calls for better and more comprehensive demographic data covering the entire age span, including working and old ages. Therefore, precise identification of age-reporting problems is crucial for producing reliable population exposures and demographic rates in the region. Further research is needed to develop plausible and transparent adjustment methods allowing solving distortions of survey- or census-based measures.

Appendix

Table A1: Digit-specific modified Whipple's indexes (W_i) and total modified Whipple's indexes (W_{tot}) for India, Turkey and Pakistan.

Country and year(s)	Digit-specific Modified Whipple index W_i										Total modified Whipple index
	W_0	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{tot}
India											
2005-06	2.069	0.499	1.091	0.811	0.688	2.031	0.807	0.658	1.017	0.524	**4.210**
2015-16	1.643	0.673	1.165	0.921	0.820	1.671	0.845	0.777	1.061	0.686	**2.817**
2019-21	1.491	0.690	1.190	0.959	0.871	1.516	0.881	0.815	1.088	0.771	**2.300**
Turkey											
1993	1.529	0.753	0.955	1.312	0.910	1.265	0.858	0.959	1.028	0.741	**1.959**
1998	1.306	0.795	1.042	1.209	1.028	1.142	0.950	0.919	1.079	0.830	**1.311**
2003	1.172	0.862	1.035	1.242	1.048	1.118	0.887	0.944	1.049	0.952	**1.019**
2008	1.168	0.823	1.072	1.161	1.052	1.135	0.920	0.932	1.096	0.913	**1.098**
2013	1.166	0.965	1.031	1.206	0.993	1.053	0.938	0.998	1.023	0.900	**0.685**
2018	1.251	0.811	1.150	1.124	1.017	1.130	0.938	0.958	1.092	0.827	**1.230**
Pakistan											
1990-91	2.677	0.405	0.837	0.725	0.604	2.439	0.724	0.548	0.817	0.356	**6.100**
2006-07	2.048	0.542	0.973	0.867	0.776	1.909	0.831	0.654	1.038	0.579	**3.774**
2012-13	1.847	0.481	1.174	0.900	0.794	1.796	0.832	0.726	1.076	0.615	**3.544**
2017-18	1.504	0.668	1.183	1.008	0.860	1.545	0.870	0.813	1.127	0.711	**2.446**

Data Source: Standard DHS data sets for India (2005-2021), Turkey (1993-2018) and Pakistan(1990-91) (ICF, 1985-2023)

Table A2: Digit-specific Original Whipple's indexes (WI_i) and total Re-modified Whipple's indexes (RW_{tot}) for India, Turkey, and Pakistan.

Country and year(s)	Digit-specific Whipple index WI_i										Total Re-modified Whipple index
	WI_0	WI_1	WI_2	WI_3	WI_4	WI_5	WI_6	WI_7	WI_8	WI_9	RW_{tot}
India											
2005-06	2.012	0.425	0.908	0.777	0.725	2.086	0.870	0.678	1.033	0.486	**4.262**
2015-16	1.596	0.584	0.978	0.882	0.850	1.719	0.899	0.794	1.058	0.641	**2.746**
2019-21	1.461	0.603	1.001	0.891	0.886	1.547	0.936	0.840	1.104	0.731	**2.225**
Turkey											
1993	1.407	0.664	0.813	1.270	0.937	1.365	0.883	0.942	1.018	0.701	**2.122**
1998	1.197	0.678	0.865	1.169	1.079	1.237	1.011	0.921	1.073	0.770	**1.531**
2003	1.079	0.754	0.864	1.236	1.096	1.207	0.919	0.939	1.016	0.891	**1.268**
2008	1.080	0.707	0.898	1.128	1.098	1.221	0.981	0.948	1.085	0.853	**1.225**
2013	1.132	0.902	0.925	1.162	0.989	1.084	0.937	0.976	1.010	0.883	**0.777**
2018	1.213	0.728	1.016	1.036	1.015	1.164	0.978	0.956	1.102	0.792	**1.090**
Pakistan											
1990-91	2.559	0.350	0.720	0.707	0.643	2.546	0.773	0.550	0.822	0.328	**6.215**
2006-07	1.973	0.450	0.773	0.815	0.807	1.978	0.905	0.678	1.075	0.546	**4.054**
2012-13	1.736	0.392	0.922	0.866	0.853	1.904	0.920	0.759	1.090	0.558	**3.460**
2017-18	1.401	0.548	0.935	0.973	0.913	1.651	0.956	0.850	1.123	0.650	**2.349**

Data Source: Standard DHS data sets for India (2005-2021), Turkey (1993-2018) and Pakistan(1990-91) (ICF, 1985-2023)

Table A3: Digit-specific modified Whipple's indexes (W_i), digit-specific Original Whipple's indexes (WI_i), total modified Whipple's indexes (W_{tot}), and Re-modified total Whipple's indexes (RW_{tot}) for simulated data Series.

Data Series	Digit-specific Modified Whipple index W_i										Total modified Whipple index
	W_0	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{tot}
S1	1.123	0.963	1.103	1.264	0.843	0.957	0.952	1.024	1.023	1.018	0.839
S2	0.953	1.217	1.082	0.833	1.153	1.099	1.066	0.960	0.825	1.121	1.166
S3	1.014	1.132	0.953	1.103	1.173	0.904	0.974	0.974	1.102	0.979	0.740
S4	1.050	1.164	1.109	1.072	1.102	0.806	1.201	1.051	0.865	0.950	1.129
S5	0.983	1.116	1.024	1.160	1.071	0.955	1.041	1.015	0.952	1.012	0.548
S6	0.998	1.058	1.150	1.044	1.053	1.007	0.981	0.968	1.032	1.011	0.408
S7	0.983	1.075	1.127	1.081	1.041	1.019	0.990	0.993	1.060	0.964	0.473
S8	0.969	1.091	1.104	1.103	1.014	1.005	1.026	0.992	0.985	1.028	0.426
S9	0.964	1.098	1.112	1.110	1.044	1.011	0.980	1.006	1.035	0.963	0.510
S10	0.999	1.033	1.116	1.130	1.044	1.001	0.992	0.990	1.015	1.007	0.365
Data Series	**Digit-specific Whipple index WI_i**										**Total Re-modified Whipple index**
	WI_0	WI_1	WI_2	WI_3	WI_4	WI_5	WI_6	WI_7	WI_8	WI_9	RW_{tot}
S1	1.229	1.012	1.036	1.060	0.723	0.867	0.867	1.012	1.084	1.108	1.084
S2	0.962	1.167	0.962	0.741	1.073	1.088	1.073	0.978	0.820	1.136	1.073
S3	1.053	1.093	0.869	0.957	1.037	0.869	0.973	0.973	1.140	1.037	0.718
S4	1.058	1.118	1.025	0.936	1.034	0.800	1.179	1.015	0.880	0.955	0.859
S5	0.958	1.050	0.913	1.060	1.034	0.980	1.044	1.012	0.945	1.005	0.410
S6	1.030	1.023	1.042	0.944	0.989	0.975	0.968	0.963	1.038	1.028	0.323
S7	0.988	0.999	0.997	0.970	0.984	1.015	1.005	1.005	1.068	0.970	0.185
S8	0.975	1.045	0.986	0.992	0.958	0.998	1.019	1.001	0.989	1.038	0.204
S9	0.962	1.040	1.005	1.014	0.998	1.014	0.984	1.002	1.021	0.960	0.192
S10	1.000	0.982	1.005	1.018	0.995	1.000	0.990	0.990	1.015	1.005	0.086

References

Bachi, R. (1951). The tendency to round off age returns: measurement and correction. *Bulletin of the International Statistical Institute, 33*(4), 195-222.

ICF. (1985-2023). *Demographic and Health Surveys (various) [Datasets]* Funded by USAID.

Myers, R. J. (1940). Errors and bias in the reporting of ages in census data. *Transactions of the Actuarial Society of America*, *41*(2), 395-415.

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

Shryock, H. S., & Siegel, J. S. (1976). The materials and methods of demography. *J AcademicPress, New York. Shryock*.

Singh, M., Kashyap, G. C., & Bango, M. (2022). Age heaping among individuals in selected South Asian countries: evidence from Demographic and Health Surveys. *Journal of Biosocial Science*, *54*(4), 725-734.

Spoorenberg, T., & Dutreuilh, C. (2007). Quality of age reporting: extension and application of the modified Whipple's index. *Population*, *62*(4), 729-741.

UnitedNations. (1955). *Methods of appraisal of quality of basic data for population estimates, Manual II, Series A, Population Studies No 23* (23). U. Nations.