IJPH - 2012, Volume 9, Number 1

ITALIAN JOURNAL OF PUBLIC HEALTH

Digit preference in Iranian age data

Aida Yazdanparast[1], Mohamad Amin Pourhoseingholi[2], Aliraza Abadi[3]

ABSTRACT

BACKGROUND: Data on age in developing countries are subject to errors, particularly in circumstances where literacy levels are not high. A common error in age reporting is the tendency of rounding the ages to the nearest figure ending in '0' or '5' or to a lesser extent, to the nearest even number. Because of this tendency, commonly known as "digital preference", age heaping occurs at certain ages.
The aim of this study was to study this phenomenon and both Myers' and Whipple's Indexes were employed to identify the digit preference in Iranian national census, 2005.
METHODS: Myers' and Whipple's Indexes were employed to study the pattern of digit preference. The Myers' Blended Index shows heaping at ages ending in 0 and 5 years, and the pattern of heaping is pronounced for both urban and rural populations.
RESULTS: The quality of age reporting for the 2005 census data was poor if compared to the 1995 census data. Digit preference occurred most often in the female population compared to male one, and in rural areas compared to urban ones.
CONCLUSIONS: It can be concluded that both males and females tend to misreport their ages before age 60 especially in rural areas. So, whenever any data gathering regarding age information occurs, the ID card should be used regardless of person's self report.

Key words: Digit preference, Myers' Blended Index, Whipple' s Index, Age data

(1) Department of Statistics, Allameh Tabatabaii University, Tehran, Iran; (2) Research Center for Gastroenterology and Liver Diseases, Shahid Beheshti University of Medical Sciences, Tehran, Iran; (3) Department of Biostatistics Shahid Beheshti University of Medical Sciences, Tehran, Iran

CORRESPONDING AUTHOR: *Mohamad Amin Pourhoseingholi, 7th Floor of Taleghani Hospital, Research Center for Gastroenterology and Liver Diseases.*
Email: aminphg@gmail.com

INTRODUCTION

Age structure is a crucial component in health and demographic analysis as it provides a quick and ready tool for mapping the broad contours of demographic history and makeup of a population. Similarly, the future demographic events are influenced to a large extent by the present sex-age structure, other things being constant (1). Ewbank (1981) discussed at length the effect of age misreporting on the parental survival technique for estimating mortality (2). He did a simulation exercise to demonstrate the effect that age exaggeration has on estimated life expectancy. The results showed that an age exaggeration of approximately 2.5 years will bias the estimated life expectancy upward by approximately the same amount (2). Good age-reporting is a crucial prerequisite for accurate estimates of age-specific fertility rates, which relate births to the age of the mother at the time of birth. If women's ages are misstated, even an accurate enumeration of the total births by each woman will result in distortions in age-specific fertility rates and, if age misreporting is systematically related in any way to marital status and/or parity, there will be systematic biases in fertility estimates (3).

Though, conceptually, the collection of information about age seems to be a simple straightforward task, the fact is that age returns in the censuses were found to be far from the true ages for a large part of the population. Apart from differential under-enumeration in various ages, the age data suffers from distortion owing to preferences for certain ages and digits

due to social, cultural and legal habits, as well as norms observed in a society (1).

A common error when reporting age is the tendency of rounding the ages to the nearest figure ending in '0' or '5', or to a lesser extent, in even numbers. Because of this tendency, commonly known as "digital preference", age heaping occurs at certain ages (4). This error is quite common in many less developed countries (5).

The aim of this study was to chart the occurrence of this phenomenon, employing Myers' and Whipple's Indexes to identify the digit preference in datasets obtained from national census of the Iranian population in 2005, and its comparison to secondary data from previous censuses (censuses that have been done in 1985 and 1995).

METHODS

We studied age data from the national Iranian census conducted by the Statistical Centre of Iran covering all the society's individuals and units. It was carried out in 2005 and the population composition was described according to age, sex, and residence status (urban or rural) in the relevant publication (6).

Two standard indices that were used for this purpose were the Whipple's and Myers' indices. Whipple's Index assumes uniform distribution of population in a five-year range and aims to detect heaping on terminal digits '0' and '5' in the range from 23 to 62 years. This index varies between 100, representing no preference for '0' or '5' and 500, indicating that only ages ending in '0' and '5' were reported (7, 8).

The Whipple's Index is usually calculated as:

$$\text{THE WHIPPLE'S INDEX} = \frac{\left(\sum_{i=0}^{7} N_{25+k*5}\right) * 100 * 5}{\sum_{i=23}^{62} N_i}$$

where Nx is the population of age x in completed years.

The value of the Whipple's Index in a population with perfect age reporting, as well as no large changes in fertility, mortality and migration for a long time, would be 100. The United Nations recommended a standard for measuring age heaping as described in Table 1 (8). The choice of 23 and 62 as the limits of the age band to be examined in the classic Whipple's Index calculation is arbitrary but has been found to be most suitable for the practical purpose of measuring age heaping in general in a population of all ages (3).

TABLE 1

THE UNITED NATIONS' RECOMMENDATION FOR MEASURING AGE HEAPING AS IDENTIFIED BY WHIPPLE'S INDEX.

Whipple's Index	Quality of data	Deviation from
<105	Very accurate	5%
105-110	Relatively Accurate	5-9.99%
110-125	OK	10-24.9%
125-175	Bad	25-74.99%
>175	Very Bad	≥75%

The Myers' Blended Index was developed to detect preference for all terminal digits from 0 to 9. This index is calculated through the following steps:

1. Select the age range for which the digital preference has to be measured. For instance, age 10-89 years.
2. This range is then divided into two overlapping age ranges: 10-89 years, 20-89 years.
3. Population totals are calculated for ages ending in each of the 10 digits and then recorded.
4. Apply weights to each digit selected (weights 1 and 9 for 0 digit, weights 2 and 8 for 1, etc.) and convert the distribution into percent.
5. Find the deviations from 10 percent. The deviations from 10 percent indicate the preference or non-preference of digits.
6. A summary index of deviations for all ages is calculated by dividing the sum of the deviations by 2, or it is one half of the sum of the deviations from 10 percent.

The method yields a reference index for each terminal digit as well as a summary index of preference for terminal digits. The theoretical range of Myers' Blended Index is from 0 to 90. An index of 0 represents no heaping and an index of 90 represents a heaping of all reported ages at a single digit, say five (9, 10).

TABLE 2

THE DISTRIBUTION OF LAST DIGITS FOR THE AGE DATA ACCORDING TO GENDER AND RESIDENCE (IRANIAN CENSUS, 2005 (6)).

Last Digit	% Male	% Female	% Rural	% Urban	% Total
0	11.50	11.93	11.47	12.22	11.71
1	10.15	10.09	10.15	10.06	10.12
2	10.36	10.30	10.31	10.38	10.34
3	10.18	10.08	10.16	10.09	10.13
4	9.77	9.61	9.76	9.56	9.69
5	10.84	10.93	10.78	11.11	10.88
6	9.54	9.43	9.58	9.28	9.48
7	9.59	9.53	9.61	9.44	9.56
8	9.39	9.37	9.41	9.31	9.38
9	8.62	8.69	8.72	8.50	8.65
Total	100	100	100	100	100

RESULTS

Table 2 indicates the percentages of last digits in Iranian age data according to gender and residence, seen in the 2005 census which illustrates up to 11.7% of age distribution for last digit=0 and 10.9% for last digit=5. The minimum percentage is for last digit=9 and maximum belongs to last digit=0.

Figure 1 compares the percentages of last digits between census 2005 and census 1995, indicating that the percentage of ages that ended in 0 and 5 in the 2005 census were higher than those reported in the 1995 collection.

Table 3 shows the degree of digit preference bias that was assessed using a modification of Myers' Index for the whole population separated for male, female and residence, that indicated a higher index for females than males in both urban and rural populations, which means that age was more accurately reported among males than females, respectively 3.06 and 3.35.

FIG. 1

THE DISTRIBUTION OF LAST DIGITS FOR THE AGE DATA FOR IRANIAN CENSUSES, 1995 AND 2005.

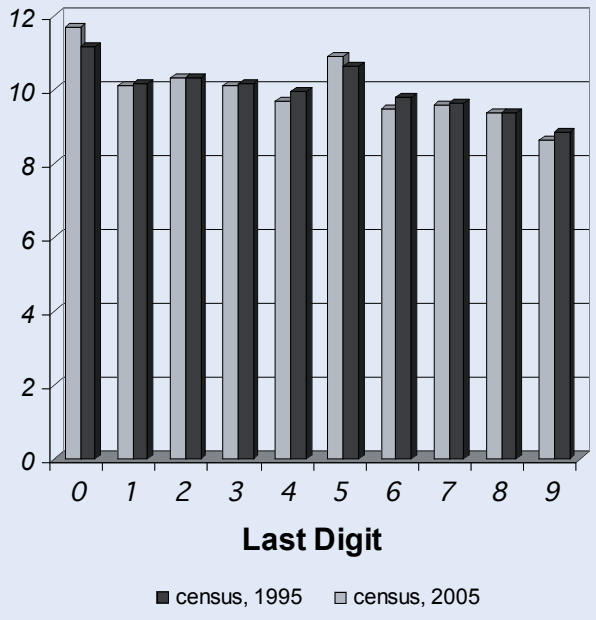

TABLE 3

RESULTS OF MYERS' BLENDED INDEX FOR THE AGE DATA ACCORDING TO GENDER AND RESIDENCE (IRANIAN CENSUS, 2005 (6)).

	Rural	Urban	Total Population
Male	2.73	3.77	3.06
Female	3.04	4.01	3.35
Total	2.88	3.88	3.20

The pattern of heaping is pronounced from age 20 onwards, and this is true for both urban and rural populations. Besides, the total measure of Myers' Blended Index is 3.20 for the Iranian census, 2005.

Table 4 shows Whipple's Index for the whole population separated by male, female

and residence. According to the standard for age measuring, this index for the whole population (111.58) shows that the quality of data is ok, and implies that age reporting is good and more accurate in urban than rural populations.

These results indicate that males have a higher tendency of age heaping than females in rural areas, whilst the reverse was observed in urban areas.

Myers' Blended Index for the total population in the 1995 census was 2.645 whilst it was 3.2036 in the 2005 census implying that age reporting was better in the 1995 census.

DISCUSSION

It can be deduced from the analysis that the quality of age reporting for the 2005 census data was poor when compared to the 1995 census data (Figure1). However, it was of better quality than the 1985 and 1975 census data (6).

This may suggest that both males and females tend to misreport their ages before age 60. Frequently, the elderly population either does not know their age at all or will tend to report their ages in bigger age bands such as 60-70, 70-80 etc. It is possible that the enumerator is often forced to estimate the age of a person based on physical appearance or hearsay in absence of any reliable documents or observance of socio-cultural norms which allow the individual or member of the household to know their ages precisely (1).

There are two other groups for whom recording of age proved rather difficult, women being one of them. Although, frequently, women may be in a position to recall when they were married or when they gave birth to a child, it is difficult for them to state their own date/year of birth unless they are literate. In addition, the assessment of age by an enumerator may also be difficult for young women as, in certain sections of this population or for cultural reasons, they may not be permitted to appear during the enquiry, unless the enumerator is a lady. The other groups, which may suffer from these inaccuracies, are infants and children particularly those not attending school (1).

The preference for these digits among males may be attributed to the greater tendency of overestimating age, whilst, for females, it may be due to an underestimation of their age. This could also be due to the fact that men were often not available at the time of the census and, therefore, female respondents had to report on behalf of men in the census. It is highly likely that the female respondents may not have correctly reported the age of males during the census (11). However, the magnitude of digit preference bias seems to be reducing with the passage of time. This is especially true in the case of females. The possibility of increased female literacy as a factor underlying this reduction is pointed out.

The absence of significant digit preference at ages divisible by five or ten, however, is not necessarily proof of data accuracy since other kinds of errors in age misreporting may also distort the data quality. One way of addressing this issue is to examine the reported population at very old ages relative to the total elderly population (8).

As shown by Coale and Kisker (1986), the proportion of those aged 95 or over among people aged 70 or over in 23 countries with accurate data was always less than six per thousand. Comparatively, this proportion in 28 countries with poor data ranged from one percent to 10 percent (12).

TABLE 4

RESULTS OF WHIPPLE'S INDEX ACCORDING TO GENDER AND RESIDENCE (IRANIAN CENSUS, 2005 (6)).

	Rural		Urban		Total Population	
	Whipple's Index	Quality of Data	Whipple's Index	Quality of Data	Whipple's Index	Quality of Data
Male	115.22	Ok	108.50	Relatively accurate	110.44	Ok
Female	119.16	Ok	105.81	Relatively accurate	112.75	Ok
Total	117.19	Ok	109.26	Relatively accurate	111.58	Ok

The limitation of this study is that we did not have access to the original database from other censuses (conducted in 1975 and 1985) in order to calculate the indexes in detail and develop a full comparison among all censuses, and we only reported the published indexes which were released by the Iranian National Statistics Centre.

In conclusion, both males and females in the Iranian population tend to misreport their ages before age 60 especially in rural areas. So whenever any data gathering regarding age information takes place, it is recommended to refer to an ID card in preference to the person's self report.

ACKNOWLEDGEMENTS: *Authors thanks Dr Ghodratolah Roshanaee for his kind help in data gathering and the reviewers' comments.*

References

(1) Choudhury DR, Deputy Registrar General (C&T). Office of the Registrar General & Census Commissioner, Census of India 2001:1-7.

(2) Ewbank, D. Age Miss reporting and Age-Selective Under enumeration. Patterns and Consequences for Demographic Analysis (Report / Committee on Population and Demography). Washington DC: National Academy Press, 1981.

(3) United Nations. Indirect Techniques for Demographic Estimation. New York: United Nations Publication, 1995.

(4) Pakistan Social and Living Standards Measurement Survey 2004-05, Federal Bureau of Statistics.

(5) Beckett M, DaVanzo J, Sastry N, Panis C, Peterson C. The Quality of Retrospective Reports in the Malaysian Family Life Survey. Santa Monica, California: RAND, 1999:7-10.

(6) Iran Statistical Year Book (1385), 2005. Available from: www.sci.org.ir/portal/faces/public/census85/census85.natayej. [Accessed on june 2011].

(7) Spoorenberg T. Quality of age reporting: extension and application of the modified Whipple's Index. Population (English edition). 2007;4(62):729-41.

(8) Zeng Y, Vaupel J. Oldest-Old Mortality in China. Demography Res 2003,8:215-44.

(9) Myers RJ. Errors and biases in the reporting of ages in census data. Transactions Acturial Soc America 1940;41:395-15.

(10) Shryock H, Siegel J. The Methods and Materials of Demography. Chapter 8. San Diego: Academic Press, 1976.

(11) Naseem I, Gubhaju B, Niyaaz H. Rapid Fertility Decline in the Maldives: An Assessment (Demographers' Notebook).Asia-Pacific Popul J 2004;19:57-75.

(12) Coale A, Kisker EE. Mortality Crossovers: Reality or Bad Data? Population Studies 1986;40:389-401.