Detection of Digit Preference and Age Misreporting by using Demographic Techniques

Rahul Kanti Barua

MPRHGD Program

Department of Social Relations

East West University, Dhaka

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Population, Reproductive Health, Gender and Development (MPRHGD) at East West University, Dhaka, Bangladesh

February 2015

APPROVAL PAGE FOR GRADUATE THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF REQUIREMENTS FOR THE DEGREE OF MASTER OF POPULATION, REPRODUCTIVE HEALTH, GENDER AND DEVELOPMENT (MPRHGD) AT EAST WEST UNIVERSITY (EWU), DHAKA, BANGLADESH

RAHUL KANTI BARUA

STUDENT ID NO: 2010-2-97-002
DEGREE: MPRHGD
DEPARTMENT: Social Relations

Title: Detection of Digit Preference and Age Misreporting by using Demographic Techniques

Thesis Committee
Dr. Rafiqul Huda Chaudhury
Honorary Professor
Adviser & Coordinator, MPRHGD, EWU

.... 14/2/15
Signature & Date

Dr. Lutfun Nahar
Associate Professor
Department of Social Relations, EWU

.... 14/2/15
Signature & Date

DECLARATION

I declare that this thesis entitled "Detection of Digit Preference and Age Misreporting by using Demographic Techniques" is my own work and all sources that I have used or quoted have been indicated and acknowledged by means of complete references. Any mistakes or inaccuracy is my own.

Signature.................................. **Date**....08.02.2015....

(Rahul Kanti Barua)

DEDICATION

I would like to dedicate this thesis paper to my beloved parents, mother in law, my wife and son, whose continuous support and encouragement helped me to keep patience to complete this research work.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisor Dr. Rafiqul Huda Chaudhury, Honorary Professor, Adviser and Coordinator, MPRHGD Program, East West University for his cordial guidance, important suggestion and inspiration.

I also convey my gratitude to my respectable teacher Dr. Lutfun Nahar, Associate Professor, Department of Social Relations, East West University for her observation and constructive suggestions to improve the quality of this thesis preparation and writing.

I am grateful to Bijoy Prasad Barua, *Ph.D.*, Professor, Department of Social Relations, East West University for his inspiration and leadership to complete this thesis.

I am extremely grateful to Dr. Kim Streadfueld from International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B) for providing me single year, age and sex data from Matlab HDSS area population in the year of 2012 for my thesis work.

I am thankful to my two friends, Dr. Abdur Rahim and Dr. Avik Barua, for extending their cooperation by providing me with outpatients' age and sex data from the Chittagong Medical College Hospital for my research.

I also like to extend my special thanks to MPRHGD colleagues, East West University for their continuous encouragement, useful advice and suggestions throughout the thesis proposal, data collection, analysis and writing period.

Finally, I would like to express my heartiest thanks to my wife Papry Barua for her assistance of data entry and continuous encouragement of completing my thesis.

Nevertheless, my heartiest thanks to Mr. Timathy White and Dr. Sumana Das Gupta, who have finally reviewed my thesis and gave their valuable suggestion and input to improve the write-up.

ABSTRACT

Age misreporting and digit preference[1] is very common error all over the world. In developing countries this error is more critical than the developed countries (Talib et al., 2003). The age misreporting and digit preference can be found frequently in population census, survey or other epidemiological and clinical studies. There are many studies in the area of age and sex data. However, this research is unique in the sense as no previous study attempted to evaluate simultaneously the age and sex data of medical outpatients, particularly of Chittagong Medical College (CMC) in Bangladesh and that of the population of Health and Demographic Surveillance System (HDSS) in Matlab.

The rationale for using two data sets of the population in this research is to compare the quality of age collected by two sources- one with a long history of proper data record keeping and verification and the other has no such history. There is no proper record of keeping and verification system of age and sex in outpatients of CMC Hospital. However, in Matlab HDSS, there is a systematic data management system from its establishment in 1966. Since then, the Matlab HDSS has maintained the registration of births, deaths, and migrations, in addition to carrying out periodical censuses. A group of trained staff were involved in data collection and periodic visit of all households in the area. This data is then maintained continuously in a database system. It is expected that age and sex data collected by Matlab HDSS would be of better quality than that of the data collected by CMC Hospital as the former collects these data more rigorously and meticulously than the later. This hypothesis is also supported by data.

We know that demographic data are usually classified by age and sex, and errors in age reporting are more frequent than in sex reporting. Different demographic techniques were used to evaluate both the data sets. These techniques include Whipple's Index (WI) that indicates age data which show systematic heaping for particular ages such as those ending with '0' and '5', Myer's Blended İndex (MI) that measures the extend of digit preference for all the digits, and Age-sex Accuracy Index that determines the accuracy of age reporting.

[1] 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. This tendency is commonly known as "digital preference".

The result of Whipple's Index has shown very rough age data reporting for both male and female outpatients in CMC Hospital, but age data, as expected, have shown accurate reporting for both male and female in the population of Matlab HDSS area. For Myer's Index, about 64 percent of male outpatients and 72 percent of female outpatients in the CMC Hospital reported their ages with incorrect digits. The most preferred digits are '0' and '5', while the most avoided digits by both sex are '9' and '1'. But in case of population in Matlab HDSS area, there is no digit preference and this holds for both male and female.

Likewise, the calculated age and sex accuracy index for outpatients of CMC Hospital is 188, which classified the age data as highly inaccurate, according to the United Nations (UN) scaling and it is about more than four times higher than the population of the Matlab HDSS area (41.1). The female age data is better than male outpatients and the respective sum of the absolute deviation of both male and female age ratio from the unity (100) is 245.42 and 239.76 respectively. The fluctuation for male outpatients found in the age group between 15-19 to 35-39 and for female age ratios is in the upper age group, that is, from 35-39 to 60-64 years.

The evaluation of the CMC Hospital outpatients' age data using demographic techniques, revealed inaccuracy of the age data as indicated by systematic age heaping and digit preference in the CMC Hospital data. On the other hand, age reported by the population of Matlab HDSS are found accurate using the same demographic techniques including UN scaling. Therefore, from this study, we are recommending to adopt Matlab HDSS model and scale up all over Bangladesh, where the individual household profile will be maintained by the trained staff and recorded in a systematic database system.

Table of Content

List of Tables: **Page**

List of Figures:

ACRONYMS

AAI	Age Accuracy Index
ARS	Age Ratio Scores
CMC	Chittagong Medical College
DLHS	District Level Household and Facility Survey
EP	Equal Proportions
FMP	Final Menstrual Period
HDSS	Health and Demographic Surveillance System
ICDDR,B	International Centre for Diarrhoeal Disease Research, Bangladesh
MI	Myer's Blended Index
MPRHGD	Masters of Population Reproductive Health, Gender and Development
SWAN	Study of Women's Health across the Nation
UN	United Nation
US	United State
WI	Whipple's Index

CHAPTER ONE

Introduction, Objective, Methodology and Review of Literature

1.1 Introduction:

Age and sex are the two central features of the population that received the most attention in demographic analysis (Poston, 2006). As sex is a personal characteristic of a person, usually information on sex can normally be collected with less difficulty. The age-sex structure of a population is important both for demographic analysis and developmental planning. Research on such population dynamics, as fertility, mortality, migration, nuptiality and so on, utilize the age-sex distribution as indispensable tools (Mba, 2014). This study looked at age, as it is an important study variable in demography, epidemiological and other clinical studies. It is a socio-demographic variable related to the host of descriptive studies and also a commonly assessed risk factor in analytical studies (Pardeshi, 2010).

Epidemiology has been described as the study of the distribution of health and diseases in a group of people and the study of the factors that influence this distribution (Smoller, 2004). From this, it is clear that all the bio-statistical and epidemiological studies must require accurate age grouping of patients (Johannes and Polly, 1970). In clinical studies age is one of the most commonly assessed variables, and other parameters are often analyzed and the results interpreted in relation to age. In the decision making process at the level of the physician and patient encounter, the accurate age of the patient is important because many decisions are age sensitive and misrepresentation of age may lead to inappropriate action. For example, a screening mammography is recommended to start at age 50 and patient's inaccurate age may result in having either an unnecessary or delayed test (Denic et al., 2003). Therefore, modern health care services might require accurate age and sex data for appropriate decisions, allocation of staff, disease surveillance and in the provision of required drugs and medications (Bello, 2012).

The importance of accurate age and sex data in demographic analysis cannot be overemphasized. National development plans for the provision of such need as housing, food, education, health, employment, manpower, and so on, depend on the relevant socio-demographic statistics classified by age and sex. Most population analysis like fertility and mortality are either age-sex dependent or age-sex selective (Lerche, 1983; Kpedekpo, 1982). The quality of age data is important because age-sex distribution is not only an invariable part of the survey or interview report, but also the bias introduced in studies that can lead to wrong inferences (Bello, 2012). Demographic data are usually classified by age and sex, although, errors in age reporting are more frequent than in sex reporting.

1.2 Rationale of Study:

Age misreporting is common in demographic studies (Shryock et al., 1976), but the prevalence and magnitude of age misreporting in the outpatients of the CMC Hospital were unknown. The most common phenomenon among the irregularities is the age heaping. Age data frequently display excess frequencies at round or preferred ages, such as even numbers and multiples of 5, leading to age heaping. Age heaping is considered to be a measure of data quality and consistency (Pardeshi, 2010).

In Nigeria, age heaping is one of the irregularities in census or survey reporting of age (Bello, 2012). In the United States inaccurately stated age by non-Hispanic whites were found 5 percent of the Medicare population (Kestenbaum, 1992). Among older African Americans, 37 percent misreported their ages (Elo et al., 1996). During community survey, 42 percent population misreported their ages in the Yavatmal District of Maharashtra, India (Pardeshi, 2010). Although age misreporting is a common phenomenon in population census or surveys, this misreporting has a huge implications for good planning. For example, a large number of population reporting at ages 60, 65 and 70 in a census could lead Government to allocate fund for geriatric care. This may turn out to be a bad allocation of fund for health care of a group of elderly people if they are not actually found as their numbers were based on misreporting of ages. Likewise, a screening mammography is recommended to start age 50 and patient's inaccurate age may result in having either unnecessary or delayed test (Denic et al., 2003). Thus, accurate age of patients is indispensable for proper disease diagnosis, disease surveillance and the correct provision of drugs and medications (Bello, 2012). Moreover, misreporting of patients' age may result in unnecessary or additional cost to the family, due to improper diagnosis and wrong treatment, which may have adverse consequences on family income and livelihood. This research will sensitize policy makers, medical professionals, and other stakeholders about the impact of inaccurate age data on human lives and how we can address this situation by way of collecting accurate age data presenting home based evidence.

This study investigates and measures the age misreporting and digit preference of the outpatients in Chittagong Medical College (CMC) Hospital in Bangladesh and the population of Matlab Health and Demographic Surveillance System (HDSS) area. The reasons of using above two data sets of the population in this research is because in the outpatients of CMC Hospital data, there is no proper record keeping and verification system of age and sex data. Moreover, a non-trained staff is being used for keeping records. However, in Matlab HDSS, there is a systematic data collection and management system. Since 1966, Matlab HDSS has maintained the registration of births, deaths, and migrations, in addition to carrying out periodical censuses (International Centre for Diarrhoeal Disease Research, Bangladesh, 2014). A group of trained staff was involved in data collection and periodic visit of all households in the area. This data was recorded in a systematic order and a database was created.

1.3 Objectives of the study:

The objectives of this study are:

i. To evaluate the level of accuracy of age and sex data of CMC Hospital outpatients and Matlab HDSS population
ii. To investigate the digit preferences of male and female in both CMC Hospital outpatients and Matlab HDSS population
iii. To measure the age heaping in CMC Hospital outpatients and Matlab HDSS population

1.4 Hypothesis:

The accuracy of age data is expected to be considerably better in Matlab HDSS population than in CMC Hospital outpatients'. This is expected in view of the fact that there is no systematic process of record keeping and verification of outpatients' age and sex data in CMC Hospital. Besides there is no proper orientation or guidance for the staff about the importance of accurate age and sex data. Whereas, in Matlab HDSS, there is a systematic data collection, record keeping and verification system. For the last 50 years, Matlab HDSS has maintained the registration of births, deaths, and migrations, and also carrying out periodical censuses (International Centre for Diarrhoeal Disease Research, Bangladesh, 2014). In Matlab, a group of trained staff were involved in data collection and periodic visit of all households in the area and maintaining a proper data collection and keeping records in a systematic database. Therefore, we expect to find quality age data better in Matlab population than that in CMC Hospital outpatients.

1.5 Methodology:

This paper intends to apply the demographic techniques of Whipple's Index (WI), Myer's Blended Index (MI) of digit preference and UN Age-sex Accuracy Index to evaluate the age and sex data of outpatients of the Chittagong Medical College (CMC) Hospital, 2013, and the population of Matlab HDSS, 2012.

1.6 Operational definitions of the Demographic techniques:

In this part operational definition of demographic techniques that were used to assess the quality of age and sex data in the thesis paper is described below.

Whipple's Index:

Whipple's Index is applicable where age is reported in single-years. It gives the relative preference for digits '0' and '5', while reports age in the interval 23 and 62 years (Kpedekpo, 1982). It is computed as:

$$Whipple'index = \frac{\sum(P_{25} + P_{30} + P_{35} + \ldots + P_{60})}{\frac{1}{5}\sum(P_{23} + P_{24} + P_{25} + \ldots + P_{62})} \times 100 \longrightarrow (1)$$

To evaluate ages ending with '0', that is 30, 40, 50 and 60, the index is calculated as:

$$Whipple's\ index = \frac{\sum(P_{30} + P_{40} + P_{50} + P_{60})}{\frac{1}{5}\sum(P_{23} + P_{24} + P_{25} + \ldots + P_{62})} \times 100 \longrightarrow (2)$$

To evaluate ages ending with '5', that is, 25, 35, 45 and 55, the index is calculated as:

$$Whipple's\ index = \frac{\sum(P_{25} + P_{35} + P_{45} + P_{55})}{\frac{1}{5}\sum(P_{23} + P_{24} + P_{25} + \ldots + P_{62})} \times 100 \longrightarrow (3)$$

If there is no heaping at age reporting ending with '0' and '5', the index will have a value of 100. If there is complete heaping, the index will have a value of 500. Between these extremes, the following scale for estimating the reliability of the data is used (Kpedekpo, 1982):

Quality of Data	Whipple's Index
Highly Accurate	Less than 105
Fairly Accurate	105 – 109.5
Approximate	110 – 124.5
Rough	125 – 174.9
Very Rough	175 +

Myer's Blended Index

This index is used for evaluating single-year age-sex data. It gives the extent of digit preference for all digits 0, 1, 2, 3,…, 9. It can be used to report errors for all ages 10 – 89 years (Kpedekpo, 1982). The underlying assumption of this method is that, in the absence of systematic irregularities in the reporting of age, the blended sum at each terminal digit should be approximately equal to 10 percent of the total blended population. If the sum at any given digit exceeds 10 percent of the total blended population, it indicates over selection of ages ending in that digit (digit preference). On the other hand, a negative deviation (or sum that is less than 10 percent of the total blended population) indicates under-selection of the ages ending in that digit (digit avoidance). If age heaping is non-existent, the index would be approximately 0 (Kpedekpo, 1982). The procedure for computation is as follows:

1) Sum all the population ending in each terminal digit over the whole range from the ages 10 – 89;
2) Sum all the population ending in each terminal digit over the whole range from the ages 20 – 89;
3) Multiply the sums of ages at each terminal digit in (1) above by co-efficient 1, 2, 3, 4, 5, 6, 7, 8, 9, 10;
4) Multiply the sums of ages at each terminal digit in (2) above by co-efficient 9, 8, 7, 6, 5, 4, 3, 2, 1, 0;
5) Add the product of (3) and (4) above to obtain the blended sum at each terminal digit;
6) Add up the blended sum in (5) above;
7) Find the percentage of the blended sum at each terminal digit to the total of the blended sum;
8) Find the deviation of the percentage distribution from 10

Age-Sex Accuracy Index (Joint Score)

This index measures the level of quality of age-sex population data. It employs the age ratios and the sex ratios simultaneously (Kpedekpo, 1982), and computed as:

Joint score = 3 × (sex ratio score) + (male and female age ratio scores) ⟶ (4)

Age Ratios

Age ratio is usually defined as the ratio of the population in the given age group to one half of the population in the two adjacent groups. Mathematically, let $_5P_x$ be the age group from age X to age X+5, $_5P_{x-5}$ and $_xP_{x+5}$ be the preceding and the following age groups respectively, then,

$$Age\ ratio = \frac{_5P_X}{\frac{1}{2}(_5P_{X-5} + _5P_{X-5})} \times 100 \longrightarrow (5)$$

The computed age ratio is then compared with the expected value, which is usually 100. The discrepancy at each age group is a measure of net age misreporting. An overall measure of the accuracy of an age distribution, called an age accuracy index, is derived by taking the average deviation (regardless of the sign) from 100.0 of the age ratios and summing over all the age groups. The lower this index, the more accurate the data on age. The age ratios are usually calculated for males and females separately and can be calculated for each age group (except the youngest and the oldest) provided the intervals are equal. An age ratio under 100 implies either that members of the age group were selectively under enumerated or that errors in age reporting resulted in misclassifying persons who belong to the age group. A ratio of more than 100 suggests the opposite of one or the other or both of these conditions. Generally, age ratios should be studied in a series of age groups, preferably for the entire span of age for which they can be calculated (Kpedekpo, 1982).

1.7 Data source and limitation of the study:

The primary data used in this research has been collected from hospital record register book 2013 of the Chittagong Medical College Hospital in Bangladesh and Matlab HDSS area population in 2012. Despite the risk of misclassification of age data in medical studies, analysis of the prevalence and magnitude of age misreporting in medical studies is rarely available in Bangladesh.

1.5 Reviews of literature

This section has reviewed the literature of various research studies and reports on age misreporting and digit preferences that are published in different national and international journals. Moreover, relevant research reports are also studied. Besides, Bangladesh Census reports, Bangladesh Bureau of Statistics (BBS), Matlab population data, different survey reports on age misreporting and digit preferences were also reviewed.

This research is a replica of the similar study on "Error Detection in Outpatients' Age Data Using Demographic Techniques" conducted by Yousuf Bellow (2012), where he evaluated accuracy of age reporting by the outpatients in General Hospital Dutsin-ma, Katsina state, Nigeria, using demographic techniques, which includes Whipple's Index, Myer's Blended Index and UN Age-sex Accuracy Index. In his research, he has shown very rough age data reporting for both male and female outpatients. For the Myer's Index, about 86 percent of male outpatients and 88 percent of female outpatients reported their ages with incorrect digits. The most preferred digits he has found are '5' and '0', while the most avoided digit by both sexes was '1'. Similarly, Whipple's Index has also found very rough results in his study and identified considerable age heaping ending with '5' and '0'. Furthermore, UN scaling results were also found to be inaccurate where more critical inaccuracy was observed in female than male age data.

Pardeshi (2010) assessed the quality of age data collected during a community survey in the district of Yavatmal, Maharashtra in India. By applying Whipple's Index, Myers' Blended Index, Age Ratio Scores (ARS) and Age Accuracy Index (AAI) in measuring the age heaping and digit preference, he found very poor quality of age data collected in the survey. He has found 42 percent of the population reported incorrect digit and the preference terminal digit were '0' and '5'. Pardeshi (2010) proposed to apply the innovative methods in data collection along with measuring and minimizing errors using statistical techniques should be used to ensure accuracy of age data.

Another study by Borkotoky and Unisa (2014) examined the quality data on large scale survey data. In their study, age misreporting observed and it differs by region to region and individual characteristics. They also identify some major factors that is, illiteracy, rural residence, poor economic conditions that are associated with age misreporting. The study concludes that "age misreporting, inconsistency and incomplete response are three sources of error that need to be considered before drawing conclusion from any survey".

Denic, Khatib and Saadi (2004) describe that age misreporting is common in demographic studies. In their study, they analyzed the quality of age data in the single-year age distribution and terminal digit preference in cancer patients from developing countries. The study found low quality age data in cancer patients from the Indian subcontinent and Middle Eastern countries. Additionally, it was found that females of all nationalities had a better quality of age data than males and the preference for age ending with digits '0' and '5' was found in all populations. By this study, they concluded that the quality of age data is poor in the developing countries.

Mason and Cope (1987) analyzed the characteristics of people with discrepancies between reported age and reported date of birth or with missing data in their study on 'sources of age and date-of-birth misreporting in the 1900 U.S. census. They concluded that four sources of age data could be attributed to age misreporting in any census or survey. These four sources are, firstly, ignorance of the actual ages among respondents; secondly, miscommunication between interviewers and informants; thirdly, the distortion of age to meet preconceptions or social norms about the relationship of age to other social characteristics or events; and finally, errors in recording or processing.

Ueda (1980) in a similar study reported that one of the major types of errors most commonly found in the sex-age data derived from censuses or surveys is the false reporting of age. In many cases, the erroneous reporting of age is attributable to the ignorance of respondents. In most cases, ages are being reported on some particular digits, "0" and "5".

Talib et al. (2007) in their study "Age Reporting Behavior, A Case Study of 1991 and 2000 Population and Housing Censuses, Malaysia" used a combination of methods, that is, Pyramid Chart, Whipple-type Index and test differences of the terminal digits to analyze the quality of single years of age population. The study found that misstatements in age reporting were due to digit preferred and digit avoided in both censuses.

Mukherjee and Mukhopadhyay (1988) analyzing the Turkish Census 1945, 1955, and 1970, found sever misreporting in age for both male and female and heaping at age ending with the terminal digit "0" and "5". Similarly, Kabir and Chowdhury (1981) examined the population census of Bangladesh, where their analysis found errors in age reporting and it was due to digit preference. Moreover, there was strong tendency to report age digit ending with "0" and "5", with subsidiary heaping at ages ending with "8" and "2" respectively. Bailey and Makannah (1991) evaluated the quality of age and sex data for the successive census 1963, 1974 and 1985 for Sierra Leone. In their study, they found inaccurate age data where the tendency was age ending in the digit zero than digit 5; digit 8 was also popularly reported in the age data. This study also found that the most avoided digit was 1.

Edmonston and Bairagi (1981) in their study, among Bengali populations in Bangladesh, found digital preference more manifest with adolescence and early adulthood, 0 and 5 were the most frequent ending digit. This age misreporting was found critical for females than male. In their study, a tendency was to deflate the ages of girls until after menstruation and marriage, when their ages are adjusting to those more suitable for married women. Young unmarried females were heaped at 10-12; young married at 15. The Single year age data may be improved by analytical approaches such as standard smoothing techniques or developing alternative age groupings to minimize the distortion, or by more careful and sensitive data collection techniques.

In a similar type of study, Nagi, Stockwell and Snavley (1973) revealed that age heaping is a major source of inaccuracy in the age statistics in many of the developing nations in the African continent, particularly among Islamic populations. This phenomenon was found to be more pronounced among women than men, and it tends to increase with age. However, Bairagi, et al. (1982) points out that misreporting also occurred in the early age of the population especially in the rural area. The misstatement for young children in rural Bangladesh increase monotonically with age and systematic errors in age misstatement displays modest overstated for the first years of life and more pronounced understatement for ages 4, 5 and 6.

Some researchers found that higher tendency of age heaping occurs in the older age category of the population. Hill et al. (1997) in their study found that the age misreporting remains considerably high for older African American. Aimee and Samuel (1991) have also found misreporting is most severe at an older age. They found evidence of very pervasive overstatement of age at advanced ages. In their study, the evidence of increasing age misstatement with old age was consistent with the observation that literacy rates also declined with age. However, their findings revealed that the age misstatement is associated with illiteracy and low educational attainment.

Age at marriage especially in the developing countries also contributes to age misreporting. In some developing countries, marriage at a very young age still exists. Indonesia is one example of a country characterized by relatively young age at marriage for females (Savitridina, 1997). Interviewers have some motivation to shift the age of women who are within the boundaries of the 15 to 59 interval to be below or above the minimum age of respondent eligibility. There may also be some shifting of birth to be outside the maximum age of eligibility for the health questions (Pullum, 2005).

Gunasekera (1984) assessed age data in the 1981 census in Sri Lanka and compared to 1971 census. In his study, a remarkable improvement in accuracy between the two censuses were found were reasonably accurate age data in the last census. The improvement was greater in rural areas and females.

Omondi-Odhiambo (1985) in a study on "Age estimation in clinical and public health research, explores problems associated with age reporting in Africa" conclude that two causes of poor age reporting were identified. First one is ignorance, especially found in rural communities and secondly is deliberate falsification, more prevalent in urban communities. The terms age heaping, age preference, and a digital preference were common among men than women. To improve age reporting this study suggested to use the calendar of Historical Events Method, which relates the birth of an individual to some notable event, and the Family Relationships Estimation Method, which pieces together age estimates for all members of a family or household. This method requires some knowledge of marriage patterns in African society such as average age at marriage and birth intervals. The remaining methods are the Individual Maturation Estimation Method (using events such as puberty, marriage.

circumcision and 1st birth as milestone events) and the Community Comparative Estimation Method, which compares data from an individual with that of another in the community.

Locker and Mason (2006) reported that digit preference is intended for health care and health service management. The digit preference bias affects the recording of the time patients arrive and leave emergency departments. Their result found that some departments show the considerable digit preference (0 and 5) bias in the recording of the time of departure from the emergency department. Such bias may cause difficulty in assessing changes in the performance of departments.

Thavarajah, White and Mansoor (2003) assessing the traditional hypertension blood pressure methodology found that the traditional blood pressure (BP) methodology is subject to observer error such as terminal digit preference and single number preference leading to inaccuracies in measurement. A high percentage (60-90%) of terminal BP readings digit being zero has been reported from general medical and hospital-based clinics.

Considering the above literature review, it is observed that although age misreporting and digit preference is an old issue, but it still exists in population census, survey, in medical or any clinical enrollment, even at birth and death certificates and also in social security records. So, in this thesis, the findings are not different rather this study validated the result of the earlier studies. However, this study has created an opportunity to demonstrate that data quality including age data can be improved even in developing countries like Bangladesh using innovative methods and strategies as are employed by Matlab HDSS.

1.9 Scope of this study:

It is hoped that the findings of this study may prompt the Government, NGO or private clinics to conduct further studies in this field. This research also can help the policy makers, researchers, and civil society to find out alternative methods or corrective measures to minimize the misreporting of age data that can ensure the age and sex data quality for the betterment of accurate age structure associated with correct diseases diagnostics including measuring the reproductive age of women, or identify the age burden diseases for the future generation of Bangladesh. This study also re-emphasizes the birth registration as mandatory and advocate to create demographic profile or individual database profile that will contribute to fulfilling the dream of our digital Bangladesh.

CHAPTER TWO

Age and Sex data of Outpatients of CMC Hospital

This chapter mainly shows age and sex distribution of outpatients at Chittagong Medical College (CMC) Hospital and assesses its quality by applying different demographic techniques; namely Whipple Index (WI), Mayer's Blended Index (MI) and UN joint ratio scores.

Table 1 presents the data on age and sex distribution of outpatients in Chittagong Medical College (CMC) Hospital for the month of January 2013 that were collected from the hospital record register office. .

Table 1: Age and Sex Distributions of Outpatients in CMC Hospital, January 2013

Outpatients age (Years)	Sex of outpatients		
	Male	Female	Total
0-4	6	3	9
5-9	5	2	7
10-14	17	9	26
15-19	33	14	47
20-24	71	26	97
25-29	77	49	126
30-34	56	62	118
35-39	75	65	140
40-44	74	88	162
45-49	58	75	133
50-54	67	72	139
55-59	57	44	101
60-64	54	24	78
65+	63	23	86
Total	713	556	1269

2.1 Whipple's Index:

Whipple's Index (WI) is applied to measure the level of heaping at ages ending with digit '0' and '5'. Therefore, Tables 2 (a) and 2 (b) presents the age distribtutions between 23 to 62 years for the male and female outpatients who visited the Chittagong Medical College (CMC) Hospital in January 2013.

From the analysis of data in Table 2 (a), the WI for heaping at male ages ending with '0' and '5' is 246.3, and the quality of the age data is considered to be very rough. In that regards, WI for each of the digits ('0' and '5') is then calculated separately to determine the magnitude of error for each one of them. The WI for ages ending with '0' is 133.4 that qualified the age data rough and then the WI for ages ending with '5' is 113.0 that improved the quality of the age data from rough to approximate data.

Table 2: Age Distribution of Male and Female outpatients of CMC Hospital

Table 2 (a): Age Distribution of Male outpatients of CMC Hospital

Age	Number of patients	Age	Number of patients
23	14	25	20
24	11	30	24
25-29	77	35	33
30-34	56	40	42
35-39	75	45	31
40-44	74	50	41
45-49	58	55	37
50-54	67	60	36
55-59	57	Total	264
60	36		
61	1		
62	10		
Total	536		

Table 2 (b): Age Distribution of Female outpatients of CMC Hospital

Age	Number of patients	Age	Number of patients
23	5	25	11
24	9	30	35
25-29	49	35	34
30-34	62	40	52
35-39	65	45	48
40-44	88	50	57
45-49	75	55	30
50-54	72	60	23
55-59	44	Total	290
60	23		
61	0		
62	0		
Total	492		

The analysis of data in Table 2 (b) suggests that the WI for heaping at female ages ending with '0' and '5' is 294.7, and the quality of the age data is also considered very rough. The WI for ages ending with '0' is 169.7, which again qualified the age data is very rough. The WI for ages ending with '5' is 125.0, which is only a slight improvement that qualified the age data from very rough to rough data.

For comparing both male and female outpatients' age data, female, age heaping is greater than the male outpatients in both '0' and '5'. Thus, the analysis of age data revealed that female outpatients are generally misreporting the age compared to the male outpatients. It has also found that for both male and female outpatients, they preferred '0' digit than '5'.

Furthermore, the results of above WI are also validated by the findings in Figure 1. The Figure I shows percentage of preference for digits ending with '0' and '5' increases with the increase of age except age group 40-49 of digit '5' in male outpatients and the highest age heaping is found in the age group 60-69 years. On the other hand, the percentage of digit preference ending with '0' and '5' increases with the increase of aging except age group 50-59 of digit '5' and the highest preference is in the age group 60-69. It is also found from the Figure 1 that for both male and female outpatients, the percentage of digit preference ending with '0' and '5' in the age group 20-29 almost double in the age group 30-39 where it is 5 times more in the '0' digit for the female outpatient age group.

Figure 1: Percentage of Heaping of digit ending with '0' and '5' of CMC Hospital outpatients

1(a). Age distribution of Male outpatients **1 (b). Age distribution of Female outpatients**

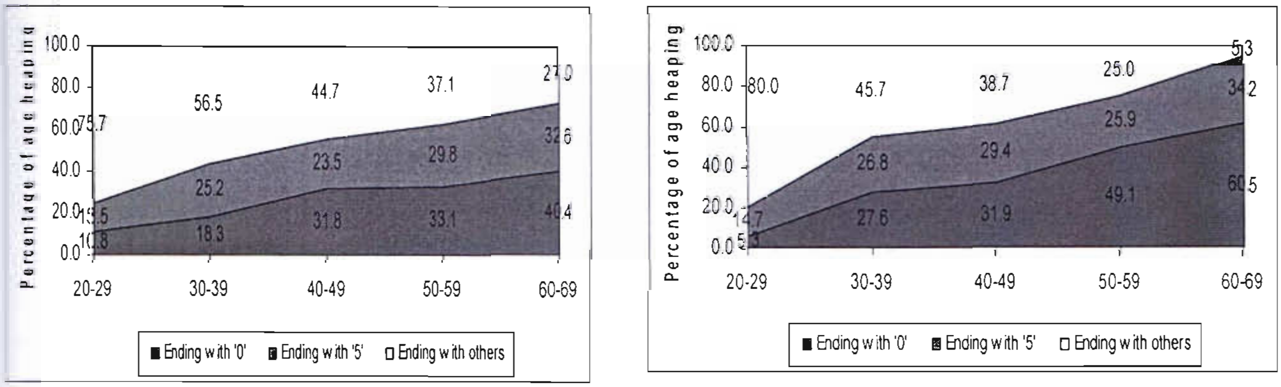
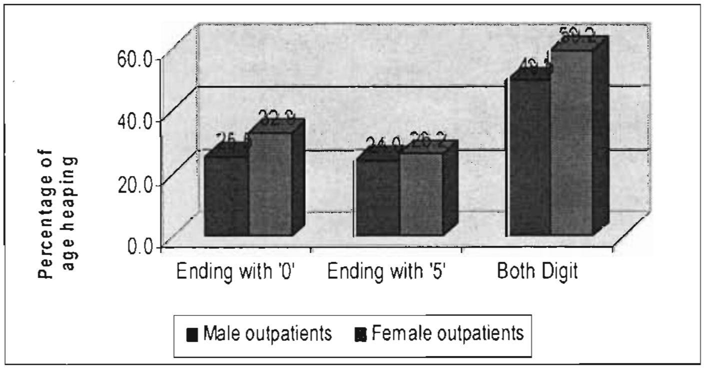


Figure 2: Percent of age heaping of male and female outpatients of CMC Hospital

Again, from Figure 2, it has been shown that almost 50 percent male are heaping their age and it is almost 60 percent of female outpatients. Thus, the female is higher in age heaping than male outpatients.



2.2 Myer's Blended Index

Myer's Blended Index of digit preference is used for evaluating single-year age-sex data by giving the extend of digit preference for all the digits 0, 1, 2,..., 9. The computation of Myer's Index for male outpatients in January 2013 of the Chittagong Medical College (CMC) Hospital is presented in Table 3.

Table 3: Myer's Blended Index for Male outpatients of CMC Hospital

Terminal Digits	Sum of age 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended sum	% Distribution	Deviation from 10
0	183	1	183	179	9	1611	1794	26.41	16.41
1	30	2	60	27	8	216	276	4.06	-5.94
2	81	3	243	77	7	539	782	11.51	1.51
3	36	4	144	36	6	216	360	5.30	-4.70
4	32	5	160	26	5	130	290	4.27	-5.73
5	158	6	948	153	4	612	1560	22.96	12.96
6	35	7	245	32	3	96	341	5.02	-4.98
7	44	8	352	36	2	72	424	6.24	-3.76
8	76	9	684	62	1	62	746	10.98	0.98
9	22	10	220	19	0	0	220	3.24	-6.76
Sum	697			647			6793	100.00	63.74

From Table 3, the results reveal the over selection of ages ending with digits '0' and '5' with the respective preferences of 26.4 percent and 23.0 percent. However, the ages ending with '9' have the highest avoidance, followed by ages ending with '1' and '4'. The next respective avoidance digits are '6', '3' and '7'.

Table 4 shows the computation of Myer's Index for female outpatients. The result has also shown the over selection of ages ending with digits '0' and '5' with the respective preferences which are 31.5 percent and 24.6 percent. Similarly, the ages ending with '9' have the highest avoidance digit, followed by ages ending with '1' and '3'. The other respective avoidance digit numbers are '4', '7', and '6'

Table 4: **Myer's Blended Index for Female Outpatients of CMC Hospital**

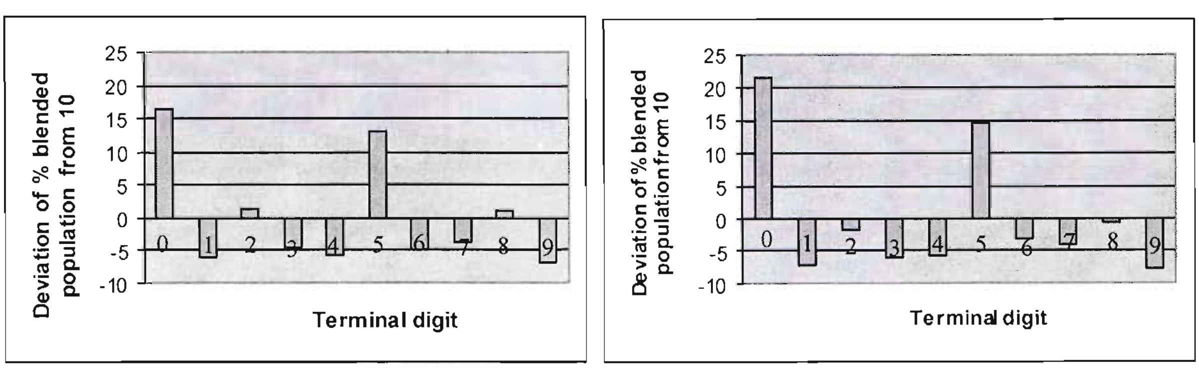
Terminal Digits	Sum of age 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended sum	% Distribution	Deviation from 10
0	177	1	177	175	9	1575	1752	31.53	21.53
1	19	2	38	16	8	128	166	2.99	-7.01
2	46	3	138	45	7	315	453	8.15	-1.85
3	22	4	88	22	6	132	220	3.96	-6.04
4	25	5	125	23	5	115	240	4.32	-5.68
5	137	6	822	136	4	544	1366	24.59	14.59
6	39	7	273	34	3	102	375	6.75	-3.25
7	34	8	272	32	2	64	336	6.05	-3.95
8	52	9	468	50	1	50	518	9.32	-0.68
9	13	10	130	9	0	0	130	2.34	-7.66
Sum	564			542			5556	100.00	72.24

By analysis from Tables 3 and 4. the results indicated that, about 63.4 percent of male outpatients and 72.2 percent of female outpatients reported ages with incorrect final digits. Female outpatients are the higher misreporting than male outpatients. Moreover, for both male and female outpatients, the digit '0' is higher preference than digit preference '5'.

Figures 3 (a) and 3 (b) describe the deviations of the percentage of the blended population of 10 among each of the terminal digits. The most preferred terminal digits while reporting ages were '0' and '5' for both male and female outpatients. For both male and female, '0' digit preference is higher than digit preference '5'. It is also evident that female digit preference is higher in both cases, that is, digit '0' and '5', than male.

Figure 3: Myer's Blended Index for Male and Female outpatients of CMC Hospital

Figure 3(a): Myer's Blended Index for Male outpatients **Figure 3(b): Myer's Blended Index for Female outpatients**



2.3 Age-Sex Accuracy Index (Joint Score)

Table 5 shows the computation of age-sex accuracy index for male and female outpatients of CMC Hospital.

Table 5: Results of Age Ratios, Sex Ratios and Joint Score of CMC Hospital

Age group	Male Number	Age Ratio	Deviation from 100	Female number	Age Ratio	Deviation from 100	Sex Ratio	First Differences
0-4	6			3			200.00	-50.00
5-9	5	43.48	-56.52	2	33.33	-66.67	250.00	61.11
10-14	17	89.47	-10.53	9	112.50	12.50	188.89	-46.83
15-19	33	75.00	-25.00	14	80.00	-20.00	235.71	-37.36
20-24	71	129.09	29.09	26	82.54	-17.46	273.08	115.93
25-29	77	121.26	21.26	49	111.36	11.36	157.14	66.82
30-34	56	73.68	-26.32	62	108.77	8.77	90.32	-25.06
35-39	75	115.38	15.38	65	86.67	-13.33	115.38	31.29
40-44	74	111.28	11.28	88	125.71	25.71	84.09	6.76
45-49	58	82.27	-17.73	75	93.75	-6.25	77.33	-15.72
50-54	67	116.52	16.52	72	121.01	21.01	93.06	-36.49
55-59	57	94.21	-5.79	44	91.67	-8.33	129.55	-95.45
60-64	54	90.00	-10.00	24	71.64	-28.36	225.00	-48.91
65+	63			23			273.91	
Total	713			556				
Total (irrespective of sign)			245.41			239.76		637.75
Mean			20.45			19.98		49.06
Joint Score			**187.60**					

From Table 5, it can be observed that the Index is extremely high (187.6), and age data is highly inaccurate according to the UN scaling. The inaccuracy is slightly critical in male age data than female and the respective sum of the absolute deviation of both male and the female age ratio from the unity (100) is 245.42 and 239.76.

Figure 4: Age Ratios by Sex for 5 years age group of CMC Hospital outpatients

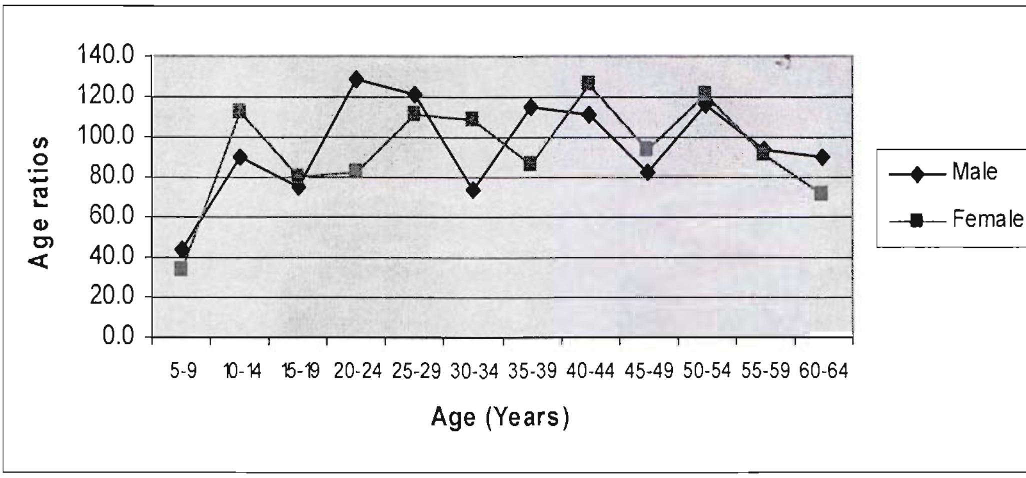


Figure 4 shows fluctuations in the age group (5-9) to (60-64) years for both male and female age ratios. For male age ratios, serious fluctuation is found in the age group 15-19 to 35-39 and for female age ratios is also found massive fluctuation in the upper age group, that is, from 35-39 to 60-64 years.

The fluctuation indicates large differences between frequencies of populations in adjacent groups. The maximum positive deviation in male is 29.1 percent in the age group 20-24, which is a negative deviation in females (17.5 percent) in the same age group. Then the maximum positive deviation for in female is 25.7 percent in the age group 40-44, which is two times higher than males (11.3 percent) in this same age group. Therefore, the maximum negative deviation is in the 5 – 9 years age group, 56.5 percent in males and 66.7 percent in females. An age ratio under 100 implies either that members of the age group were selectively under enumerated or that errors in age reporting resulted in misclassifying persons who belong to the age group. A ratio of more than 100 suggests the opposite of one or the other or both of these conditions (Kpedekpo, 1982).

2.4 Summary of findings:

In this chapter, we have evaluated age data of outpatients of CMC Hospital and applied different demographic techniques, that is, Whipple's Index (WI), Myer's Blended Index (MI), and Age-sex Accuracy Index. The results of each index show the very poor quality of age data in the outpatients of CMC Hospital. WI and MI revealed digit preference of age ending with '0' and '5' in the outpatients age data where the digit ending with '0' is more preferred than '5' for both male and female outpatients. Moreover, female is higher in age heaping than male outpatients, and the ages ending with '9' have the highest avoidance digit, followed by ages ending with '1' and '3'. On the other hand, Age-sex Accuracy Index measuring in UN scaling found very highly inaccurate age data in the outpatients of the CMC Hospital. In conclusion, we can say that there are age misreporting and digit preferences in the age data of the outpatients of CMC Hospital and very high rate of inaccuracy exists in both male and female outpatients.

CHAPTER THREE

Age and Sex data of Matlab HDSS area population

This chapter mainly presents the findings and discussion on quality of mid-year population in Matlab Health and Demographic Surveillance System (HDSS) areas by age and sex, where the same demographic techniques were applied as were those applied to CMC Hospital outpatients and these techniques are; Whipple Index (MI), Mayer's Blended Index (MI) and UN joint ratio scores.

Matlab has a very high quality data management system since its establishment in 1966 (International Centre for Diarrhoeal Disease Research, Bangladesh, 2014). A number of trained staff are involved in data collection involving registration of births, deaths, and migrations, in addition to carrying out periodical censuses. These data are maintained in a database system in the computer. In this research, it could be a good opportunity to recommend evidence based strategy to improve data quality, particularly age accuracy by comparing age data quality of Matlab HDSS population with age data quality of CMC Hospital outpatients. Table 6 presents the data on age and sex distribution of Matlab HDSS, 2012.

Table 6: Distributions of population of Matlab HDSS by age and sex, 2012

Age group	Sex		
	Male	Female	Total
0-4	12300	11991	24291
5-9	12715	12477	25192
10-14	12456	12298	24754
15-19	9542	10512	20054
20-24	6805	10480	17285
25-29	5809	9208	15017
30-34	5797	8270	14067
35-39	5743	7663	13406
40-44	5726	7629	13355
45-49	6238	7780	14018
50-54	6633	6413	13046
55-59	4421	4513	8934
60-64	3247	3599	6846
65+	7149	8412	15561
Total	104581	121245	225826

3.1 Whipple's Index:

Whipple's Index (WI) is applied to measure the level of heaping at ages ending with digit '0' and '5'. Tables 7 (a) and 7 (b) present the age distributions between 23 to 62 years for the male and female population of Matlab HDSS area in 2012.

From the analysis of Table 7 (a), the WI for male age heaping digit ending with '0' and '5' is 101.0, and the quality of the age data is highly accurate. In that regard, WI for each of the digit ('0' and '5') is then calculated separately to determine the magnitude of heaping at each one of them. The WI for male ages ending with '0' and '5' is 47.2 and 53.8 respectively, represented that the age data is highly accurate, that is, no error was found in the Matlab HDSS age data.

Table 7: Age Distribution of Male and Female Population of Matlab HDSS

Table 7(a): Age Distribution of Male Population, Matlab HDSS

Age	Number of population	Age	Number of population
23	1341	25	1254
24	1252	30	1101
25-29	5809	35	1372
30-34	5797	40	1130
35-39	5743	45	1290
40-44	5726	50	1322
45-49	6238	55	937
50-54	6633	60	698
55-59	4421	Total	9104
60	698		
61	666		
62	749		
Total	45073		

Table 7(b): Age Distribution of Female Population, Matlab HDSS

Age	Number of population	Age	Number of population
23	2153	25	1870
24	2136	30	1734
25-29	9208	35	1710
30-34	8270	40	1395
35-39	7663	45	1564
40-44	7629	50	1377
45-49	7780	55	1028
50-54	6413	60	721
55-59	4513	Total	11399
60	721		
61	733		
62	768		
Total	57987		

From the analysis of Table 7 (b), the WI for female age heaping ending with '0' and '5' is 98.3, and the quality of the age data is also found highly accurate. The WI for ages ending with '0' and '5' is 45.1 and 53.2 respectively are also found highly accurate.

For comparing the male and female population, the magnitude of error is almost same in both the digit ending with '0' and '5'. Thus, no age heaping or digit preferences exist for male and female in the Matlab HDSS population.

3.2 Myer's Blended Index

Myer's Blended Index of digit preference is used for evaluating single-year age-sex data by giving the extend of digit preference for all the digits 0, 1, 2,..., 9. The computation of Myer's Index of male population in Matlab HDSS is presented in Table 8.

Table 8: Myer's Blended Index for Male population, Matlab HDSS

Terminal Digits	Sum of age 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended sum	% Distribution	Deviation from 10
0	8859	1	8859	6181	9	55629	64488	9.40	-0.60
1	9103	2	18206	6525	8	52200	70406	10.26	0.26
2	8915	3	26745	6497	7	45479	72224	10.52	0.52
3	8424	4	33696	5959	6	35754	69450	10.12	0.12
4	8080	5	40400	5763	5	28815	69215	10.08	0.08
5	7669	6	46014	5607	4	22428	68442	9.97	-0.03
6	7165	7	50155	5208	3	15624	65779	9.58	-0.42
7	7307	8	58456	5285	2	10570	69026	10.06	0.06
8	7038	9	63342	5234	1	5234	68576	9.99	-0.01
9	6876	10	68760	5179	0	0	68760	10.02	0.02
Sum	79436			57438			686366	100.00	2.12

From Table 8, the results reveal that there is no over selection of ages ending with digits '0', '5' and other digits, as well as under selection of age (digit avoidance). Therefore, age heaping is nonexistence because the index is approximately close to '0'.

Table 9 shows the computation of Myer's Index for the female population in Matlab HDSS. The results show that there are no over selection of ages ending with digits '0', '5' and other digits, as well as under selection of age (digit avoidance). Therefore, age heaping is nonexistence because the index is approximately close to '0'.

Table 9: Myer's Blended Index for Female population, Matlab HDSS

Terminal Digits	Sum of age 10-89	Coefficient	Ages 10-89 Coefficient Product	Sum of ages 20-89	Coefficient	Ages 20-89 Coefficient Product	Blended sum	% Distribution	Deviation from 10
0	10511	1	10511	7889	9	71001	81512	9.49	-0.51
1	10444	2	20888	7869	8	62952	83840	9.77	-0.23
2	10975	3	32925	8621	7	60347	93272	10.86	0.86
3	9977	4	39908	7573	6	45438	85346	9.94	-0.06
4	9898	5	49490	7555	5	37775	87265	10.16	0.16
5	9238	6	55428	7031	4	28124	83552	9.73	-0.27
6	9161	7	64127	7037	3	21111	85238	9.93	-0.07
7	9204	8	73632	7172	2	14344	87976	10.25	0.25
8	8727	9	78543	6646	1	6646	85189	9.92	-0.08
9	8536	10	85360	6468	0	0	85360	9.94	-0.06
Sum	96671			73861			858550	100.00	2.55

From the analysis of Tables 8 and 9, results indicated that there is no existence of age heaping for both male and female population in the Matlab HDSS area, 2012.

3.3 Age-Sex Accuracy Index (Joint Score)

Table 10 shows the computation of age-sex accuracy index of male and female population in Matlab HDSS.

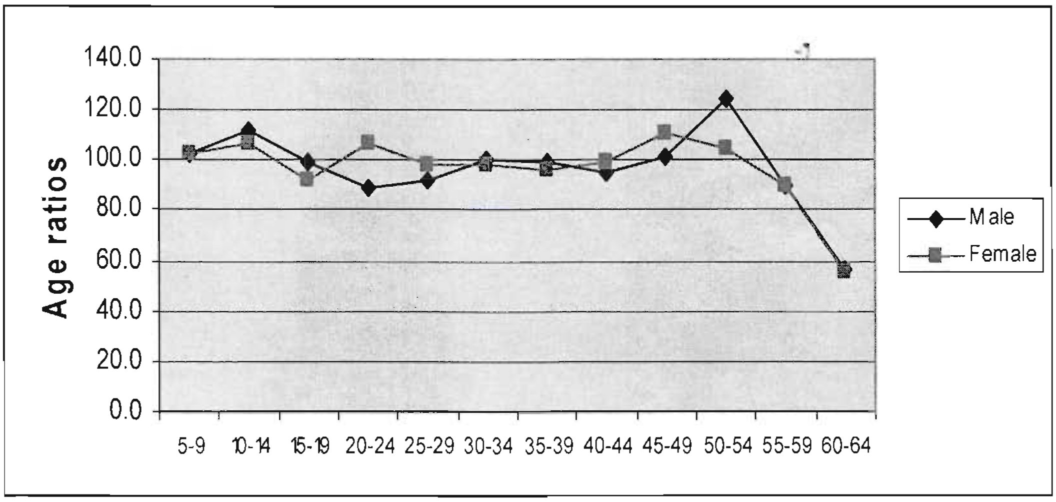
From Table 10, it can be observed that the joint ratio Index is 41.11; here, the reliability of age data is near 40, which indicates the accuracy of age data, according to the UN scaling, in the Matlab HDSS. Analysis revealed that female age data is better than the male age data and the respective sum of the absolute deviation of both male and the female age ratio from the unity (100) is 119.6 and 101.6.

Table 10: Results of Age Ratios, Sex Ratios and Joint Score in Matlab HDSS

Age group	Male Number	Age Ratio	Deviation from 100	Female number	Age Ratio	Deviation from 100	Sex Ratio	First Differences
0-4	12300			11991			102.58	0.67
5-9	12715	102.72	2.72	12477	102.74	2.74	101.91	0.62
10-14	12456	111.93	11.93	12298	106.99	6.99	101.28	10.51
15-19	9542	99.08	-0.92	10512	92.30	-7.70	90.77	25.84
20-24	6805	88.66	-11.34	10480	106.29	6.29	64.93	1.85
25-29	5809	92.19	-7.81	9208	98.22	-1.78	63.09	-7.01
30-34	5797	100.36	0.36	8270	98.04	-1.96	70.10	-4.85
35-39	5743	99.68	-0.32	7663	96.40	-3.60	74.94	-0.11
40-44	5726	95.58	-4.42	7629	98.80	-1.20	75.06	-5.12
45-49	6238	100.95	0.95	7780	110.81	10.81	80.18	-23.25
50-54	6633	124.46	24.46	6413	104.34	4.34	103.43	5.47
55-59	4421	89.49	-10.51	4513	90.15	-9.85	97.96	7.74
60-64	3247	56.13	-43.87	3599	55.69	-44.31	90.22	5.23
65+	7149			8412			84.99	
Total	104581			121245				
Total (irrespective of sign)			119.60			101.57		98.28
Mean			9.97			8.46		7.56
Joint Score			41.11					

Figure 5 shows that male age ratio and female age ratio are almost similar, except for age groups 20-24 and 50-54. In the age group 20-24, the male age ratio (88.7) is lower than female age ratio (106.3) and in the age group 50-54, female age ratio (104.3) is lower than male age ratio. The age ratio appeared to fluctuate at the higher ages, that is, age group from 45-49 to 60-64; there is a sharp fall in the age group 60-64 for both male and female age ratios. This fluctuation indicates large differences between frequencies of populations in adjacent groups. The maximum positive deviation in males is 24.5 percent in the age group 20-24, and 4.3 percent for females in the same age group. The maximum positive deviation in females is 11 percent in the age group 45-49, which is 1 percent in male in the same age group. On the other hand, the maximum negative deviation is 44 percent in the age group 60-64 for males, which is almost same for females (44.3 percent) in the same age group.

Figure 5: Age Ratios by Sex for 5 years age group of Matlab HDSS



3.4 Summary of findings:

Application of demographic techniques; Whipple Index (WI), Mayer's Blended Index (MI) and UN joint ratio scores, show that there are no age heaping and digit preferences at ages ending with terminal digits '0' and '5' in Matlab data. More specifically the single year age data of the Matlab HDSS population is accurate in quality for both male and female. Moreover, in UN scaling, the joint score of the Matlab HDSS population is also found accurate in age data. In conclusion, we can say that there is no age heaping and digit preference in the population of Matlab HDSS in 2012.

CHAPTER FOUR

Analysis of CMC Hospital outpatients and Matlab HDSS population

This chapter mainly exhibits the comparative analysis of the findings of the Chittagong Medical College (CMC) Hospital outpatients and the mid-year population in Matlab Health and Demographic Surveillance System (HDSS) by age and sex, where different demographic techniques were applied, that is Whipple Index (WI), Mayer's Blended Index (MI) and UN joint ratio scores.

4.1 Whipple's Index:

From Table 11, WI of male ages ending with the digit '0' and '5' is 246.3 for CMC Hospital outpatients and 101.0 for population in Matlab HDSS area. When comparing age data of CMC Hospital male outpatients with male population of Matlab HDSS, the quality of the age data is considered to be very rough for outpatients of CMC Hospital, whereas the quality of the age data of Matlab HDSS population is highly accurate. On the other hand, WI of female ages ending with '0' and '5' is 295.0 for CMC Hospital outpatients and 98.3 for population in Matlab HDSS. Again, when comparing with CMC Hospital female outpatients and Matlab female population, the quality of the age data is also considered very rough for Hospital outpatients, whereas the quality of age data is highly accurate in Matlab population.

Moreover, WI for each digit ('0' and '5') is then compared separately for male and female population of both CMC Hospital outpatients and Matlab HDSS population. The WI of CMC Hospital outpatients for male ages ending with '0' and '5' is 133.4 and 113 respectively, while the corresponding values are 169.7 and 125.0 for female ages ending '0' and '5' respectively. This indicates age data is rough for both male and female outpatients of CMC Hospital. However, WI for Matlab population, the male ages ending with '0' and '5' are 47.2 and 53.8 respectively, while for female ages ending with '0' and '5' are 45.1 and 53.2 respectively that suggests the age data of Matlab population is very accurate.

Table 11: Whipple's Index for outpatients of CMC Hospital and Population of Matlab HDSS

Whipple's Index	Ending with '0'		Ending with '5'		Both '0' & '5'	
	Male	Female	Male	Female	Male	Female
Outpatients of CMC Hospital	133.4	169.7	112.9	125.0	246.3	294.7
Population of Matlab HDSS	47.2	45.1	53.8	53.2	101.0	98.3

4.2 Myer's Blended Index:

Myer's Blended Index of digit preference is used for evaluating single year age-sex data by giving the extent of digit preference for all the digits 0, 1, 2... 9. The comparative analysis of computation for Myer's Index for male and female outpatients in CMC Hospital and Matlab HDSS population is presented in Table 12.

Table 12: Myer's Blended Index for outpatients of CMC Hospital and Population of Matlab HDSS

Terminal Digits	Deviation from 10 (Outpatients of CMC Hospital)		Deviation from 10 (Population of Matlab HDSS)	
	Male	Female	Male	Female
0	16.4	21.5	-0.6	-0.5
1	-5.9	-7.0	0.3	-0.2
2	1.5	-1.8	0.5	0.9
3	-4.7	-6.0	0.1	-0.1
4	-5.7	-5.7	0.1	0.2
5	13.0	14.6	0.0	-0.3
6	-5.0	-3.3	-0.4	-0.1
7	-3.8	-4.0	0.1	0.2
8	1.0	-0.7	0.0	-0.1
9	-6.8	-7.7	0.0	-0.1
Sum	63.7	72.2	2.1	2.6

In Table 12, the results show the over selection of ages ending with digits '0' and '5' for male outpatients of CMC Hospital is 16.4 and 13.0 respectively. On the other hand, age ending with digits '0' and '5' for female outpatients in the CMC Hospital is also found over selection of age preference which is 21.5 and 15.0 respectively. However, there is no over selection of age preference, that is, almost near to zero of age ending with digits '0' and '5' for male and female population of Matlab HDSS.

But in the population of Matlab HDSS, no over selection of age preference is found, that is, almost zero of age ending with digits '0' and '5' for the female population.

The results, also show that about 63.7 percent of male outpatients and 72.2 percent of female outpatients of the CMC Hospital reported ages with incorrect final digits; this is very limited about 2 percent in Matlab HDSS.

4.3 Age-Sex Accuracy Index (Joint Score)

According to the UN scaling, from Table 13, it has been observed that the accuracy of age data is comparatively more accurate in Matlab HDSS population than CMC Hospital outpatients. The rate of accuracy score in the Matlab HDSS population (41.1) is about five times higher than outpatients of the CMC Hospital (187.6).

Table 13: UN Joint score for outpatients of CMC Hospital and Population of Matlab HDSS

Outpatients of CMC Hospital (Joint Score)	187.6
Population of Matlab HDSS (Joint Score)	41.1

4.4 Causes of Misreporting:

In respect to Bangladesh, mainly the poor and low income groups of people come for treatment in the government medical hospitals. These poor and low income groups are mostly from rural areas and from urban slums, where limited or poor health facilities are available and people are mostly illiterate and unaware about the importance of accurate age factors in proper diagnosis and determining age burden of diseases. In this study of CMC Hospital outpatients', it has been observed that about 57 percent outpatients were from the rural areas and 43 percent from urban areas of Chittagong. Dr. Anne and Dr. Khalid (2000) in their study revealed that about 40 percent outpatients of the hospital in Sylhet reported that their monthly income is below 1000 taka and 53 percent reported their monthly income is in between 1001 and 5000 taka. They also found that the majority (61 percent) of the outdoor patients were illiterate. Furthermore, female mostly misreported than male, in that case young unmarried women heaped their age of 10-12 and young married women at the age 15 (Edmonston and Bairagi, 1981). In rural areas, misstatement was occurring at the early age and it increases monotonically with age (Bairagi, 1982).

In summary, it is evident that education is one of the major factors of age misreporting, where Aimee and Samuel (1991) revealed that there is an association of age misstatement with literacy and low educational attainment. Then the other important factor of age misreporting is unawareness as because, people have lack of knowledge about the importance of accurate age, which has the huge impact on their daily lives; that is, proper treatment and medication, disease diagnosis, identifying the age burden diseases, marriages and other social issues.

4.5 Implications of Misreporting:

In clinical studies age is one of the most commonly assessed variables, and other parameters are often analyzed and the results interpreted in relation to age. Therefore, modern health care services require accurate age and sex data for appropriate decisions, disease surveillance and in the provision of required drugs and medications.

The following are the most possible impact of age misreporting causes:

- Misreporting could be put into the wrong age subgroup that affect the population age group of census and survey.
- In the decision- making process at the level of the physician and patient encounter, the accurate age of the patient is important because many decisions are age sensitive and misrepresentation of age may lead to inappropriate action. For example, a screening mammography is recommended to start at age 50 and patient's inaccurate age may result in having either an unnecessary or delayed test (Denic et al., 2003).
- Females are mostly prone to hiding the age, especially those who have an early marriage (as per the government rules the exact marriage age for male and female is 21 and 18 respectively), and it will definitely be critical when women are pregnant and have a tendency of hiding their actual age to the doctors as because they and their families might have to face legal questions regarding child marriage and early pregnancy. For avoiding these difficulties women and their families misreport their age to doctors.
- In the older patients' misreporting propensity is higher; as a result, there is a possibility of misclassifying age burden of diseases.
- The misreporting of patients' ages might lead to wrong treatment with all its accompanying consequences.

4.6 Recommendation:

- Birth registration should be mandatory in the country's demographic profile for future generation of Bangladesh;
- Advocate to increase the mass awareness about the importance of accurate ages in the people's daily life;
- In all of the medical and other purposes, age data should be mentioned with date of birth in the record book and validate if there is any confusion;
- The age data might be improved by analytical approaches such as standard smoothing techniques or by more careful and sensitive data collection techniques;
- To improve age reporting, calendar of Historical Events method should be used, which relates the birth of an individual to some notable event, and the Family Relationships Estimation method, which pieces together age estimates for all members of a family or household.
- As age and sex data is found highly accurate in Matlab HDSS population, in that case, we are recommending to adopt and scale up the Matlab HDSS model of data collection all over the Bangladesh.

4.7 Conclusion

In this study Whipple's Index. Myer's Blended Index and Age-sex Accuracy Index were applied to evaluate the age and sex data of CMC Hospital outpatients and the population of Matlab HDSS. By applying WI, digit ending with '0' and '5' for male was 246.3 in CMC Hospital outpatients, which is above two times higher than Matlab HDSS population (101.0). On the other hand, for female, digit ending with '0' and '5' was 295.0 of CMC Hospital outpatients that are three times higher than Matlab HDSS population (98.3). The findings indicate that Matlab HDSS data is more accurate than CMC Hospital data for both male and female population, and it is true for each digit ('0' and '5') when compared separately.

In MI, ages ending with digits '0' and '5' for male outpatients in the CMC Hospital was 16.4 and 13.0 respectively, whereas in Matlab HDSS population it was almost zero, which means no over the selection of age. The age data ending with digits '0' and '5' for female outpatients in CMC Hospital was 21.5 and 15.0 respectively, but it was also near to zero in population of Matlab HDSS, that is, no over selection of age. The results reveal that age data of Matlab HDSS population is more accurate than age data of outpatients in CMC Hospital.

Moreover, in UN scaling, the score of the outpatients of the CMC Hospital (187.6) was more than four times higher than Matlab HDSS population (41.1), which indicated more accurate age data in Matlab HDSS population than CMC Hospital outpatients

Finally, we can conclude by saying that in all cases, that is, WI, MI and Age-sex Accuracy Index, the results suggest that Matlab HDSS data is more accurate than CMC Hospital outpatients. This research has provided an evidence to reduce the age misreporting and digit preference by adopting the Matlab HDSS model.

REFERENCES

Abd. Latib Talib, Mohd Sofi Ali, Marlini Sahul Hamid and Khamsiah Mohd Zin, (2007). *Age reporting Behaviour: A Case Study of 1991 and 2000 Population and Housing Censuses, Malaysia.* Department of Statistics Malaysia.

Amiee R. D, Samuel H. Preston (1991). *Age misreporting and its effects on adult mortality estimated in Latin America.* Population Bulletin of the United Nations, 31/32

Anne Cockcroft and Khalid Omer (2000). *Hospital improvement initiative in Bangladesh,* Final report

Bailey M, Makannah TJ (1996). *An evaluation of age and sex data of the population censuses of Sierra Leone: 1963-1985.* Genus; 52(1-2):191-9

Bairagi R, Aziz K.M, Chowdhury M.K and Edmonton B. (1982). *Age misstatement for young children in rural Bangladesh.* Demography, vol 19, no. 4

Becker S. Diop-Sidibé N. (2003). *Does the use of the calendar in surveys reduce heaping?* Stud Fam Plann.; 34 (2):127-32.

Borkotoky K, Unisa S (2014). *Indicators to examine the quality of large scale survey data: an example through district level household and facility survey.* 5;9 (3):e90113. doi: 10.1371/journal.pone.0090113.

Chuks J. Mba (2014). *Examining the Accuracy of Age-sex data, An Evaluation of recent Sub-Saharan African population census.* Community and Change in Sub-Saharan African Demography.

Crawford S, Johannes C, Bradsher J, Stellato R, Sherman S, Samuels S (2000). *Digit preference in year at menopause. Data from the study of women's health across the nation.* Ann Epidemiol. ; 10 (7):457.

Denic S, Khatib F, Saadi H, (2003). *Quality of age data in patients from developing countries.* Journal of Public Health, 26(2), 168–171.

Dudley L. Poston (2006). *Age and Sex.* Handbook of population, Chapter 1: 19-56

Edmonston B, Bairagi R. (1981). *Errors in age reporting in Bengali populations.* Rural Demogr. ; 8(1):63-87.

Ekanem II, Ayeni O (1976). *Models for measuring digital preferences in the census of developing countries.* Jimlar Mutane. ; 1 (2):65-87.

Elo I.T., Preston S.H., Rosenwaike I., Hill M. and Cheney T.P. (1996). *Consistency of age reporting on death certificates and social security administration record among elderly African- Americans.* Soc Sci Res, 25, 292–307.

Gunasekera HR (1984). Colombo, Sri Lanka, University of Colombo, Demographic Training and Research Unit, 40 p.

Haandrikman K. (2004). *Using a local time-path calendar to reduce heaping in durations of postpartum Amenorrhea, breastfeeding, postpartum Abstinence and contraceptive use.* Time Soc. 13, 339-62.

Hayes S.J. (2008). *Terminal digit preference occurs in pathology reporting, irrespective of patient management implication.* J Clin Pathol. ; 61(9):1071-2.

Hill, M.E, Preston, S.H and et al. (1997). *Age-linked institutions and age reporting among older African Americans.* Social Forces, vol. 75, no. 3

International Centre for Diarrhoeal Disease Research, Bangladesh, (2014). Retrieved from www.icddrb.org

Ira Rosenwaike, Mark E. Hill, (1996). *Accuracy of Age Reporting among Elderly African Americans: Evidence of a Birth Registration Effect.* Research on Aging 18: 310-324

Irma T. Elo, Samuel H. Preston, Ira Rosenwaike, Mark Hill and Timothy P. Cheney (1994). "*Consistency of Age Reporting on Death Certificates and Social Security Administration Records among Elderly African-American Decedents*", Population Aging Research Center, University of Pennsylvania working Paper Series No. 95-03

Johannes I. and Polly F., Bancroft's, (1970). *Introduction to Biostatistics, (2nd Edition),* Harper and Row publishers, New-York

Kabir M, Chowdhury K. (1981). *The pattern of age reporting errors in the districts of Bangladesh.* Rural Demogr. ; 8 (2):33-54.

Kestenbaum B. (1992). *A description of the extreme age population based on improved Medicare enrollment data,* Demography, 29, 565–580.

Knodel J, Chayovan N. (1991). *Age and birth date reporting in Thailand.* Asian Pac Popul Forum. ; 5(2-3):41-50, 64-76.

Kpedekpo G.M.K. (1982). *Essentials of Demographic Analysis for Africa.* Hernerman Educational Books Inc., New Hemisphere.

Larson A. (1987). *Age and sex reporting in SUPAS 85.* Majalah Demografi Indones. ; 14 (28):47-63.

Lerche C. O. (1983). *Social and Economic Statistics for Africa, (2nd Edition).* Longman group, London.

Locker TE, Mason SM. (2006). *Digit preference bias in the recording of emergency department times.* Eur J Emerg Med.; 13(2):99-101.

Mason, K.O and Cope, L.G. (1987). *Sources of age and date of birth misreporting in 1900 U.S. census.* Journal of Demography, vol 24, no. 4

Materia E, Mehari W, Mele A, Rosmini F, Stazi MA, Damen HM, Basile G, Kifle T, Miuccio G, Ferrigno L. (1993). *Census in a rural area of Ethiopia: methodology and results.* Eur J Epidemiol. ; 9(1):5-9.

Mukherjee BN, Mukhopadhyay BK, (1988). *A study of digit preference and quality of age data in Turkish censuses.* Genus. ; 44 (1-2):201-27.

Nagi M.H., Stockwell, E.G and Snavley, L.M. (1973). *Digit preference and avoidance in the age statistics of some recent African censuses: some patterns and correlates.* International Statistical Review, vol. 41, no. 2

Omondi-Odhiambo (1985). *"Age estimation in clinical and public health research, explores problems associated with age reporting in Africa".*

Pardeshi G. S. (2010). *Age heaping and accuracy of age data collected during a community survey in the Yavatmal district, Maharashtra.* Indian J Community Med, 35(3) , 391-395.

Roberts, J. M and Brewer, D.D. (2001). *Measures and test of heaping in discrete quantitative distributions.* Journal of Applied Statistics, vol, 28, no. 7

Pérez-Fernández GA, Grau-Abalo R. (2012). *Determination for preference of the terminal digit in arterial hypertension. Utopia in underdeveloped countries.* Arch Cardiol Mex. ; 82 (1):73-4.

Pullum T. W. (2005). *Assessing the quality of age and date reporting in the demographic and health surveys.* Paper prepared for UT Population Research Center brownbag.

Rini Savitridina. (1997). *Determinants and consequences of early marriage in Java, Indonesia.* Asia – Pasific Population Journal, vol.12, no. 2

Shirley L., Ravi V. and Margaret M., (2004). *An evaluation of the age and sex data from the census population of Canada, provinces and territories, 1971 to 2001*

Shryock H.S., Siegel J.S. and Associates, (1976). *The Methods and Materials of Demography.* Condensed edition by Edward G. Stockewell, Academic Press, New York

Smoller S.W. (2004). *Biostatistics and Epidemiology, (3rd Ed.).* Sponger-Verlay, New York

Thavarajah S, White WB, Mansoor GA., (2003). *Terminal digit bias in a specialty hypertension faculty practice.* J Hum Hypertens. ; 17(12):819-22.

Ueda, K. (1980). *Basic demographic data, their accuracy and use for subnational population projection in Asian and Pacific countries.* Tokyo: South East Asian Medical Information Center

Wiki/Whipple, (undated), http://en.wikipedia.org/wiki/Whipple's_index

Yusuf Bello (2012). *Error Detection in Outpatients' Age Data Using Demographic Techniques.* Int. J. Pure Appl. Sci. Technol., 10(1), pp. 27-36