Data quality of reported age-sex structure from a community diagnosis program of a hilly region of Nepal

Bhandary S,[1] Shakya SR,[2] Sinha NP[3]

[1]Patan Academy of Health Sciences, [2]Kathmandu University School of Medical Sciences/
WHO- IPD Survillence Office, [3]Kathmandu University School of Medical Sciences

ABSTRACT

Introduction: Indicators of health are extensively based on the age-sex data and thus warrants a detailed assessment of data. The main purpose of this study is to assess and then adjust the various types of bias and error that can occur during any census or survey.

Method: This study is based on a cross-sectional *de jure* census of Ugrachandi Nala VDC, Kavre, Nepal, which was conducted during August 30 – September 5, 2005. Digit and gender preference is analysed using Whipple Index Myers Index and, Age-Sex Accuracy Indices. The data is adjusted using Beer's Method and Hill Technique. Microsoft Excel 2003 and SPSS 11.5 programs were used for data entry and analysis respectively.

Result: The Whipple Indices revealed that the single year data as rough and Myers Index confirmed that 0 and 5 are the most preferred digits. Age-Sex Accuracy Indices disclosed the quniquennial age data as highly inaccurate. Single-year age data adjusted using Beer's Method and quinquennial age data adjusted using Hills Technique are found to be an ideal age-sex structure.

Conclusion: Although the reported age shows serious problems of digit preference and age displacement in Ugrachandi Nala VDC, its data quality is better than the Nepal and Kavre district. This study shows that bias and errors related to the reported age and gender in a census and survey can be identified and adjusted using various demographic tools. Thus, it is suggested to use adjusted age-sex structure for calculating demographic, epidemiological and other health indicators.

Keyword: age-sex, data quality, census, community, kavre, nepal.

CORRESPONDENCE

Mr. Shital Bhandary
Assistant Professor, Department of Community Health Sciences
Patan Academy of Health Sciences
E-mail: shitalbhandary@pahs.edu.np

Introduction

The quality of reported or declared age data obtained from the past national censuses has not been very accurate[1] in Nepal but it has improved a lot in the national surveys.[2] As most of the indicators of health are based on the age-sex data, assessment and adjustment of reported age and gender is crucial to evaluate and formulate relevant health plans and policies at national, regional, district and local levels and for programs to achieve the Millennium Development Goals (MDGs) by 2015.[3]

For this purpose, it is necessary to assess and adjust the primary data obtained from a census or a survey for their quality. It is not only sufficient to adjust the primary data but it is also indispensable to use the same adjustment technique in order to compare indicators at two or more settings. This process in turns ensures the quality, comparability and relevancy of the indicators being compared.

Generally, population structure is constructed in terms of age and gender and there exist various methods to assess and adjust the single as well as grouped age-sex data. Yet, selection of the best, relevant and unique adjustment method is difficult as well as critical. It requires intense technical caution and involves advance computing processes. Therefore, this study aims to suggest one such possibility for the raw data obtained from the 2005 census of the Ugrachandi Nala VDC in a very simple manner. It also intends to present the importance of age-sex data evaluation and adjustment before using it to calculate the indicators of health and the Millennium Development Goals.

Method

This study is based on a *de jure* census count of 986 households and its 5,722 inhabitants residing in the Ugrachandi Nala Village Development Committee (VDC), Kavrepalanchowk district, which was conducted during August 30 – September 5, 2005. *De jure* census includes all the inhabitants who belong to their usual place of residence[4]. The fourth batch Bachelor of Medicine and Bachelor of Surgery (MBBS) students of School of Medical Sciences, Kathmandu University conducted this census as their usual residential Community Diagnosis Program (CDP). Core questionnaire was designed and administered in English but the questions were asked in Nepali, Newari and Tamang languages with the aid of students, faculty, teachers and local native speakers.

The CDP questionnaires were first pre-tested in the premises of Nala Health Post with rural inhabitants and Nala Bazaar with urban dwellers. Technical and other limitations observed and identified from the pre-tested questionnaires were modified during a week long orientation program just before the field work. Affirmative verbal consent triggered the *de jure* census. The data collection norms and procedures during the study period were constantly monitored by the CDP coordinator and two residential supervisors stationed at two different locations with two groups of students. Information on age of the household members was collected from the head of the household or from a person who was able to respond to the questions of the 2005 CDP questionnaire. Enumerators were not able to verify the reported age with the official documents like passport, citizenship, birth certificates or religious and mythological manuscripts prepared at birth of the child, which was the limitation of this study.

In the present study, reported single-years-of-age was assessed using the line diagram, Whipple's Index[4] and Myers Index[5] and adjusted using the Beer's Method[6]. Reported five-years age data (quinquennial age groups) was evaluated in terms of age ratio score, sex ratio score and age-sex accuracy index and adjusted using Arriaga[7], Strong[7] and Hill[10] methods. Male and female age ratio scores and sex ratio score are used to evaluate the accuracy of reported age and the data quality by gender whereas the age-sex accuracy index was used to measure the extent of age-sex reporting bias and errors for the quinquenial age groups.

Data were first entered in the MS Excel 2003 spreadsheets based on the code books of the study questionnaire. They are later transported to the Statistical Package for Social Sciences (SPSS) 11.5 datasets for further analysis using SPSS programming language.

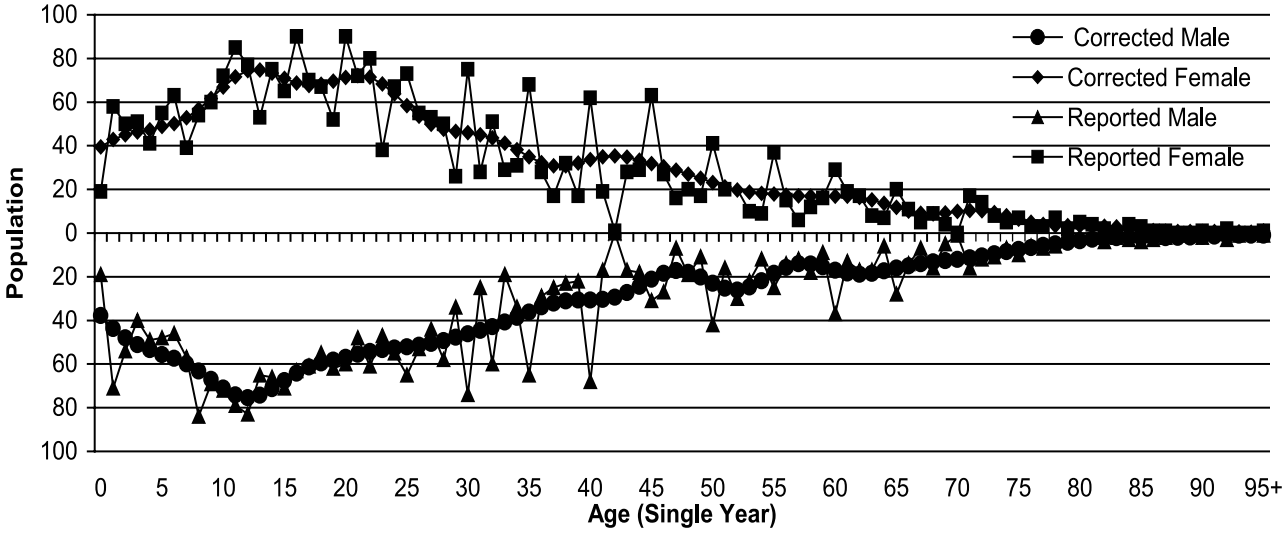
Result

The study found 986 occupied households and 5,800 inhabitants during the study period. However, complete age and sex information was obtained from the 5,722 inhabitants only. Out of them, nearly 96 per cent of the household heads were Hindus followed by Buddhsts (2.5 %), Muslims (1.7%) and Christians (0.1 %). More than

half (51.2 %) of the people were Brahmins, 36.6 per cent were Newars, 5.1 per cent were Chhetris, 1.7 per cent were Tamangs and rest were from other castes or ethnic groups. Seven out of twelve and seven out of twenty were nuclear and joint families whereas seventeen out of two-hundred were extended families. The male and female adult literacy rates were found to be 83.91 and 40.01 per cent respectively.

Figure 1 shows the unadjusted and adjusted population distribution by single-years-of-age and gender. Peaks and troughs are seen for ages ending at 0 and 5 for the reported single year data. The adjusted population distribution, on the other hand, shows a very smooth distribution of population by single-years-of-age and gender.

Figure 1. Reported and Corrected Age-Sex Distribution by Single Age
Ugrachandi Nala VDC, Kavre, Nepal: 2005



Source: 2005 CDP Database, Ugrachandi Nala VDC, Kaver, Nepal

Table 1. Accuracy of reported and adjusted Age-Sex Structure

	Reported Population		Adjusted Population	
	Whipple's Index	Myers Index	Whipple's Index	Myers Index
Male	167.63	10.41	99.99	2.03
Female	176.24	10.12	99.85	1.89
Total	172.03	10.68	99.92	2.16

Table 1 shows the Whipple Index for total, male and female reported population as 172.03, 167.63 and 176.24 respectively. Conversely, the Whipple Index for the Beer's adjusted total, male and female population distribution is obtained as 99.92, 99.99 and 99.85 respectively.

Figure 2. Preference for Digits (Myer's Index)
Ugrachandi Nala VDC, Kavre, Nepal: 2005

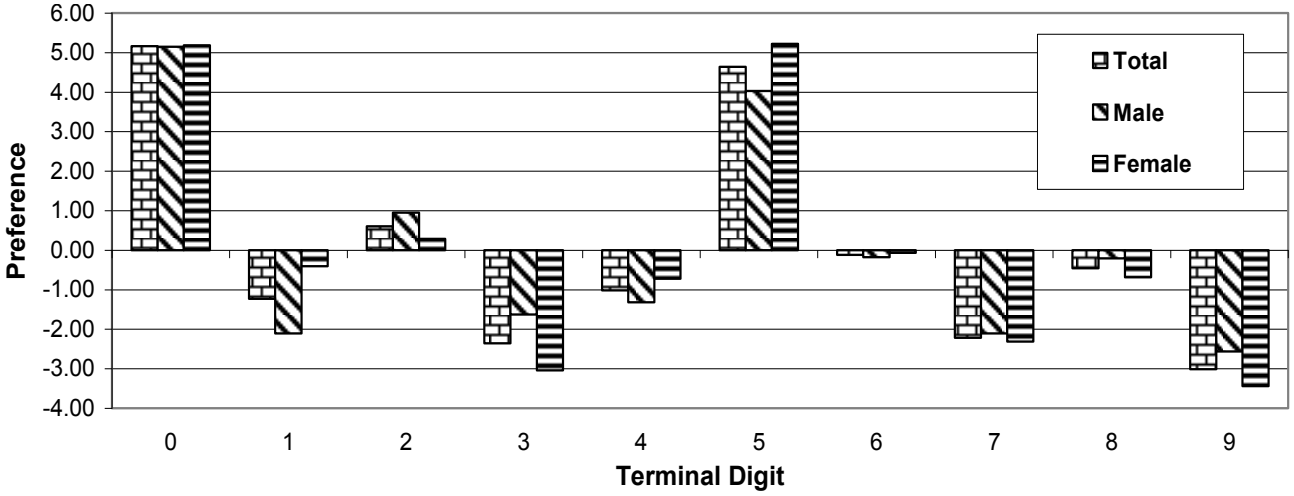


Figure 2 shows the terminal digit preference trend obtained from the Myers Blended Method. It reveals that 0 and 5 are the most preferred digits whereas 9, 3, 7 and 1 are the least preferred digits. Slight preference on digit 2 is seen but study population also tends to reject the terminal digits 4, 8 and 6.

Figure 3: Unadjusted and Adjusted Sex Ratios
Ugrachandi Nala VDC, Kavre, Nepal: 2005

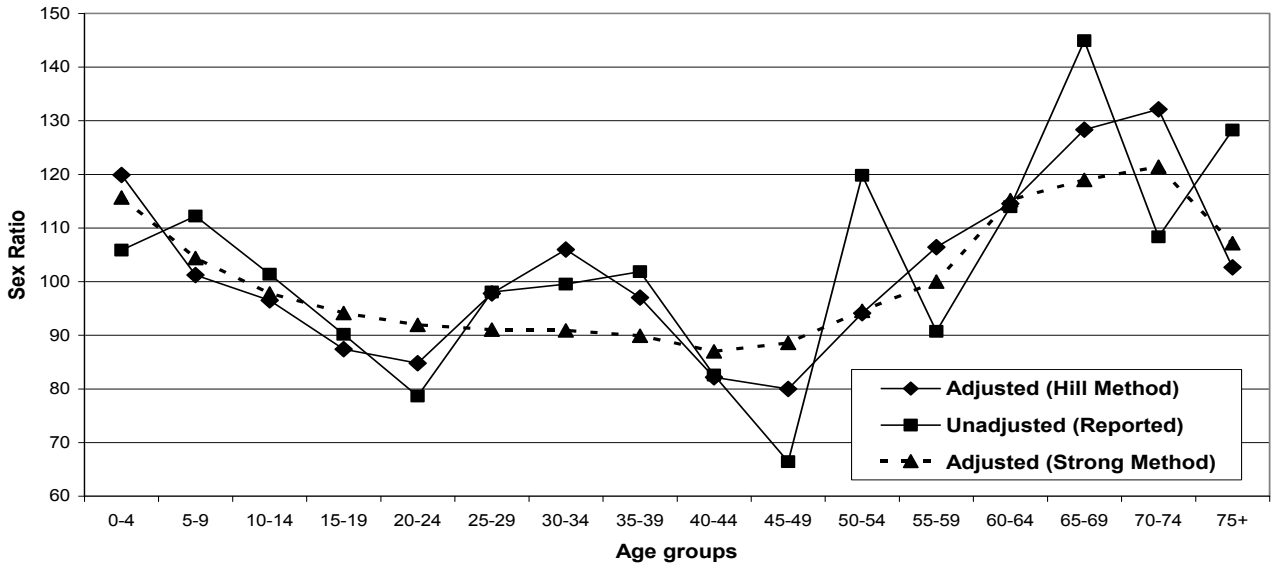


Table 1 also shows the Myers Index for total, male and female reported population as 10.4, 10.1 and 10.7 respectively. The Myers Index for the adjusted population is obtained as 2.03, 1.89 and 2.16 for both sexes, males and females respectively.

Figure 3 shows sex ratios for reported and adjusted age-sex distribution. For the reported population, it reveals the deficit of males at ages 15 – 29 and especially at ages 20 – 24. Equal or more males were found at ages 30 – 39 before declining again from ages 40 – 59 the only exception being at ages 50 – 54. The sex ratios still show some lopsided trends for Hill adjusted population, which is not seen for Strong adjusted age-sex distribution.

Table 2: Summary of Indices Measuring the Accuracy of Grouped Data

	Reported Population	Complete Age-Sex Adjustment Method		
		Arriaga	Strong	Hill
Sex Ratio Score	19.82	9.76	3.89	9.75
Male Age Ratio Score	14.32	3.57	2.22	3.06
Female Age Ratio Score	9.72	5.88	1.94	4.00
Age-Sex Accuracy Index	83.49	38.73	15.82	36.32

Table 2 presents the various indicators that are used to measure the accuracy of reported and adjusted grouped age-sex structures. It shows the lowest sex ratio score for the Strong method and an identical sex ratio score for Arriaga and Hill methods. Similarly, age ratio scores for both sexes are smaller for the Strong method than Arriaga and Hill method. The age-sex accuracy index is found to be 83.49 for the reported age-sex data and 38.73, 36.62 and 15.82 for adjusted population structure based on Arriaga, Hill and Strong methods respectively.

Discussion

Whipple Index[4] shows the degree of age heaping for terminal digit 0 and 5. In our study, age heaping was not only occurred at different ages but it was found to be more pronounced among females than males; which is consistent with findings from the past Nepalese censuses.[1,11] Past studies also revealed that the causes and patterns of age or digit preference vary from one culture to another, but preference for ages ending in "0" and "5" is quite widespread.[12] Similarly, age heaping was also found to be most pronounced among populations or population subgroups having a low educational status.[4] As the female adult literacy rate is 2.1 times less than males in our study, it might be the main cause of the deficient single-years-of-age data. Whipple Indices in our study population were found to be lower than the 2001 national and Kavre's population, both being greater than 200. [1, 11] These results indicate that the overall digit preference pattern of the study population is in fact better than the national and district population.

Myers Index[5] shows the degree and magnitude of digit preference. The results (Figure 2) once again confirmed that the age heaping at terminal digit 0 and 5 is quite common in the study population. It has been observed that in some cultures certain numbers are specially avoided, e.g., 13 in the West and 4 in the Orient.[4] Our study revealed that Ugrachandi Nala people rejected the terminal digit 8 as well. This was an unusual trend as the past Nepalese censuses and survey showed some preference for this terminal digit[1, 11]. Yet the Myers Indices for the reported age-sex data by gender revealed that the degree of terminal digit preference in Ugrachandi Nala VDC is lower than the reported gender disaggregated 2001 national population structures.[1, 11, 13](Table 1).

However, the magnitude of digit preference was still very high in our study (Table 1). Therefore, it warranted the adjustment of reported gender disaggregated single-years-of-age data before using it to compute the indicators of health and to formulate health plans, policies and programs at local level. Consequently, Beer's Method[6] was used to adjust the single-years-of-age data in the present study. The Whipple's and Myers Indices for Beer's adjusted total, male and female population structures were consistently lower than the reported age-sex data (Table 1). Thus, the Beer's adjusted single year population structures are suggested as the ideal single-year-of-age population structures at all levels.

The choice of an ideal grouped population structure is critical and it is obtained on the basis of the lowest and plausible sex ratio score[4,7] age ratio score[4,7] and age-sex accuracy index[8,7] among different available adjustment techniques. Age-sex adjustment methods like Carrier-Farrag[4], Karup-King[7], United Nations[8] can only adjust the population 10 years and above. Nonetheless, complete adjustment of grouped age-sex data is possible from Arriaga[7], Hill[10] and Strong method.[7] Thus, it is logical to choose the ideal grouped population structure from one of these three methods.

Unadjusted sex ratios (Figure 3) showed an irregular trend for the study community. This type of erratic trend was very unusual because sex ratios showed a consistent deficit of males from ages 15 – 49 and 15 – 59 respectively for Nepal and Kavre district in the 2001 census.[13] Displacement of age started from age 30 – 39 but it worsened for age 40 – 54. These values indicate that the age displacement up to 5 years or more had occurred in the present study and demands correction of the raw data.

Thus, Hill adjusted population structure showed reasonably correct values of sex ratio at different quinquennial age groups whereas the sex ratios obtained from the Strong method suited well with the sex ratio pattern of Nepal and Kavre.[1, 11, 13] Moreover, sex ratio

scores for the Strong adjusted age-sex structure is closer to zero and was also better than the Arriaga and Hill adjusted population distributions. These results indicate that the population structure adjusted using Strong method is better than the Arriaga and Hill methods.

Similarly, male and female age ratio score were also found to be consistently lower for the adjusted age-sex structure obtained by the Strong method than the Arriaga and Hill method. However, age ratio score for Hill adjusted age-sex data was somehow better than the Arriaga adjusted population structure. This indicates the superiority of Strong method over the Arriaga and Hill methods for the study community. Since sex ratio and age ratio are very sensitive to the migration and mortality in higher age groups[4,8], extreme smoothing of age data can distort the true pattern of the population dynamics. Thus, severe adjustment of age-sex data is not recommended until and unless huge migration and heavy mortality fluctuation is observed in the community studied.

The age-sex accuracy index is the best indicator of the quality of grouped age data by gender. The age-sex accuracy index was obtained as 21.1 [1, 11] and 27.71[13] for Nepal and Kavre district at 2001 against the 83.49 in our study (Table 2). This indicates a highly erratic quinquennial age-sex data in our study population. Age heaping and displacements were very common observations for single-year-of-age data in the Nepalese censuses[1, 11] but when the single-years-of-age are combined to form quinquennial age groups then the quality of the data improved.[1] It is because some of the peaks and troughs, as seen in Figure 1, were eliminated in the process[11], which was not the case in our study. As the grouped data was found to be inaccurate, it was adjusted for its age reporting bias and errors. In our study, the age-sex accuracy index obtained from the Strong method (15.82) was lower than the Arriaga method (38.73) and Hill technique (36.32). The ideal grouped population structure is obtained from the adjusted grouped age-sex structure having lowest age-sex accuracy index. This could be, sometimes, misleading factor as the strongest smoothing method based on strong regression coefficients may distort the true pattern of population structure. Thus, it was necessary to choose a realistic age-sex structure based on medium value of age-sex accuracy index.

Thus, we can not take the population structure obtained from the Strong method as an ideal grouped age-sex data because this method is only recommended when all the data quality indicators indicate deficient age-sex data[7], which is not true in our study. For instance, Whipple's and Myers Indices already revealed that the reported single-years-of-age data is in fact better than national and district level data. Thus, adjusted group age-sex data obtained by the Arriaga method or Hill technique should be selected as the ideal grouped population structure. As the sex ratio scores and age ratio scores for the Hill adjusted population distribution were lower than the Arriaga adjusted age-sex data (Table 2), Hill smoothing technique is preferred over the Arriaga method. Moreover, age-sex accuracy index for Hill adjusted age-sex structure was also slightly less than the Arriaga adjusted population distribution. Based on these results, it is suggested the age-sex structure adjusted by the Hill technique as the ideal quinquennial population distribution for the study population.

Conclusion

The Whipple's Index and Myers Index indicate that the reported age-sex structure is rough and there is an extreme heaping or preference for terminal digits "0" and "5". Yet, their values are lower than Nepal and Kavre's data. This means that Ugrachandi Nala VDC has comparatively good data than the national and district level. Age-sex accuracy index revealed that the reported quinquennial age-sex structure is highly inaccurate. It may have caused by the actual trend of the age-sex structure or, more firmly, by the severe displacement of age by both sexes in the present study.

Nonetheless, this study showed that deficient single-years-of-age and limited quinquennial age data can be adjusted using Beer's correction method and Hill smoothing technique respectively. But, the age-sex adjustments should be avoided as long as one can confirm or control the data quality. In other words, the adjustment technique is not very useful when people declare or report their true age or when enumerators can verify the reported age with the official or religious/ mythological documents prepared at the time of the birth. For this purpose, ordinary people should be made aware about the importance of revealing true information and consequences of bias data as they impact the formulation of national, regional, district and local level planning, policies and programs. Consequently, the needy people or group of people may be deprived from all the

requirements they really deserve from the biased age-sex structure. Thus, it is suggested to verify the age data with the valid official documents like birth certificate or the religious and mythological documents prepared at the time of birth. If they are not available then the reported age should be verified by means of citizenship certificate, marriage certificate, school registration, passport, etc.

Same types of age reporting bias and errors have been observed in the past Nepalese censuses[1,11] and a lot of resources were mobilized to obtain a relatively reliable age-sex population structure. Thus, it seems like that the quality of declared or reported age can be improved in any censuses if the task of age recording is given to one special group who are well trained for this specific purpose, which in turn, demands for the additional time, money and manpower.

Based on these results, we can conclude that the population structure obtained from the Beer's method and Hill technique are the ideal single-years-of age and quinquennial age-sex structures respectively for the Ugrachandi Nala VDC, Kavre, Nepal for the 2005. All the indicators related to the Millennium Development Goals including the indicators of health for this VDC should be computed using these adjusted age-sex structures and, all the local level plans, policies and programs should be based on them. It is suggested to replicate similar procedures for all the regions, zones, districts and local level entities so that firm, relevant and plausible plans, policies and programs could be assessed and devised at the respective level to achieve the Millennium Development Goals by 2015.

Acknowledgement

The authors would like to thank District Health Office, District Administrative Office, Kavre and Nala Health Post for their invaluable support. We would also like to thank all the students of fourth batch MBBS program, KUSMS for their hard work and all the inhabitants of Ugrachandi Nala VDC for their consent and active participation. Thanks are also due to Mr. Prem Raj Shakya, who proved to be an excellent supervisor for the community diagnosis programs. We appreciate the presence of other KUSMS faculty members and administrative staffs. We are truly thankful to Prof. Kun-Young Sohn and his energetic wife for their motivation and guidance to the students during their stay. Last not the least, we would like to express our deepest gratitude to the School of Medical Sciences, Kathmandu University for its tremendous support for this program.

References

1. Bhandary S. Inventory of Existing Data and Their Quality. Population Censuses of Nepal. Kathmandu: UNFPA/CBS, 2003:1–113.

2. Pant PD. Assessment of The Quality of Data, Comparative Tables from the NFHS 1996 and NDHS 2001, Socio-Economic and Demographic Indicators from Selected Surveys and the Tables of Indicators Based on the NLFS 1998/99, Kathmandu: UNFPA/CBS, 2003:1 – 64.

3. Government of Nepal and United Nations Country Team. Nepal Millennium Development Goals: Progress Report 2005, Kathmandu, 2005: Nepal:1 –119.

4. Shryock HS, Siegel JS and Assoc The Methods and Materials in Demography, San Diego: Academic Press Inc, 1976:29 –113 and 531 – 59.

5. Myers RJ. Errors and Bias in the Reporting of Ages in Census Data. Transaction of the Actuarial Society of America 1940;41,2(104):395-415.

6. Beers HS. Six-Term Formulas for Routine Actuarial interpolation: The Record of the American Institute of Actuaries. 1945;34,Part 1(69):59.

7. Arriaga EA, Jahnson PD, Jamisons E. Population Analysis with Micro-Computers, Vol I and II, Bureau of Census, Washington D.C. 1994.

8. United Nations. Methods of Appraisal of Quality of Basic Data for Population Estimates, Manual II, Population Studies, Series A, New York. 1955:23:42-3.

9. United Nations. Accuracy Tests for Census Age Distributions Tabulated in Five-Year and Ten-Year Groups, Population Bulletin 1952;2:59-79.

10. United Nations. Manual X: Indirect Techniques for Demographic Estimation, Population Studies (81), New York, 1983.

11. Pantha R, Sharma BR. Population Size, Growth and Development, Population Monograph of Nepal, Vol I. 1st ed. Kathmandu: Union Press. 2003:37–86.

12. Gupta AD. A Technical Note on Age Grouping, National Sample Survey 12, Indian Statistical Institute, Calcutta, 1958.

13. Central Bureau of Statistics. Population Census 2001 National Report, 1st ed Kathmandu, CBS 2002:24–32.