Original Article

# Quantifying Age Heaping and Age Misreporting in a Multicentric Survey

**D. R. Basannar, Sumeet Singh, Jyoti Yadav[1], Arun Kumar Yadav**

Department of Community Medicine, AFMC, [1]Independent Researcher, Pune, Maharashtra, India

## Abstract

**Introduction:** Demographic indices known as the age-heaping indexes were used to explore the patterns of age misreporting in a multicentric survey. **Methods:** The data of 3252 individuals were analyzed, and measurement of errors in age for the sampled data has been evaluated by Whipple's Index (WI), Myer's Blended Index, and United Nations Age–Sex Accuracy Score which comprises Sex Ratio Score, Male Age Ratio Score (ARS), and Female ARS. **Results:** Out of total 3252 participants, 828 (25.5%) were female. The mean statistical division age of our population was 34 (8.5) years and ranged from 15 to 65 years. The percentage of female ages ending with digits 0 or 5 is 23.55% and percentage of male ages ending with digits 0 or 5 is 23.28%. The calculated WI was 117.75 and 116.34 for males and females, respectively. The calculated Myer's Index for females and males is 10.53 and 25, respectively. **Conclusion:** The study provides evidence that probably age-heaping bias is less of problem in the conducted study.

**Keywords:** Age-heaping bias, Myer's Blended Index, Whipple's Index

## Introduction

Researchers have to commonly deal with random and systematic errors. They have to justify that they did all possible to minimize random error and take precautions to prevent systematic error.[1] While random error arises due to the natural fluctuation or variation in the accuracy, systematic error arises from an innate flaw in the selection or measurement.[2] Random error is likely to distort study measurements in either a positive or negative direction,[3] whereas systematic error results in an incorrect estimate of the measure of association or effect.[4]

Age is a sociodemographic variable, and accuracy of age depends on numerous factors such as literacy, occupation, socioeconomic status, and so on. Misstatement of age is a common example of content error in census and surveys. Age data display excess frequencies at round or attractive ages, so age heaping is considered to be the measure of data quality and consistency.[5] For examining the quality of population censuses, the one way that seems particularly adequate is an assessment of the extent and nature of deficiencies stemming from the rounding of ages.[6] Rounding at preferred digits continues to be a major characteristic of the age statistics of developing countries which in turn depends on the existing system of counting.[7]

A large number of multicentric surveys are being continuously conducted among Armed Forces personnel, for which the data on age are regularly collected. Armed Forces personnel differ in many ways from the census population of India in being having gone through mandatory criteria of qualification, rigorous training, and specialized military education. As of now, there is no study in the Armed Forces population to estimate age heaping. This would have an implication on all the surveys undertaken as well as future surveys as the study is based upon the randomization on the last digit of age. Keeping in view the importance of age heaping, the present study has been undertaken to assess the level of age-heaping bias among 3252 serving personnel in a study of seroprevalence of COVID antibodies who are deployed in different geographical locations across the country.

**Address for correspondence:** Dr. Arun Kumar Yadav,
Department of Community Medicine, AFMC, Pune, Maharashtra, India.
E-mail: arunyadavpsm@gmail.com

**Access this article online**

Quick Response Code:

**Website:** www.ijcm.org.in

**DOI:** 10.4103/ijcm.ijcm_1179_21



**How to cite this article:** Basannar DR, Singh S, Yadav J, Yadav AK. Quantifying age heaping and age misreporting in a multicentric survey. Indian J Community Med 2022;47:104-6.

**Received:** 03-09-21, **Accepted:** 12-02-22, **Published:** 16-03-22

## Methods

The study was conducted on the secondary data collected for a multicentric study for serosurvey in Armed Forces.[8] The principal investigator of each center collected the data on the age at each center by asking the completed age in years.

Measurement of errors in age for the sampled data has been evaluated by Whipple's Index (WI), Myer's Blended Index, and United Nations Age–Sex Accuracy Score.

### Whipple's Index

WI is the simplest and most widely used index and is a measure of age heaping on ages ending in 0 or 5.[9] The Whipple score is calculated by dividing the sum of the populations at single ages by one-fifth of the sum of the populations from age 23–62 and multiplied by 100. The total score ranges from 100 to 500, with a score <105 being highly accurate, and it decreases as the score rises, with >175 being very rough. Symbolically,

$$\text{Whipple's index for the 5 - year range} = \frac{\sum(P25 + P30 + P35 + ..... + P60)}{1/5\sum(P23 + P24 + P25 + ..... + P62)} \times 100$$

### Myer's Blended Index

It is one of the most accurate measure heaping toward certain digits by blending the population in a particular way that each terminal digit has an equal sum, based upon the assumption that the population is equally distributed among the different ages.

It is calculated for the age above 10 years and expressed as percentages. The first step is the sum of populations ending in each digit over the whole range starting with the lower limit of the range, for example, 10, 20, 30, 40…; 11, 21, 31, and so on. It is followed by ascertaining the sum excluding the first population combined in the first step, for example, 20, 30, 40.; 21, 31, 41, and so on. The next step is to weigh the sums in steps 1 and 2 and add the results to obtain a blended population. This distribution is then converted into percentages and takes the deviation of each percentage from 10.0, which is the expected value for each percentage. Finally, the summary index of preference for all terminal digits is derived as one-half of the sum of the deviations from 10.0%, each without regard to signs.

The combined score of Myer's Index ranges from 0 to 180, whereas for individual digit, it varies from 0 to 90. Accuracy is inversely proportional to the score. The lower the score, the higher the accuracy and the deviation indicates the heaping.

Symbolically, it is obtained by arranging the matrix of the population in a standard format, as shown in Table 1.[4]

### Age Ratio Score

Age Ratio Score (ARS) is defined here as the ratio of the population in a given age group to one-third the sum of the population in that age group and in the preceding and following groups, multiplied by 100. It is calculated for age up to 74 years. Symbolically,

$$\text{ARS for } 5\,P_a = \frac{{}_5P_a \times 100}{1/3\left[{}_5P_{a-5} + {}_5P_a + {}_5P_{a+5}\right]}$$

When the quality of data is good, there is a gradual change from one age group to another. However, data show fluctuations in the case of poor quality.

### United Nations Age–Sex Accuracy Score

It is mainly used for comparative analysis in cross-sectional studies. There are no minimum or maximum values, and the weight of three attached to sex ratios is arbitrary based on empirical findings. Symbolically,

U. N. Joint Score (JS) =3 (sex reassignment surgery) + Male ARS + Female ARS (FARS).

JS of <20 is considered accurate, 20–40 is inaccurate, and over 40 is highly inaccurate.

For the study, age and sex data were collated in MS Excel, and all the scores and indexes were calculated in MS Excel itself. The two authors independently calculated all the scores. The discrepancy was resolved with mutual discussion.

## Results

Out of total 3252 participants, 828 (25.5%) were female. The mean (standard division) age of our population was 34 (8.5) years and ranged from 15 to 65 years. The age- and sex-wise distribution of the population is shown in Figure 1. The percentage of female ages ending with digits 0 or 5 is 23.55% and percentage of male ages ending with digits 0 or 5 is 23.28%. The calculated WI was 117.75 and 116.34 for males and females, respectively. The calculated Myer's Index for females and males is 10.53 and 25, respectively. The calculated age ratios and sex ratios are shown in Table 2. The Sex Ratio Score, Male Age Ratio Score, FARS, and UN JS Index are 39.99, 27.29, 20.61, and 167.86, respectively.

## Discussion

This study was conducted to ascertain age-heaping bias in a multicentric study among Armed Forces personnel. We found that as per WI, the data quality is fairly accurate and Myer's Index illustrates low age heaping.

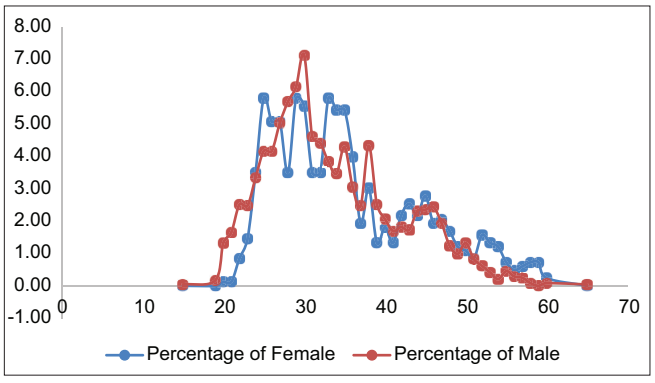


**Figure 1:** Age and sex distribution of study population

**Table 1: Matrix for Myer's Index**

| | Starting at 10+*a* | Starting at 20+*a* | Column 1 | Column 2 | Number (1×3+2×4) | Percent distribution | Deviation of percentage from 10 (6-10) |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | $X_1$ | $Y_1$ | 2 | 8 | $Z_1$ | $Z_1/\sum Z_i$ | $P_1$ |
| 2 | $X_2$ | $Y_2$ | 3 | 7 | $Z_2$ | $Z_2/\sum Z_i$ | $P_2$ |
| - | - | - | - | - | - | - | - |
| - | - | - | - | - | - | - | - |
| 9 | $X_9$ | $Y_9$ | 10 | 0 | $Z_{10}$ | $Z_{10}/\sum Z_i$ | $P_{10}$ |
| Total* | $\sum X_i$ | $\sum Y_i$ | | | $\sum Z_i$ | | $\sum P_i$ |

*Total irrespective of sign

**Table 2: Age ratios and sex ratio**

| Age | Male | Female | Male AR | Female AR | Sex ratio |
|---|---|---|---|---|---|
| 20-24 | 279 | 50 | | | |
| 25-29 | 611 | 209 | 144.10 | 169.23 | 292.3 |
| 30-34 | 569 | 197 | 112.12 | 116.22 | 288.8 |
| 35-39 | 404 | 130 | 100.75 | 92.86 | 310.8 |
| 40-44 | 233 | 83 | 75.04 | 79.05 | 280.7 |
| 45-49 | 217 | 80 | 137.78 | 120.30 | 271.3 |
| 50-54 | 82 | 50 | 67.49 | 93.46 | 164.0 |
| 55-59 | 26 | 27 | 61.18 | 103.85 | 96.3 |
| 60-64 | 3 | 2 | | | |

15-19 added to 20-24 and 65-69 added to 60-64

As mentioned earlier, Armed Forces personnel differ from the census population, so the gradual change in ARS is not expected as all study participants are from the selected age group, i.e., from age 18 to 45 years. In comparison with the National Census Report 2011 which showed poor quality of age data due to heaping, this study illustrates better quality of age data which can contribute to more literacy level of participants compared to average literacy rate of India and the meticulous system of data collection in Armed Forces.

A study conducted among 4304 people in 823 households in a community survey in the Yavatmal district of Maharashtra state showed that the collected age data are of poor quality with age heaping at ages with terminal digits.[5] One's unawareness of its own age manifests as age-heaping bias. The chances of misreporting of age are rarely stated in the research. Although approximation of the age is common, but this approximation is not random or the individual do not choose age randomly but they have a preference for the digits. With some individual's preference is for the digits 0 and 5, while other having preference for even numbers. Hence, age heaping indicates lack of awareness of age in the society, and a strong correlation has been found with age-heaping bias and illiteracy and can also be used as an indicator of human capital.[10]

The misclassification of the age by bias can lead to wrong assessment of denominator in age groups and calculation of wrong epidemiological rates, leading to erroneous planning and policies.

## Conclusion

In our study population, the accuracy of age data was acceptable. This may be because our population is more educated and has yearly events such as Annual Medical Examination, Periodic Medical Examination, and career-oriented courses where individuals would recall their age. Nevertheless, age heaping is less of problem in survey conducted in this particular study, but it is strongly recommended that checking errors while acquiring age data should be done periodically and the researcher should have an inbuilt system in the data collection tool to minimize age rounding/heaping, especially in the community field surveys where age heaping and misreporting can significantly distort the measures of association/treatment effect.

### Financial support and sponsorship

Nil.

### Conflicts of interest

There are no conflicts of interest.

## References

1. Vetter TR, Mascha EJ. Bias, confounding, and interaction: Lions and Tigers, and Bears, Oh My! Anesth Analg 2017;125:1042-8.
2. Bowling A. Research Methods in Health: Investigating Health and Health Services. 4th ed. Buckingham: Open University Press; 2014. p. 536. Available from: https://eprints.soton.ac.uk/371922/. [Last accessed on 2021 Jul 22].
3. Hulley SB, Newman T, Cummings S. The anatomy and physiology of clinical research. In: Designing Clinical Research. 4th ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams&Wilkins; 2013. p. 2-13.
4. Glasser SP. Bias, confounding, and effect modification. In: Essentials of Clinical Research. 2nd ed. Cham, Switzerland: Springer; 2014. p. 362-73.
5. Pardeshi GS. Age heaping and accuracy of age data collected during a community survey in the yavatmal district, maharashtra. Indian J Community Med 2010;35:391-5.
6. Szołtysek M, Poniat R, Gruber S. Age heaping patterns in Mosaic data. Historical Methods. J Quant Interdiscip Hist 2018;51:13-38.
7. Stockwell EG, Wicks JW. Age heaping in recent national censuses. Soc Biol 1974;21:163-7.
8. Ghosh S, Yadav AK, Rajmohan KS, Bhalla S, Sekhawat VS, Prashant J, *et al.* Seropositivity of severe acute respiratory syndrome coronavirus 2 infection among healthcare workers of the Armed Forces medical services, India: A multicentric study. Med J Armed Forces India 2021;77:S359-65.
9. Spoorenberg T, Dutreuilh C. Quality of age reporting: Extension and application of the modified Whipple's index. Population JSTOR;62:729-41.
10. A'Hearn B, Baten J, Crayen D. Quantifying quantitative literacy: Age heaping and the history of human capital. J Econ Hist 2009;69:783-808.