Khalida P. Zaki and M. Javed A. Zaki

A Comparative Study of Age Reporting in Pakistani Censuses and Surveys : 1951-1981[5*]

Introduction

IT is generally recognized that the *statistical* information on population characteristics, presented in censuses and surveys suffer from inaccuracy, in developing as well as developed countries. The age data too are subject to great inaccuracies.

In Pakistan, in every census or survey the question on age was asked of the respondent or head of the household, who reported this information by recalling his or her memory for each member of the household. Alternatively, it is estimated by the enumerator. It is never recorded on the basis of a birth certificate or some other reliable evidence. However, in the presence of inadequate vital registration, majority of people do not possess birth certificates. Therefore, the errors are to be expected in the censuses and surveys. Other factors which contribute to the error in age data are given below : (Shryock and Seigel, 1975).

a. Coverage errors
b. Failure to record ages
c. Misreporting of ages
d. Ignorance of correct ages

*Helpful comments from Elaine Loebner, **Robert D,** Retherford, **Peter** Zingle, Samuel **Baum, Kozo** Ueda and Nasim Iqbal Farooqui **on an** earlier version **of** this paper **are** gratefully acknowledge. **However,** the sole responsibility rests on the authors.

e. A general tendency to report ages, ending on certain "preferred" digits
f. A tendency to exaggerate length of life at advanced ages
g. Misreporting of ages for purely personal reasons.

Age is the most important variable in the study of mortality, fertility, nuptiality and other factors of population change. The importance of census data by age in the studies of population growth is even greater when adequate vital registration system is lacking. Accordingly, age was recommended with higher priority by the United Nations for inclusion in the 1960 and 1970 censuses. The presence of errors in the age data may vitiate any projections, demographic estimates or judgements based on these data.

The purpose of the present study is to explore the nature and extent of errors jn age data for Pakistan. The analysis will be based on comparisons of several censuses and surveys conducted in Pakistan since 1951.[1]

The quality of data from surveys is considered to be better than that of censuses as it is argued that surveys in general, utilize the services of trained and well paid enumerators. Since census is a gigantic activity, it does not enjoy as good a supervision and commitment from its staff as do the surveys. The relative judgements about the quality of data from censuses and surveys will also be made on the basis of the above comparison. The pattern of age reporting for males and females, levels, trends and differentials in reporting of ages, will also be discussed.

The first population census of Pakistan was conducted in 1951, the second in 1961 and the third in 1972. The latest census was undertaken in March, 1981.[2] All the censuses adopted a uniform question on age. The question on age was simple. For example, in 1961 census it was pharsed : "what is your age on 31st of January 1961?" and the answers were recorded in completed years.[3]

There were a number of improvements made in the census taking procedures in 1972 and 1981 over that of 1961 and 1951. Enumerators and their supervisors were very thoroughly trained and special manuals of instructions for enumerators and supervisors were issued. Simpler questionnaires were used in 1972 and 1981 as compared to the earlier censuses. The use of modern and scientific

1. However, 1951 census data are analysed at the national level only, because it becomes a job loo cumbersome to adjust for geographic boundry changes 10 bring it up in line with 1961 and 1972 provincial figures by 5 years of age groups.

2. **The figures from** 1981 census arc only available for 5-years of age groups. (Census **Bulletin number-7).** 1983

3. The explanation of completed years was given a; follows : If a person was 20 years and 10 months on 31st January 1961, was entered as 20 years not 21 years. In cases where ages were not known, were estimated by the enumerators by refering to any past event. It generally helps in the estimation of ages to enumerate the children in a household starring from the youngest child.

methods of collecting and presenting data was made, which was expected to lead to improve the quality and reliability of data.

One would also be interested to analysis the differences in the accuracy of age reporting on the basis of the question on age; that is, whether the question on age was based on completed years or on the year of birth. Since all the surveys and censuses included in this study recorded age in completed years except the Pakistan Fertility Surveys (PFS) (1975), where the question on age was based on year of *birth,* the possibility to make any generalization in limited.

There are very few studies previously done in Pakistan on this topic. The present study is quite unique and extensive in nature because several (fourteen) data sets of single year age returns or five year age intervals have been examined both for males and females. This study may be expected to contribute considerably to the existing knowledge of age data in Pakistan.

The Data

The data base comprises the censues of population undertaken in Pakistan in 1951, 1961,1972 and 1981, Population Growth Estimation Projects (PGE), 1962 through 1965, Population Growth Survey (PGS), 1968, 1969, 1971 and 1976, Housing, Economic and Demographic Survey (HED), 1973 and Pakistan Fertility Survey (PFS) of 1975.[4]

Techniques of Analysis

A number of techniques of analysis can be used to identify the errors in age data, tabulated by single year or 5 year age groups. The method used in the current analysis for grouped data involved the study of sex ratios and age ratios by sex leading towards the computation of United Nations Age Sex Accuracy

4. To *maintain* the geographic ident'ty between censuses and several surveys, the population of the Provincially Administered *Tribal Areas* (PATA) of North West *Frontier* Province, Centrally Administered Tribal Areas (CATA) and Federal Capital Territory of Islamabad (FCT1) have not been included in this analysis. However, the population of Tribal Areas adjoining Hazara district have been included in the censuses of 1961 and 1972. The given ago data for 1951 also excluded 2,666,378 persons of Agencies and Special Added Areas under the Deputy Commissioners in the district of Peshawar and Dera Ismail Khan in the Frontier Regions. Therefore, this study represents Pakistan comprising of its four provinces!.i.e, NWFP *(excluding* Malakand Division), *Punjab Sind* and Baluchistan, representing about 93 percent of the total population of Pakistan. The population total for Pakistan in 1961 and 1972 were obtained by summing up each age group and sex group from the available District Census Reports (DCRs). This was necessitated due to the unavailability of the 1972 census age distribution corresponding to the 1961 distribution at the national and at provincial levels ai the time of this study.

Index (UN Index). For single year data, the well known methods such as, Whipple's and Myers' indices have been computed to examine the overall inaccuracy and digital preference in the Pakistani age data[5]. The above indices determine the quality of data through the description of a certain range. According to UN Index, the data will be categorized as "accurate", inaccurate" or highly inaccurate" if the UN Index is below 20, 20-24 or above 40 respectively. Whipple's Index varies between 100, representing no preference for digits "0" or "5", and 500, indicating that only digits "0" and "5" were reported. The theoretical range of Myers' Index is between 0 to 180, 0 representing no age heaping and 180 indicating that all ages were reported at a single digit[6]. Generally the higher the value of these indices the greater the 'digital perference' or 'age heaping' in the data will be. There are several other methods. (See Shryock and Seigel (1975) for the definitions and computational procedures involved in these techniques.)

Results

The following section presents the results of the current research in two parts. Part I describes the findings for the data grouped by 5-years of age groups, and Part II deals with indices based on single year age distributions. Part III presents a small exercise on effects of age errors on mortality and fertility estimates in PGS 1976.

Findings for Grouped Data

Percentage Distributions

The reported percentages by main age groups for Pakistani males and females are given in Table 1. This Table is based on several censuses and surveys for 1951 to 1981 period. Age pyramids based on the 1961 and 1972 census returns have broad bases indicating of a high birth rate in the country. It is seen from Table 1, that about 40 to 46 percent of males and females are below age 15. Censuses and surveys show somewhat similar picture. Table 1 shows that the

5. There arc two other techniques developed by Bachi and Ramachandran (2 and 10). Bachi's Index is extended form of Whippie's Index. Rachi's index is applied to examine the concentration on each terminal digit, on which ages have been reported. Ramachandran's technique is also similar to Myers' method. Bachi's and Ramachandran indices are not widely used for the reasons: that they are too complicated whereas Whippiles and Myers'Indices are relatively simpler. Therefore, only Whipple's and Myers'methods are used in this study and fulfill our purpose quite well.

6. Sometimes Myers' Index is divided by a number 2 which is termed as Myers' summary index: the theoretical range for this index would be to 90.

TABLE 1 -PERCENTAGE OF MALES AND FEMALES IN MAIN AGE GROUPS IN PAKISTANI, CENSUSES AND SURVEYS, 1951-1981

Main groups by sex	*1951*	*1961*	*1972*	*1981*	*1962*	*1963*	*1964*	*1965*	*1968*	*1969*	*1971*	*1976*	*1973*	*1975*
	Censuses[a] 1951 1961 1912 1981				*PGE*	*PGE*	*PGE*	*PGE[b]*	*PCS*	*PGS*	*PGS*	*PGS[c]*	*HED**	*PFS[a]*
Males														
Pre-School (0-4)	13.09	15.61	14:11	15.11	16.33	16.29	15.00	15.17	15.63	**15.90**	**16.02**	15.09	12.70	15.33
Under age 15	42.19	41.86	43.10	44.55	42.91	43.17	41.78	42.89	43.32	44.43	45.55	45.17	42.42	44.33
Working age (15-59)	52.1S	50.73	49.33	48.17	48.84	49.81	51.55	50.81	49.63	47.97	46.53	47.43	49-36	41.96 (15-49)
Old age (60+)	5.70	7,40	7.56	7.30	7.27	7.02	6.68	6.31	6.96	7.24	7.92	7.40	8.04	13.72 (50+)
Females														
a. Pre-school (0-4)	14.50	17.29	!6.05	17.08	17.66	17.31	**16.06**	16.12	16.25	16,49	**16,72**	15,42	15.05	15.42
Under age '5	44.46	43,17	44,45	45.82	43.38	43.41	43.29	43.32	43.58	44.19	45.65	44.49	44.23	44.30
Working age (15-59)	49.7?	50.43	49.13	48.13	50.46	50.75	50.79	50.79	50.12	49.59	47.85	48.85	49.54	44.23 (15-49)
Oid Age (60+)	5.68	6.42	6.37	6.05	6.15	5.82	**5.92**	**5.93**	5.98	6,22	6.50	6.66	6.23	11,21 (50+)
b. **Females of Reproductive** age (15-49)	**43.97**	44,98	43.88	42.92	45.31	45.54	**45.50**	45.64	44.50	44.24	**41,60**	43.37	43.78	44.49
Age Dependency Ratio[1]	95.78	98.64	103.01	107.73	**99.50**	98.99	**95.32**	96.88	95.27	104.4)	111.82	107.87	102,07	**131,62** (50+)

SOURCES : *a :* **Pakistan Census Organisation,** *b :* **PIDE, 1971,** *c :* **Statistical Division, d: Population council 1976.**

1. **Percentages are based on total of males and females separately.**
2. **Age Dependency ratio is defined as follows; ADR=Population of age 15 and 60+ divided by population ages (15-59), multiplied by ICO.**

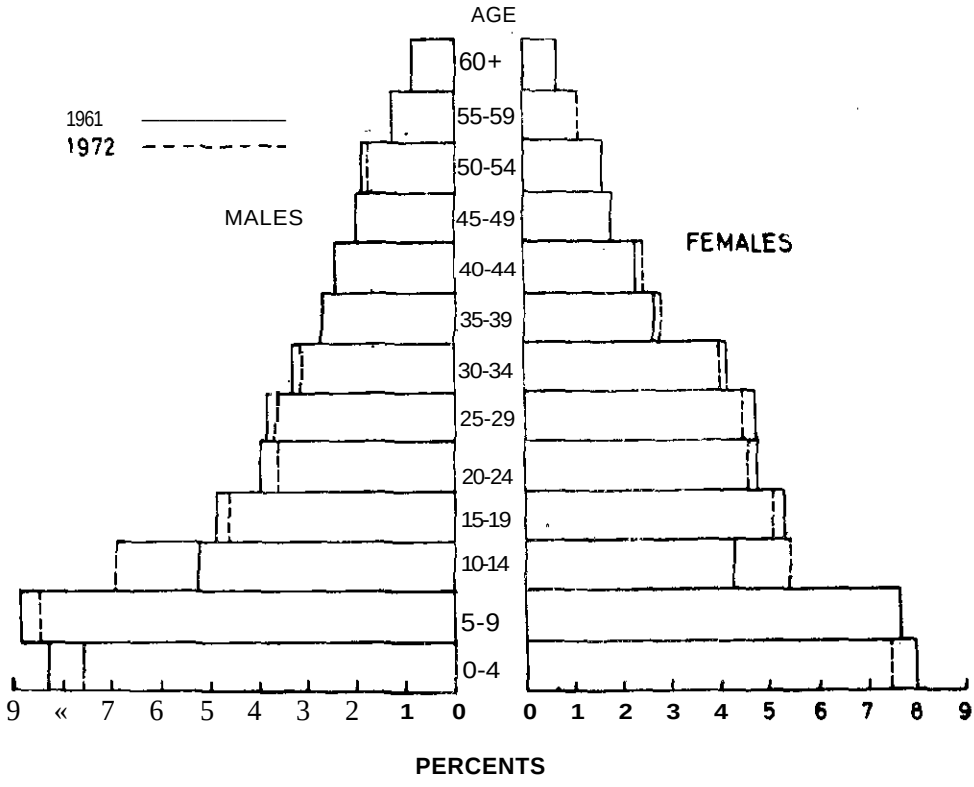

Fig. 1. Percent of total population in age groups.

population of males and females in pre-school group (0-4) have been varying since 1951 through 1981. Their proportion is lower both for males and females in 1972 (census) than in 1961 or 1981 (census). PGE (1962-64) showed a declining trend, whereas PGS (1968-1971) indicated a larger size of population, as compared to the PGS 1976. Therefore, it is difficult to determine whether fertility has begun to decline in Pakistan, because the cummulative percent of males and females under ago 15 ranges between 40 to 46 percent and has been roughly stable throughout. The proportion below 15 is greater in 1981 census than it was in 1961 and 1972 censuses.

The proportions in age groups 15-59 were lower in 1981 (census) but higher in 1976 (PGS), whereas they had gone up during PGE periods (1962-1964). The old age groups (60-1) have also been fairly stable except in the 1951 census. Similar has been the comparative position of females of reproductive age; their proportions varied during 1962-1964 (PGE), increased in 1976 (PGS) and then

declined in 1981. *The* age dependency ratio rose from 96 in 1951 to 108 in 1981 census. It was the highest in 1971 (PGS), but declined to 108 in 1976 (PGS).

On the basis of the above findings it is very difficult to generalize about the age structure of population in Pakistan. Hence, the need for an evaluation of the quality of data.

An analysis of Pakistani data focused on age exaggeration was done by Retherford and Mirza (1981). They tested the hypothesis exaggeration of age, which increases with age. (Their purpose was to make own fertility estimates). They concluded that ages of children are reported very inaccurately in Pakistan. They observed heaping on ages 8, 10 and 12. A large deficit of children at age 2 was also observed. They suspected a rapid fertility decline, an unusually high undercount or an exaggeration of ages, or some combination of these. They argued that rapid fertility decline is very unlikely because there was no evidence of increases of contraceptive use between 1968 to 1975 (Shah, 1979 as quoted in 11). They further argued that age heaping of women have systematic upward bias, in the form of age exaggeration which increases with age. We have found this to be true in our study, for both males and females. They forwarded another explanation that age exaggeration for high parity women i.e., the young women with children tend to be shifted up from age group 15-19 into 20-24. This could be a plausible explanation for excessess at older ages relative to the stable ages. To study this phenomena the age distribution of women with parity one in HED (1973) and PFS (1975) were investigated by the present authors which are shown in Appendix Table 1. It has been observed that women aged 15-19 seem to be reported in next higher age group.[7]

Sex and Age Ratios

(a) SEX RATIOS. The sex ratios by age for several Pakistani censuses and surveys are shown in Appendix Table 2. They indicate that for Pakistan, in general, sex ratios have remained high (excess of males). In the census of 1972, from age group 0-4 the sex ratios (101) rose to a level of 127 in age groups 10-14, whereas this level was exceeded by the age group 15-19 (136) in 1964. This may be due to a systematic bias in age reporting; in a majority of surveys, the highest sex ratios are found in age group 15-19, In PFS sex ratios appear constant from age group 0-4 (108) to age group 15-19 and a deficit of males emerges (sex ratios below 100) in ages 20-34. Both in censuses and surveys the sex ratios exhibited extreme fluctuations ranging between 103 to 150 beyond the 15-19 years age group. However, minimum fluctuations were observed in 1976 (PGS) and 1981 (census).

7. **We will** see **later** that the **HED data** are more erroneous than **PFS.**

However, age to age fluctuations in sex ratios existed in all the censuses and surveys, at the national and provincial levels. In 1981, for most age groups sex ratios were below the level of 1961 and 1972; this may be an indication of better enumeration or reporting of females in 1981. The 1951 figures must have been affected by the migration at the time of partition of the Indo-Pakistan sub-continent involving violent killings. The higher sex ratios at ages 10-24 can partly be explained in terms of general understatement of the ages of older boys and young men resulting in an inflation of this age group or female mortality in the earlier child-bearing ages. Another reason may be that the women with children were reported in next higher age groups. The sex ratios peaks at the older ages are indicative of the errors in reporting and are completely against the phenomena of male/female selective mortality. Generally, persons in older ages tend to report their ages higher because socially and culturally greater respect is attached to older people. In our case, beyond age forty a saw tooth pattern of sex ratios was observed, both in censuses and surveys. The high sex ratios have also been suspected by Bean, 1974 and Burki, 1973. They described the unusually high sex ratios as the defects in data. Many other demographers (Afzal, 1973 arid Krotki, [1963] agreed that the 1961 census suffered from underenumeration. The inconsistencies of 1972 census data may therefore, seem to arise from overenumeration. The 1981 sex ratios appeared to be less erroreous relative to the previous censuses.

Sex Ratio Scores (SRS)

This summary measure gives the average of successive differences (irrespective of sign) in sex ratios by age. The lower the SRS, the better the overall quality of data. For Pakistan the range of the SRS is 5.0-14-3; the scores suggest that the 1961 and 1981 census data sets are of better quality relative to the other censuses and surveys of this study (See Table 2 and also appendix Table 2). However, according to these scores PGS of 1968, 1969 and 1976 can be classified as moderately better than PGE (1962-65), HED (1973) and PFS (1975).

(b) AGE RATIOS. The age ratios are presented in Appendix Tables 3 for males and 4 for females for censuses and surveys in Pakistan. For males and females, the age ratios gave characteristic fluctuations in both censuses and surveys. These ratios are high at ages 5-9, 25-29, 40-44 and 50-54 for both males and females, with some variations in 1951. These ratios were relatively higher at ages 20-24,30-34 in PGS for females and consistant in all the three surveys years i.e,, 1968,1969 and 1971 but somewhat different picture emerged in 1976. The high age ratios mean that excessive number of persons were reported in the age groups mentioned above, by comparison with the numbers in the groups just higher or lower. In the absence of past violent experience of births, deaths and migration and with perfect age reporting, an age ratio for a given age group should approximate to 100 in any population.

TABLE 2—SUMMARY MEASURES OF QUALITY OF GROUPED DATA IN PAKISTANI CENSUSES AND SURVEYS, 1951-1931

Year	Sex Ratio Scores	Indices for grouped data Age Ratio Scores		UN Index[1]
		Males	Females	
Census				
1951	7.0	13.3	12.7	47.0
1961	5.0	11.7	11.8	3S.5
1972	8.6	12,2	10.1	48.1
1981	*6.1*	*5.9*	5.6	29.7
Survey				
PGE 1962	13.9	10.4	16.8	68.9
PGE 1963	14.3	10.2	14.2	67.3
PGE 1964	10.2	10.8	11.8	53.3
PGE 1965	11.3	12.5	10.1	56.5
PGS 1968	6.6	10.7	15.0	45.6
PGS 1969	9.1	13.4	10.2	50.9
PGS 1971	13.2	9.7	9.1	50.4
HED 1973	13.0	12.1	7.3	58.4
PFS 1975	*9.4*	5.5	*8.7*	*42.4*
PGS 1976	8.6	7.9	*5.9*	39.7

SOURCES : [See Source in Table 1. and Appendix Tables *2*, 3 and 4],

1. This index shows the combind effect of male-female age mis-reporting.

In 1972, the age ratios were high for the age groups 10-14 for males. For females, a pattern similar to that of 1961 existed except the age group 15-19. Above age 20 the mis-reporting of ages is almost identical in the 1961 and 1972 censuses. They deviate relatively less from 100 in 1981 for both the sexes. **PGE**

projects for the years 1962-65 reflected for both the sexes that the age ratios arc consistent except in respect of the age 35-49 for males. They show fluctuations by age in all the four years of survey. A similar pattern existed for females during *the* same period (1962-1965), The overreporting for males and females is indicated in HED (1973) and PFS (1975) for age groups 5-9, 10-14, 25-29 and 40-44 and underreporting (age ratio below 100) at adjacent ages. The smaller deviations in age ratios of males and females are found in PGS relative to the censuses and other surveys.

Age Ratio Scores (ARS)

The ARS is an average of the deviations of age ratios from 100 for all age groups. It is affected by the very high age ratios. It could, therefore, be mis-leading index and must not be presented independent of age ratios. Overreporting is observed in age groups mostly beginning with fives and under reporting in the adjacent age groups in all the censuses and surveys for both the sexes. No improvement is observed for males on the basis of age ratio scores. The age reporting appears to be better for females in 1976 PGS and in 1981 Census. HED 1973 shows increased overreporting for males whereas these scores are quite low for females in HED compared to the Census of 1972. HED survey gives worse results than the previous surveys and censuses (See Table. 2).

Overall Quality of Grouped Data

United Nations Standard to describe census age sex data as "accurate" "inaccurate", or "highly inaccurate," depends upon whether the UN Index is under 20, 20-40 or above 40, respectively. By this classification the Pakistani data appear to be "inaccurate" in 1961 and 1981 censuses, and in 1976 PGS. They are "highly inaccurate" in respect of the 1951 and 1972 censuses and other surveys of this study.[8] These figures are presented at the national level in Table 2, The overall range of UN Index is 29.7-68.9 (for censuses and surveys).

The UN Index categorises the Censuses of 1961 and 1981, and PGS of 1976 as of better quality, relative to the 1951 and 1972 censuses and PGE, other PGS and HED surveys. This index amounted to 29.7 and 39.7 in the 1981 census and PGS 1976, respectively. The considerable decline of UN index in 1981 is an apparent *indication of improvement. For PGS it ranged* from 39.7 to 50.9. The UN Index for HED was 58.4.

8. **It must** be noted that the highest **age** in **PFS was recorded** as **50 above, whereas, all other** data **sources** were **analysed** upto **age 60** and **above,** hence the results must be **compared with caution.**

Quality of Pakistani Grouped Data Compared with Other Countries

UN Index for Pakistan and a few other developed and developing countries have been examined by the present authors.[9] According to the UN Index the United States (12.2), Sweden (15.1) and Japan (19. i) have accurate data sets on age. For these countries, at such an advanced stage of economic, social and scientific development alongwith high literacy, one would expect the accuracy in statistical reporting to this extent; they are also not 100 percent correct either. The comparisons reveal that the latest Pakistani data (Census 1981 and PGS 1976) are of the same quality as of the Philippine (Census, 1960) and Greece (Census, 1961). Furthermore, it is observed that the inaccuracy of data is not only inherited by Pakistani censuses and surveys, the worse are data for other countries, namely; Iran, Kuwait, Nigeria, Ivory Coast, Taiwan, Java and Bangladesh. Therefore, it is concluded that accuracy of data is expected to improve in the long run alongwith socio-economic and educational development and after the significance attached to age or to any type of statistical reporting is well understood.

Single year data

(1) Single Year Reported Age Distributions 1972

When data are grouped by 5 years or broad age groups, major inconsistencies, arising from 'age heaping' or 'digital preference' disappear or smooth out. Therefore, the examination of single year data is useful. Single year reported age distributions of males and females for Pakistan in 1972 (Census)'[0] have been shown through an age pyramid (Figure 2). The digital preference has been obsev-ed i.e., the peaks and troughs are on ages ending on digits '0' and '5', they become more pronounced with advancing age. The similar battern of digital preference exists at the provincial levels (1972) and for the surveys' data of this tudy.

(2) *'Blended Population' Distributions and 'Digital Preference'*

An examination of percent distribution of 'blended' population based on Myers' method by sex was also performed. The blended percentages give the degree of concentration on certain digits.[11]

9. Many of indices **were** computed by the present authors, **and some of them were these** taken from (13).

1O. The single years age disfributions from *1981* **census** are **not** available yet.

11. These **percentages cover a certain** age range **ia** our case the **age** was **10-79 years.**

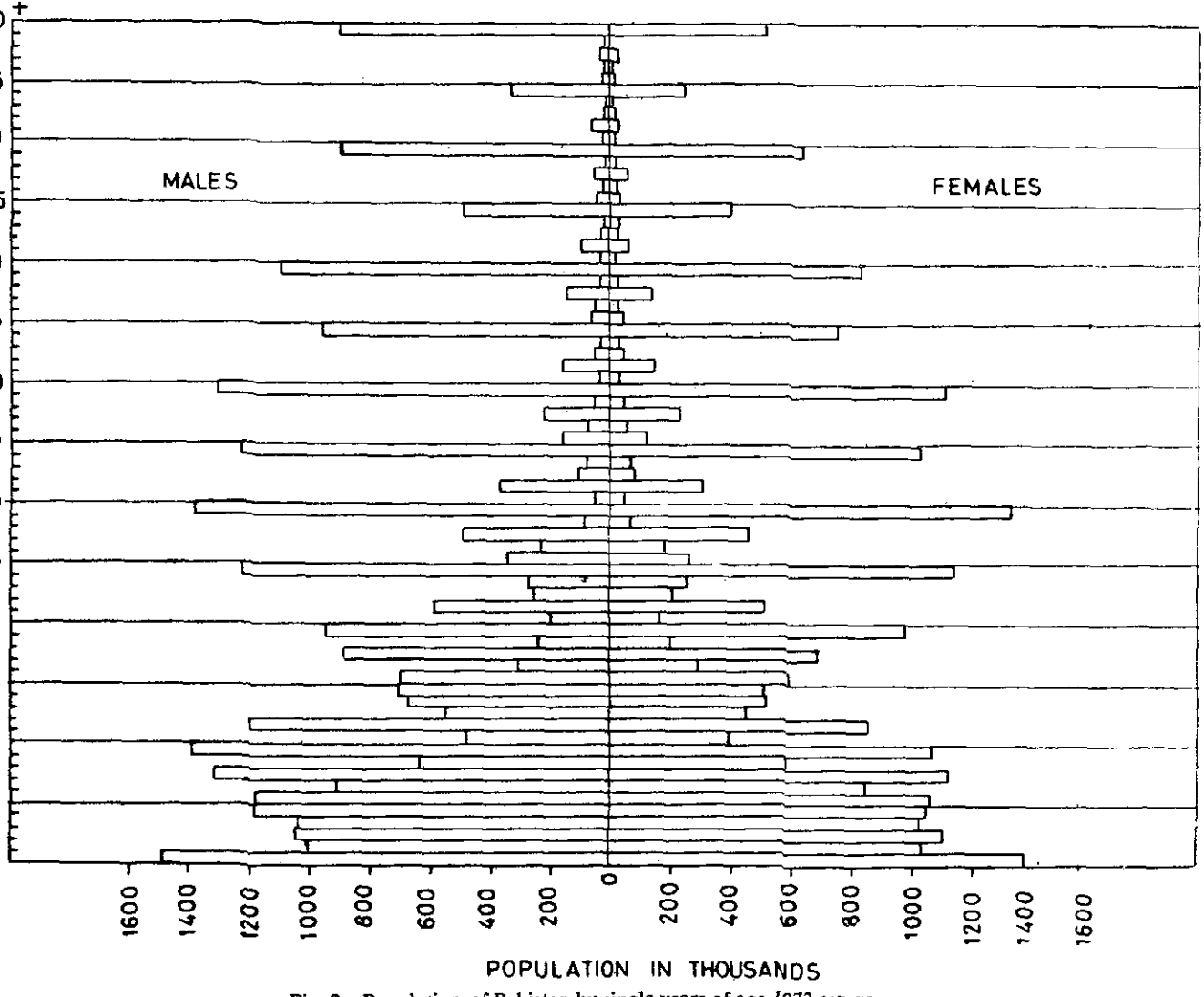

Fig. 2. Population of Pakistan by single years of age *1972 census*.

TABLE 3—PERCENT DISTRIBUTION OF MYERS' BLENDED POPULATION BY SEX IN PAKISTAN, IN SELECTED SURVEYS AND **IN** CENSUS, 1972

Terminal Digit	*PGE.-I962*		*PGS-1968*		*Census-1972*	
	Males	*Females*	*Males*	*Females*	*Males*	*Females*
0	**33.22**	**35.40**	**33.53**	**34.10**	**33.76**	**34.10**
1	**1.94**	**1.59**	**2.12**	**2.38**	**2.40**	**2.24**
2	**9.60**	**8.90**	**9.57**	**9.42**	**8.92**	**8.43**
3	**3.40**	**3.26**	**3.S7**	**3.89**	**3.69**	**3.64**
4	**4.04**	**4.12**	**4.25**	**4.53**	**4.13**	**4.32**
5	**27.46**	**25.53**	**29.91**	**24.00**	**25.82**	**25.44**
6	***6.40***	***6.22***	**6.25**	**6.63**	**6.04**	**5.72**
7	**3.32**	**3.60**	**3.37**	**3.47**	**3.S1**	**3.41**
8	**8.60**	**9.05**	**9.17**	**9.41**	**9.36**	**10.31**
9	**2.12**	**2.33**	**1.97**	**2.18**	**2.37**	**2.39**
Total	**100.00**	*100.00*	**100.00**	**100.00**	**100.00**	**100.00**

SOURCES : [See Source **in Table 1].**

It is revealed that in PGE (1962-65), PGS (1968-76) and census of 1972, 53 to 60 percent of the total population is reported with ages ending with digits '0' and '5' About 17 to 20 percent concentrated on digits 6, 4, 3 and 7. Only 4 percent was found on the least preferred digits '1' and '9'.

For illustration purposes, the above percentages for Pakistani males and females, are given in Table *3* for PGE (1962), PGS (1968) and the 1972 census.[12] It is quite evident from these digital distributions that ages for the majority of males and females are reported with terminal digits '0' and '5'- About 33 to 34 percents of males reported ages ending with '0'; the corresponding range for females is 34 to 35. The second preferred digit is '5' which has been occupied by about 26 to 30 percent of total males. The next preferred digits in the order

12. These years for PGE and PGS have been selected only on arbitrary basis.

of importance are 2, 8, 6 and 4 for both the sexes. The least preferred are digits '9' and 'T. Ages of males and females have been reported in a similar fashion. This pattern was quite a consistent pattern over the time period, 1962-72. However, the heaping is relatively lesser in 1976 PGS.

Overall Quality of Single- Tear Data

The overall judgement for the quality of single year age data will be based on Whipple's and Myers' indices. The critical range for Whipple's index is 100

TABLE 4—SUMMARY MEASURES OF QUALITY OF SINGLE YEAR DATA IN PAKISTAN CENSUSES AND SURVEYS*

	Whipp le's Index		*Myers' Index*	
	Males	*Females*	*Males*	*Females*
Census				
1951	—	—	—	—
1961	—	—	—	—
1972	282	282	79.2	79.7
Surveys				
PGE 1962	287	283	81.3	81.6
PGE *1963*	269	268	71.2	72.0
PGE 1964	274	274	76.9	78.6
PGE 1965	250	243	66.0	65.2
PGS 1968	283	27S	81.1	75.9
PGS 1969	236	285	80.9	79.7
PGS 1971	275	261	75.9	72.0
HED 1973	—	—	—	
PFS 1975	—	—	—	-
PGS 1976	252	174	67.0	65.0

SOURCES ; [See source in Table I].

1. The blank space indicate that the data were not produced by single year of age for these years.

to 500 whereas Myers' ranges between 0 to 180. The higher the value of the index, the greater the overall age mis-reporting due to digital preference or age heaping. The Whipple's and Myers' indices have been presented in Table 4 for the PGE (1962-65), PGS (1968-76) and for the *1972* census, by sex. The *range* of Whipple's index for PGE (1962-65) is 243-287, and for Myers' index is 65-82. These ranges include both males and females. From the above range, PGE 1965 data were categoricsed as the best. The second best among the PGE series appears to be the data for the year 1963. For PGS series (1968-76), the 1976 shows relatively less age mis-reporting. For PGE, both the indices declined over time for each sex, except in 1964. The sampling coverage may be the cause of this difference. However, the considerable *low value of 1965 relative to other* years is questionable. A decline of the two indices is also shown by PGS data over 1968-76, barring 1969. This behaviour of the surveys can be attributed to sampling and non-sampling errors or to methodological differences.

The values of the Whipples' and Myers' for the 1972 census are around 282 and 79 respectively. Males and females do not show any differentials in age reporting. Generally speaking, in majority of the data sources used, females indices are below the male indices, which is an indication of better *female* age reporting relative to males. However, quality of data for Pakistan has improved, Relatively a greater improvement for females has been observed. The 1972 census data by single year of age are not strictly speaking comparable with those of any other census. The 1972 census indices are higher relative to the majority of the surveys (1962-1971). Significantly, the 1981 census when compared with other censuses, for grouped data, exhibited a remarkable improvement. The PGS (1968-71), except the PGS 1976, showed no improvement over PGE (1962-1965). The 1965 PGE and 1976 PGS appear to be of the same quality,

Adjustment of Age Errors and Their Effect on *Fertility* and Mortality Rates, in Pakistan

(a) *Methods of Adjustment*

This section presents a brief discussion on currently used methods based on mathematical graduation of age data (Shryock and Seigel, 1975).

The three commonly known of such techniques are the following:

1. The Carrier-Farrage ratio method.
2. The Quadratic Interpolation method.
3. The Osculatory Interpolation method.

Shryock and Seigel applied these techniques to female age distribution of Ghana (1960)' They observed that the results obtained by the use of above techniques are similar. In this section we have applied the Carrier-Farrage ratio

method and the quadratic interpolation method. These two methods are extensively used and are expected to give satisfactory results.[13] The oscillatory interpolation method provides inadequate results when applied to the data which had greatly suffered from digital preference (Carrier-Farrage, 1959; 264).

The figures obtained by such methods can also be compared with the original counts to indicate the extent of age misreporting by 5-year age groups by each method. The estimated figures can also be substituted for reported figures as the estimated figures are corrected for net age misreporting.

(b) *Effect of Age Misreporting on Fertility and Mortality Rates in Pakistan, PGS 1976*

There are studies where distortions in age specific fertility (ASFR) and age specific mortality (ASDR) estimates have been observed in Pakistan; which may have resulted from age misreporting. Karim and Alam, (1975) examined the effect of age misreporting on fertility estimates of PGE (1962-1965) and PGS (1968-1971). They used stable population age distribution to demonstrate the discrepencies of reported age distribution. They assumed three fertility patterns (low, average and high) for Pakistan. They do not observe considerable differences in the ASFRs but in the TFRs. They concluded that PGE and PGS age distributions were simliar; hence there was not much difference in ASFR series. Any how, the substantial differences in ASFRs were observed when they were based on stable age distribution, under the assumption that the stable age distribution represents the 'real' picture of Pakistani age distribution. However, they pointed towards a systematic age pattern, existing in Pakistan. The ASFRs from Karim and Alam are in Table 5.

Retherford and Mirza (1981) analysed the 'own children' fertility estimates based on 1973, HED; mortality estimates of 1962-65, PGE., and P/F ratio fertility estimates based on 1971 PGS and 1975 PFS. They have detected systematic distortions in the above estimates also which, according to them, "are plausibly explained by a pattern of age exaggeration that increases with age". Age pattern of fertility was examined (see Table 5) and it was found that an implausible fertility pattern was observed due to the age heaping of women, which is systematically biased upwardly particularly, at age 30. Age exaggeration among older women raises estimated age specific birth rates because their ages were reported higher than what they were. They further compared the own children estimates by marital fertility with the marital fertility of Hutterites, (an American religious sect with the highest fertility over all marital fertility on record). It showed that Pakistani marital fertility exceeded the Hutterites at ages 15-19

13. There are a number of other methods also e. g., graphic methods, Use of stable-model age distributions, for details See Shryock and Seigel, 1975.

TABLE 5—AGE SPECIFIC FERTILITY RATES (ASFRs) AND TOTAL FERTILITY RATES (TFRs) REPORTED AND ADJUSTED, IN PGE, PGS AND PFS, IN PAKISTAN (1962-1976)

	*Own-children estimates**		*P/F ratio estimates**				*Record adjusted estimates***				*Mathematically graduated estimates****		
	7962-66	*1967-71*	*1971 PGS (Ua)*	*1971 PGS (a)*	*1975 PFS*	7975 *PFS (a)* (R)	*PGE (a)*	*1962-65 (X)*	*<a) PGS*	*(Reported) 1968-71*	*PGS 1976 Method I*	*PGS 1976 Method*	*PGS 1976 I*
15-19	142	149	61	105	1OS	148	75	70	55	58	56	55	54
20-24	290	324	240	300	278	308	242	170	219	170	271	281	287
25-29	305	328	263	326	330	354	268	247	265	234	348	338	33S
30-34	276	280	267	325	279	295	254	248	246	238	305	316	315
35-39	217	209	218	262	199	206	188	194	198	199	226	224	223
40-44	153	143	144	172	97	96	100	110	123	135	128	129	129
45-49	101	86	92	103	9	8	73	75	71	78	73	72	73
TFR	7.4	7.6	6.4	7.9	6.5	7.1	6.0	5.6	5.9	5.6	6.5	7.1	7.1

*SOURCES : [Table I, Table 2 and Table 6]; '*Ua*' and '*a*' refer to unadjusted and adjusted respectively.
**[Table 2, Table *4* and Table 6 (with higher fertility assumption) Karim and Alam, [6].
***[Present authors] : Methods I and II refer to Carrier-Farrage Ratio Method and quadratic interpolation method, respectively.

and 45-49, in general, and falling short at intermediate ages. Such exceedingly higher marital fertility rates at ages 45-49 are highly improbable, but they could occur if the true ages of older women are younger than reported ages.

The adjusted ASFRs using the PGS 1976 data, by the present authors, also appear in Table 5. We observed the differentials in the series of ASFRs between ages 20 to 34, heading to the raising of TFRs from 6,5 to 7,1, Therefore, it is concluded that the females age misreporting can cause considerable effect on the TFRs. It is further suspected that women having children are somewhat overreported. HED and PFS age distribution of females below age 35 with parity 1, also indicated that in HED age group 20-24 seem to be relatively overstated because the PFS data were categorized of better quality than HED. (See appendix Table-1).

Appendix Tables-5 and 6, present the reported and adjusted agedistributions along with ASDRs for males and ASDR.s and ASFRs for females. It has been observed from these tables that there does not exist any considerable differences in ASDRs when computed on the basis of adjusted age distributions for males and females both. There seem to arise some differences in ASDRs beyond age 40. This may have been due to the phenomenon of age exaggeration that increases with age. The similar observations and explanations were provided by Retherford and Mirza (1981).

However, it is suggested that the corrections of the census or survey counts should be made before the publications of the final rabies. The correction of already published census/survey tables should be avoided as far as possible.

Suggestions

Age is of the fundamental significance in demographic statistics. Information on demographic statistics are mostly tabulated and presented by age groups of population. Age plays an important role in the studies of population change and its component factors i.e., fertility, mortality, migration and nuptiality. Economic growth of a country is directly related to the population structure of the country as it determines the level of labour supply and on the basis of these, projections for future can be made. The whole planning framework for better standard of living of masses, public facilities and health programs, labour force participation, economic dependency ratios, school enrolments etc involve population statistics by age. Therefore, the economic planners and policy makers are cautioned that all the estimates based on the erroneous data would be mis-leading and hence hamper the process of economic growth. In the absence of adequate vital registration system, it is emphatically suggested that the efforts towards the evaluation of existing data sources with regard to its validity and reliability in terms of its accuracy must be encouraged.

Some improvements in age data can be expected by improving the format of question on age; whether the question is based on "year of birth" or on age "in completed years". In the censuses and surveys of this research the age was asked "in completed years" except the PFS, where the question was based on "year of birth". If the question on age is a combination of "year of birth" with reference to historical events, age "in completed years" might give better results. Though the PFS case does not show improvement, but it is relatively better than several other sources. So in some cases "year of birth" may be a better way to ask question on age because the age "in completed years" changes every year and people may not be keeping up the counting of age correctly, and they may be able to recaff the "year of birth" because it is one point of time reference. In case of Pakistan, it has been observed, that the phrasing of question on age does not make much difference. However, the question on age should be very simple so that a layman is able to understand and respond without getting confused. Any carelessness on behalf of the respondents and enumerators would produce errors in data.

The long run solution to the problems of age mis-reporting are expected alongwith the socio-economic development and improvements in the literacy level.[14] There should be compulsory registration system with emphasis on numerical age. In a developing country like Pakistan, these changes are not likely to come in the near future. In the mean time, there must be continuous effort from the government to improve the quality of data to encourage the concerned agencies, to keep the birth records through political representatives, religious leaders, hospitals and health clinics, police stations and post offices etc.

References

1. Afzal, Mohammad, 1973, 1972 Census, population expected and actual, *The Pakistan Development Review*, 12(2), 123-134.
2. **Bachi, Roberto,** (954, Measurement of the tendency to round off age returns, *Bulletin of the International Statistical Institute* (Proceedings of the 28th Session, Rome, September 6-16, 1953), 34(3), 128-138.
3. Bean, Lee L., 1974, The population of Pakistan – An evaluation of recent statistical data, *The Middle East Journal.*
4. Burki, Shahid Javed, 1973, Pakistan's population; ready or not, here they come, ***Pakistan*** *Economist,* October 30.
5. Hashmi. Sultan S, and Iqbal, Alam, 1969, *The Problem of Obtaining Age Data in Pakistan*, Karachi, The Pakistan Institute of Development Economics (Research Report No. 83).

14. An analysis of age reporting by literacy at the district level has shown a positive correlation between the two variables.

Some improvements in age data can be expected by improving the format of question on age; whether the question is based on "year of birth" or on age "in completed years". In the censuses and surveys of this research the age was asked "in completed years" except the PFS, where the question was based on "year of birth". If the question on age is a combination of "year of birth" with reference to historical events, age "in completed years" might give better results. Though the PFS case does not show improvement, but it is relatively better than several other sources. So in some cases "year of birth" may be a better way to ask question on age because the age "in completed years" changes every year and people may not be keeping up the counting of age correctly, and they may be able to recall the "year of birth" because it is one point of time reference. In case of Pakistan, it has been observed, that the phrasing of question on age does not make much difference. However, the question on age should be very simple so that a layman is able to understand and respond without getting confused. Any carelessness on behalf of the respondents and enumerators would produce errors in data.

The long run solution to the problems of age mis-reporting are expected alongwith the socio-economic development and improvements in the literacy level.[14] There should be compulsory registration system with emphasis on numerical age. In a developing country like Pakistan, these changes are not likely to come in the near future. In the mean time, there must be continuous effort from the government to improve the quality of data to encourage the concerned agencies, to keep the birth records through political representatives, religious leaders, hospitals and health clinics, police stations and post offices etc.

References

1. Afzal, Mohammad, 1973, 1972 Census, population expected and actual, *The Pakistan Development Review*, **12**(2), 123-134.
2. Bachi, Roberto, 1954, Measurement of the tendency to round off age returns, *Bulletin of the International Statistical Institute* (Proceedings of the 28th Session, Rome, September 6-16, 1953), **34**(3), 128-138.
3. Bean, Lee L., 1974, The population of Pakistan –An evaluation of recent statistical data, *The Middle East Journal*.
4. Burki, Shahid Javed, 1973, Pakistan's population; ready or not, here they come, *Pakistan Economist*, October 30.
5. Hashmi, Sultan S, and Iqbal, Alam, 1969, *The Problem of Obtaining AgeD ata in Pakistan*, Karachi, The Pakistan Institute of Development Economics (Research Report No. 83).

14. An analysis of age reporting by literacy at the district level has shown a positive correlation between the two variables.

6. Karim, Mehtab S. and Iqbal, Alam, 1975, Age reporting in Pakistan and its implicati on for fertility analysis, *The Pakistan Development Review*, **14** (1), 100-119.
7. Krotki, Karol J., 1963, Population size growth and age distribution : Fourth release from the 1961 Census of Pakistan, *The Pakistan Development Review*, 3(2), 279-305.
8. Krotki, Karol J. and Khalida, Parveen, 1976, Population size and growth in Pakistan based on early reports of 1972 Census, *The pakistan Development Review*, 15(3), 290-318. Reprinted as Population Reprint Series No. 27 (1977)by the Population Research Laboratory. University of Alberta.
9. Ogum, Godwin E. O. 1975, Comparative study of age reporting in selected censuses and surveys in Tropical Africa, *Sc.D. Thesis*, The Johns Hopkins University, U. S. A., Maryland.
10. Ramachandran, K. V. 1967, An index to measure Digit preference error in age data, *Summary in United Nations, World Population Conference*, 1965 (Belgrade). Volume III, pp. 202-203.
11. Retherford, Robert D. and Mujtaba Mirza G. 1981, *Evidence of Age Exaggeration in Demographic Estimates for Pakistan*, Working Paper No. 10, East-West Population Institute, East-West Center, Hawaii, Honolulu.
12. Shah, M. Rafique, 1975, Analysis and composition of PGE and PGS Data. In : Krotki and Sultan Hashmi, (ed) *Issues in Demographic Data Collection in Pakistan*, Pakistan Census Organizatian.
13. Shryock, Henry S. and Jacob S. Seigel, 1975, *The Methods and Materials of Demograyhy*, Volumes I and II, Washington, D. C. : U. S. Bureau of the Census.
14. Yusuf, Farhat, 1967, On the extent of digital preference in reporting of ages in Pakistan, *The Pakistan Development Review*, 7 (4), 519-526.
15. Zaki, Khalida Parveen and M. Javed A. Zaki, 1981, *A Comparative Study of Age Reporting in Selected Censuses and Surveys in Pakistan*, The Pakistan Institute of Development Economics, Islamabad, Islamabad (Research Report No. 130). (First Draft).

Appendix Table-1

AGE DISTRIBUTION OF PARITY ONE WOMEN AGES 15-35 IN HED (1973) AND PFS (1975)

Age	*HED 1973*	*PFS 1975*
Below 15	0	0.4
15-19	15.5	37.9
20-24	38.4	37.1
25-29	28.7	16.1
30-34	11.7	6.2
35+	5.8	2.5

SOURCE : [As given in Table 1].

Appendix Table-2

SEX RATIOS BY AGE IN SEVERAL PAKISTANI CENSUSES AND SURVEYS : 1951-1981

Age groups	*Censuses*				*Surveys*									
	1951	*1961*	*1972*	*1981*	*PGE 1962*	*PGE 1963*	*PGE 1964*	*PGE 1965*	*PGS 1968*	*PGS 1969*	*PGS 1971*	*PGS 1976*	*HED 1973*	*PES 1975*
0-4	105	105	101	98	105	107	108	109	109	110	110	106	98	108
5-9	112	115	111	108	111	110	105	113	112	114	114	108	112	109
10-14	114	123	127	119	127	127	127	125	120	123	124	118	126	109
15-19	106	118	120	116	123	131	136	135	118	115	121	113	130	97
20-24	117	113	107	109	97	93	105	104	105	108	101	95	116	97
25-29	135	109	112	107	98	103	106	104	110	103	104	102	107	99
30-34	134	110	108	104	105	96	112	106	109	109	106	98	108	103
35-39	135	117	116	106	122	121	118	119	118	111	118	110	105	178
40-44	132	119	116	109	116	107	118	113	109	122	131	109	118	104
45-49	130	125	122	112	137	133	120	118	121	119	141	107	112	101
50-54	131	129	133	124	119	125	137	140	116	133	135	122	139	133
55-59	108	126	118	113	151	131	122	121	122	102	73	103	101	—
60+	117	133	137	133	134	137	130	123	128	133	140	128	150	—
All ages	117	116	115	110	114	114	115	115	114	115	115	109	116	109
Sex Ratios	7.0	5.0	8.6	6.1	13.9	14.3	10.2	11.3	6.6	9.1	13.2	8.6	13.0	9.4

SOURCES : [See Source given in Table-1].

1. For age 50+

Appendix Table-3

AGE RATIOS FOR MALES IN SEVERAL PAKISTANI CENSUSES AND SURVEYS, 1951-81

Age groups	*Censuses*				*Surveys*									
	1951	*1961*	*1972*	*1981*	*PGE 1962*	*PGE 1963*	*PGE 1964*	*PGE 1965*	*PGS 1968*	*PGS 1969*	*PGS 1971*	*PGS 1976*	*HED 1973*	*PFS 1975*
0-4	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5-9	88.0	128.1	116.4	112.5	109.0	110.1	114.6	117.7	117.6	117.4	117.5	110.7	119.3	107.4
10-14	132.6	78.1	106.8	104.4	95.0	94.0	95.5	96.0	96.3	99.7	102.9	109.9	108.6	108.6
15-19	94.9	102.6	86.4	91.6	96.2	99.3	95.6	91.5	83.1	81.0	81.9	89.6	91.4	90.8
20-24	87.7	93.3	87.9	94.4	86.1	83.9	83.9	83.6	100.5	96.3	90.5	87.2	87.7	88.3
25-29	98.3	107.0	111.3	101.3	113.6	116.0	121.0	120.1	107.7	105.3	108.6	102.3	104.8	101.4
30-34	102.1	98.2	96.5	94.7	97.4	92.7	92.7	88.4	97.3	103.1	97.3	98.2	96.1	102.0
35-39	90.0	94.9	97.3	98.4	95.2	103.6	100.3	111.7	97.2	93.4	100.2	101.4	97.1	99.6
40-44	111.8	105.8	106.4	102.9	107.1	94.7	100.0	87.1	103.8	111.9	107.6	99.6	107.8	103.2
45-49	85.9	90.1	87.1	92.2	87.0	95.4	90.8	101.3	85.6	79.6	86.7	94.1	87.1	—
50-54	131.7	128.3	136.2	90.9	130.8	132.1	131.1	124.0	136.3	146.0	116.0	122.9	140.3	—
55-59	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60+	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Age Ratio Scores	13.3	11.7	12.2	5.9	10.4	10.2	10.8	12.5	10.7	13.4	9.7	7.9	12.1	5.5

SOURCE : [See Sour8e given in Table-1].

Appendix Table-4

AGE RATIOS FOR FEMALES IN SEVERAL PAKISTANI CENSUSES AND SURVEYS 1951-1981

Age Groups	*Censuses*				*PGE Survey*				*PGS Survey*				*HED*	*PFS*
	1951	*1961*	*1972*	*1981*	*1962*	*1963*	*1964*	*1965*	*1968*	*1969*	*1971*	*1976*	*1973*	*1975*
0-4	—	—	—	—	—	—	—	—	—	—	—	—	—	—
5-9	86.4	124.0	118.0	110.8	110.8	114.2	125.7	120.2	119.5	118.8	119.9	114.1	118.9	106.7
10-14	128.7	73.8	95.7	97.2	86.4	86.8	86.4	92.1	91.2	92.4	96.4	102.7	101.4	104.5
15-19	102.9	102.3	85.6	90.9	88.7	84.5	82.7	78.7	80.2	81.6	76.4	86.4	85.9	97.3
20-24	86.9	94.2	95.9	97.3	97.9	105.2	95.3	98.7	108.9	100.0	101.1	99.6	90.7	89.1
25-29	89.6	109.9	106.7	101.5	116.6	106.5	124.3	120.7	104.0	108.9	107.2	96.3	109.4	102.1
30-34	103.0	99.6	101.5	96.3	98.9	105.7	90.8	91.7	100.7	100.9	101.7	105.8	95.0	108.8
35-39	88.9	92.4	93.3	99.1	85.5	86.3	97.6	101.8	89.7	96.1	98.5	94.6	103.4	80.7
40-44	112.2	106.9	108.9	102.2	117.9	111.4	100.5	91.9	113.4	104.7	103.2	99.4	98.4	114.8
45-49	87.1	88.9	87.4	94.8	74.6	82.3	95.3	106.3	79.3	84.5	81.6	100.3	99.0	—
50-54	121.4	124.3	124.2	117.1	155.1	139.4	115.7	105.4	143.2	123.5	89.2	106.7	109.0	—
55-59	—	—	—	—	—	—	—	—	—	—	—	—	—	—
60+	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Age Ratio Scores	12.73	11.8	10.1	5.6	16.8	14.2	11.8	10.1	15.0	10.2	9.1	5.9	7.3	8.7

SOURCES : [See source given in Table-1].

Appendix Table-5

RE PORTED AND ADJUSTED MALE POPULATION AND MORTALITY RATES BY AGE IN PGS 1976

Age groups	*Reported males*	*Adjusted*[1] *males*	*Adjusted*[2] *males*	*Reported ASDR×1000*	*Adjusted*[1] *ASDR×1000*	*Adjusted*[2] *ASDR×1000*
15-19	2868799	2771710	2834879	2.51	2.59	2.54
20-24	2108890	2205979	2142792	2.39	2.28	2.35
25-29	1969407	1976553	1982099	1.70	1.69	1.68
30-34	1741069	1733923	1728377	3.84	3.85	3.86
35-39	1577957	1562451	1562489	1.02	1.03	1.02
40-44	1364943	1385449	1385411	8.27	8.18	8.18
45-49	1172169	1241667	1236166	5.39	5.09	5.12
50-54	1121681	1052183	1057684	6.56	6.99	6.96
55-59	653089	848903	847377	12.33	9.48	9.42

SOURCES : [PGS, 1976].

1. Adjusted by Carrier Farrage Ratio method.
2. Adjusted by Quadratic Interpolation method.

Appendix Table-6

REPORTED AND ADJUSTED FEMALE POPULATION, FERTILITY AND MORTALITY RATES BY AGE IN PGS, 1976

Age Groups	*Reported Females*	*Adjusted[1] females*	*Adjusted[2] females*	*Reported ASFR× 1000*	*Adjusted[1] ASFR× 1000*	*Adjusted[2] ASFR× 1000*	*Reported ASDR× 1000*	*Adjusted[1] ASDR× 1000*	*Adjusted[2] ASDR× 1000*
15-19	2538650	2617862	2663730	56.3	54.6	53.7	3.18	3.08	3.03
20-24	2224853	2145642	2099773	271.2	281.2	287.3	3.73	3.87	3.95
25-29	1929517	1987186	1984573	348.2	338.1	338.4	4.60	4.46	4.47
30-34	1780903	1723234	1725847	305.4	315.6	315.1	3.23	3.34	3.33
35-39	1436795	1449708	1453019	225.7	223.7	223.1	4.19	4.16	4.15
40-44	1256904	1243991	1240680	127.6	128.9	129.3	4.85	4.91	4.92
45-49	1091744	1097458	1093125	72.5	72.1	72.4	4.69	4.66	4.68
50-54	919959	914245	918578	—	—	—	14.58	14.67	14.61
55-59	631497	727265	726930	—	—	—	5.82	5.05	5.05
Total	2259366	12265081	12260747	—	—	—	—	—	—
TFR	—	—	—	6.53	7.07	7.09	—	—	—

SOURCES : [PGS, 1976].

1. Adjusted by Carrier-Farrage Ratio method.
2. Adjusted by Quadratic Interpolation method.