No.83
THE PROBLEM OF OBTAINING AGE DATA IN PAKISTAN:
A Study of Age Reporting of a Panel of Ever
Married Females in Yearly Enumerations
1962-1965

by

Sultan S. Hashmi   and   Iqbal Alam

September 1969

THE PROBLEM OF OBTAINING AGE DATA IN PAKISTAN:

A study of age reporting of a panel of ever married females in yearly enumerations, 1962-1965

by

Sultan S. Hashmi
and

Iqbal Alam*

(Pakistan Institute of Development Economics, Karachi
Pakistan)

Introduction

In Pakistan, like many other developing countries obtaining the age data of acceptable quality through enumerations is a serious problem. In a society where according to the 1961 census the level of literacy was as low as 16 percent $\sqrt{5}$,p.133$\sqrt{}$ and not much importance is attached to age, the problem is further aggravated. Age reporting in the 1951 Population Census of Pakistan was highly questionable. Due to unmanageable discrepencies the data were generally published in broad age groups, i.e. 0-9, 10-39, 40-59 and 60 and over $\sqrt{3}$, p.48$\sqrt{}$. In the 1961 Census the ages were reported in single year upto age 9, in conventional five years age groups from 10-59 years and in ten years age groups from 60 to 99 years and in

---

100 years and above /4, p.1-15/. Though the 1961 Census is considered to have produced better data than the 1951 Census, the quality of the age data does not appear to have improved very much. The data obtained through surveys also suffer from somewhat similar irregularities as the age data obtained in the censuses. The age data obtained in the Population Growth Estimation (PGE) experiment which provide the basis for the present analysis were also not free from irregularities. However, a major difference between the Census and the PGE data was that the former were collected by unpaid honorary enumerators while the latter were obtained by paid enumerators who were also better trained than the census enumerators. In addition, field supervision in PGE was somewhat more strict. Thus the quality of age data in the PGE may be somewhat better.

In most of the cases the enumerator plays an important role in estimating the ages as the people themselves do not have any idea about their ages /4, p. III/. Besides, in an enumeration ages for all the persons of a household are reported mostly by a male member of the household. In some studies the problem of irregularities in age reporting in Pakistani censuses and surveys has been dealt with /6, p.135, 2,8,9,10/. Indices computed under the Myer's method for female age reporting range between 65.0 and 91.2 /1,10,12/. The errors in the age data appear to be too large and non-random to yield to attempts to smooth the age data by using some of the conventional demographic techniques /7/.

In this paper an attempt has been made to investigate variation in the age reporting of a panel of ever married females (who were reported in ages 15-49 in the first PGE enumeration in January, 1962) by collating their ages reported in the 4 yearly consecutive enumerations (1962-1965). The main object of the present study is to find out how the reported

ages of females vary from one enumeration to another? What
is the direction of such variation? How are ages of women,
classified into young (15-29 years), middle aged (30-39 years)
and older (40 years over), reported from one enumeration to
another? Are ages of women of high parity exaggerated? Is
variation in age reporting of females associated with age
of eldest child? Are variations different at preferred digits
of age reporting (zero. five, even and odd) ? In the rest of
this paper attempt is made to answer these questions.

Data

In the PGE experiment, there were two systems of data
collection; (i) Longitudnal Registration under which births
and deaths occuring inside each sample area were registered
by a registrar living in the area and (ii) the Cross-sectio-
nal Survey under which population of households in sample
area was enumerated and data on age composition as well as
on various other characteristics of the population were
obtained through quarterly enumerations from 1962 to 1966.
The first enumeration in each year was a full enumeration
done independently of the previous enumerations and the
subsequent three enumerations were meant to record changes
since the first enumeration of the year. [1]/The present
study is based on a sub-sample of 25 percent of women who
were reported married and were in the reproductive ages
(15-49 years) in the first full enumeration in 1962 [_13_]·
Thus· a panel of 4117 women was obtained by using the systa-
metic sampling technique. The primary object of this sub-
sample was to undertake a study of the reproductive perfor-
mance of the selected women. For the present study ages of
these women were traced through the records of the successive

---

1/ Details of the sample and study designs are given
in [_· 12_]·

enumerations until the full enumeration in 1965. From the
present study ages of these women were traced through the
records of the successive enumerations until the full
enumeration in 1965. From the information coded the following
variables were selected for the present study:

(i)   ages of each woman as reported in the full
       enumerations, 1962-1965.

(ii)   parity reported at the time of last live/still
       birth during the period 1962-1965, and

(iii) age of eldest child reported in the first
       enumeration in 1962.

Variations in Age Reporting

Mean, standard deviation and coefficient of variation
of each woman who was enumerated at least in two yearly
enumeration visits were computed. It was noted that the
standard deviations and coefficients of variation ranged
between 0 to 23 years and 0 to 67 percent respectively
indicating that there was a great deal of variation in the
ages of females from one enumeration to the other. Zero
standard deviation was obtained for females who reported the
same age in all the visits in which they were enumerated. The
expected range of standard deviations due to aging between
1962 and 1965 for various combinations was 0.7 to 2.1. At
least all values of standard deviation below. .7 and above
2.1 signify errors in age reporting. In other words zero,
which indicates no difference between ages reported in various
enumerations, also signifies erroneous reporting. The difference
between the two reported ages of a woman was also quite wide.
For example, between 1962 and 1965 enumerations, the differences
in the two reported ages varied between -47 to +45.

TABLE I

Population of Females 15-49 in Conventional Five Year Age Groups in the 1962 and 1965 Enumerations

Age Groups in 1962	Age Groups in 1965										
	Total	Under 15	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55 and over
Total	3270	2	131	479	673	634	452	376	261	160	102
Under 15	394	1	101	197	69	16	3	3	1	—	3
15-19	616	—	24	185	260	92	24	12	8	5	6
20-24	729	—	3	68	237	259	98	43	14	5	2
25-29	537	—	—	18	72	166	147	90	28	12	4
30-34	415	—	—	8	23	67	113	108	59	24	13
35-39	365	—	—	—	10	27	58	89	89	56	34
40-44	207	—	—	1	2	6	8	30	62	56	38
45-49	7	—	—	2	—	1	1	1	—	2	2
50 and over	—	—	—	—	—	—	—	—	—	2	2

Table I shows ages as reported in 1962 and 1965 in the conventional five year age groups. Out of 4117 women selected from the 1962 enumeration there were 3270 women for whom ages were also reported in the 1965 enumeration. It is noted that after a period of three years some women have become younger and others older. A substantial proportion (20.7 percent) of women who were over 40 years in 1962 were reported under thirty five years in 1965. On the other hand (5.9 percent) of women who were under 30 years in 1962 were reported as 40 years and over after a period of three years. Comparison between other enumerations have not been shown due to limitation of space.

Inter-enumeration Visit Difference in Age Reporting

Women on the basis of their average age reported on all the visits were classified into three categories: (i) women aged 15-29 years (young), (ii) women aged 30-39 years (middle aged) and (iii) women aged 40 years and over (old). We do realise that women in age group 30-39 are not really middle aged and most of the women 40 years and over are not older and we apologise for making such a classification which has been done arbitrarily for convenience of presentation. Under this system 46.7 percent of the total 3,270 women who were enumerated in 1962 and whom we were able to trace in 1965 were classified as young, 35.0 percent as middle aged and 18.3 percent as relatively older.

Table II presents the differences in completed years between ages reported in 1962 and 1965. It shows that ages of 22.5 percent of the total women were reported higher by 1-3 years in 1965 than their ages in 1962. The expected difference between the 2 visits was 3 years, as such ages of a small fraction (269 out of 3270 or 8.2 percent) of the total women were reported accurately. Giving a margin of error of 2 years on both sides we can assert that ages of women which were reported by 1-3 years and 4-5 years higher in 1965 than the

TABLE II

Percent Distribution of Women by Age Status and Difference in Age Reporting (in completed years) between the 1962 and 1965 Enumerations

Age status 1/	Total	Percent distribution of women by difference in age reporting										
		age in 1965 more than 1962					no difference	age in 1965 less than 1962				
		1-3	4-5	6-9	10-14	15+		1-3	4-5	6-9	10-14	15+
All Women	100.0	22.5	17.2	15.2	11.1	5.8	9.8	9.1	4.6	2.7	1.5	0.6
Young	100.0	30.5	19.3	16.3	6.5	1.2	9.0	10.3	3.5	2.4	0.7	0.5
Middle aged	100.0	16.1	15.1	14.4	14.6	7.2	11.0	8.0	6.2	3.8	2.6	1.1
Older	100.0	14.2	15.6	13.9	16.1	14.7	8.7	4.0	1.7	1.7	1.3	0.2

1/   Average age: 15-29 Young
  30-39 Middle aged
  40 and over Older

TABLE III

Coefficient of Variation of Ages of Females Reported in
At least Two Enumerations by Cumulative PercentDistribution
of Ages Reported at Preferred Digit, 1962-65.

Coefficient of variation	Cumulative percent distribution at digit of age reporting ending in			
	zero	five	even	odd
Total	100.0	100.0	100.0	100.0
0	3.7	2.0	1.0	0.4
1	5.2	3.1	1.6	1.9
2	7.6	5.8	4.6	6.9
3	11.8	9.2	11.0	13.8
4	16.2	13.0	16.8	22.1
5	22.3	19.4	24.6	29.0
6	28.3	27.8	32.5	36.7
7	34.9	34.0	38.9	43.6
8	42.1	41.8	47.1	50.0
9	48.2	50.3	54.0	56.7
10	55.3	57.7	60.8	63.0
11	62.3	63.7	66.9	69.0
12	69.1	70.1	73.0	75.4
13	74.0	74.3	78.5	79.9
14	78.1	78.2	82.1	83.5
15	81.3	81.4	84.7	87.1
16	84.9	85.2	88.1	89.5
17	87.3	88.9	90.6	91.3
18	89.9	90.5	92.4	93.3
19	91.9	92.4	93.4	94.1
20	93.7	94.3	94.9	95.5
21	94.6	95.5	95.8	96.8
22	95.5	96.5	96.8	97.1
23	96.3	97.2	97.1	97.4
24	96.9	97.4	97.3	97.7
25	97.7	98.0	97.8	98.1
26	97.8	98.1	97.8	98.2
27	98.0	98.3	98.0	98.2
28	98.2	98.3	98.1	98.5
29	98.3	98.4	98.3	98.6
30	98.6	98.6	98.5	98.9
31+	100.0	100.0	100.0	100.0

ages in 1962, were more or less reported consistently. Under
this criterion, ages of approximately 40 percent of women can
be considered as reported consistently, ages of 28 percent
were under-stated and ages of 32 percent were over-stated in
1965 relative to 1962. Of course, it must be realized that
ages in 1962 were also mis-reported and the differences shown
in table II are relative differences and not the true differences.
It is noted that the percentage of women whose ages were
reported consistently decreases with age. Percentage of women
whose ages were over-stated by 6-9 years in 1965 also decreases
with age while percentage of women whose ages were over-stated
by 10 and more years increases with age. Percentage of women
whose ages were under-stated does not conform to a pattern.
However, it is noted that in general middle aged woman tend to
understate their ages more than the young and older women.

Variation in Age Reporting by Digital Preference

In Table III, cumulative percentages of ages which
were reported at digits ending in zero, five, even and odd,
are shown by the magnitude of coefficient of variation.
Percentages of ages against the values of c.v. are presented
in Graph I. It was noted that preference for reporting of ages
of females was in the expected order of digits i.e. zero,
five, even and odd. Coefficient of variation of zero percent
was obtained for those in whose ages no changes were reported
from enumeration to enumeration. Of course, if all ages were
correctly reported, true values of the c.v. would also have
varied between 1-13 percent but the expected distribution of
ages against the c.v. values would have been much different
from the observed values. It is noted that at 10 percent
level of c.v. the proportion is somewhat higher for ages
reported at digits ending in odd numbers than at other digits
and though the differences are not striking but the data
suggest that relative variation in age reporting of females
whose ages were reported at digits ending in odd numbers was

TABLE IV

Percent Distribution of Women of Each Parity by Difference in Age Reporting (incompleted years) between the 1962 and 1965 Enumerations

Parity	Total	Percent Distribution of women by differences in age reporting										
		age in 1965 more than 1962					no difference	age in 1965 less than 1962				
		1-3	4-5	6-9	10-14	15+		1-3	4-5	6-9	10-14	15+
All parities	100.0	22.5	17.2	15.2	11.1	5.8	9.8	9.1	4.6	2.7	1.5	0.6
01	100.0	30.2	18.8	12.8	8.0	3.4	12.1	10.7	2.0	0.7	0.7	0.7
02	100.0	33.2	13.7	18.0	6.7	4.3	11.0	7.3	2.4	1.3	0.6	0.9
03	100.0	27.9	18.9	15.8	8.7	3.0	9.0	9.0	3.3	1.6	0.6	0.9
04	100.0	19.9	20.5	15.7	7.3	3.7	11.0	9.2	3.3	1.6	1.6	1.1
05	100.0	20.6	20.9	16.3	12.0	3.4	8.3	9.2	5.5	3.4	3.4	0.3
06	100.0	24.1	13.8	13.8	9.6	5.0	10.0	10.3	2.8	0.9	0.9	0.4
07	100.0	21.6	15.2	15.8	18.7	2.9	6.4	9.9	7.3	4.6	1.1	0.3
08	100.0	17.4	11.9	18.3	12.8	8.3	11.9	6.4	2.9	4.1	0.6	1.8
09	100.0	21.2	7.6	22.7	16.7	7.6	9.1	10.6	8.3	2.8	1.1	0.0
10 and over	100.0	14.5	23.2	21.7	13.01	7.2	8.7	4.3	3.0	0.0	1.5	0.0
Unknown	100.0	18.5	17.0	12.9	13.2	9.3	10.0	9.4	4.7	2.9	1.5	0.7

somewhat less in comparison to ages ending in even, five and
zero respectively. Similarly at 15 percent level of c.v. the
proportion of ages reported at digits ending in odd numbers
is still higher than other digits, again indicating that
variation in age reporting of females whose ages were reported
in odd numbers was less in comparison to ages ending in even,
five and zero respectively. In general, it is noted that
variation is highest in ages reported at digits ending in zero,
next highest in ages reported at even, and least in ages
reported at odd digits. On the basis of this meagre evidence
a hypotheses can be advanced that ages of persons reported at
odd digits are relatively less inaccurate and ages reported
at zero are more inaccurate than other digits.

Differences in Age Reporting and Parity

We thought ages of women of high parity might be over-
stated and of low parity under stated. In table IV we have
presented the differences in age reporting between 1962 and
1965. It is noted that with some fluctuations the ages of
women of low parity were reported more consistently than ages
of women of high parity. A substantial proportion (32 percent)
of women are in the unknown parity group. This is due to the
reason that we were able to obtain parity only of those women
who had atleast one live or still birth during the period
1962-65. It is noted that the percentage of women whose ages
were over-stated is lowest for women of parity one and highest
for women of parity nine. The next highest difference of this
nature is noted for women of parity ten and over. Again with
some fluctuations it is evident that ages of women of high
parity were more over-stated than women of lower parity. Lowest
proportion of women whose ages were under-stated is found for
women of parity ten and over, while the highest proportion is
noted for women of parity six. Although fluctuations in the
proportion of women whose ages were under-stated is somewhat
greater, there seems to be a tendency that ages of women of
high parity are less under stated.

TABLE V

Percent Distribution of Women by Age of the Eldest Child and Difference in Age Reporting (in completed years) between the 1962 and 1965 Enumerations

Age of the eldest child (in completed years)	Total	Percent distribution of women by difference in age reporting										
		age in 1965 more than 1962					no difference	age in 1965 less than 1962				
		1-3	4-5	6-9	10-14	15+		1-3	4-5	6-9	10-14	15+
All ages	100.0	22.5	17.2	15.2	11.1	5.8	9.8	9.1	4.6	2.7	1.5	0.6
Upto 1	100.0	27.4	15.8	21.0	6.3	4.2	8.4	11.6	2.1	3.2	0.0	0.0
1	100.0	28.1	14.6	25.0	8.3	1.0	12.5	7.3	1.0	1.0	1.0	0.0
2-3	100.0	26.6	19.8	15.1	9.7	2.2	10.1	9.4	2.5	2.2	1.4	1.1
4-6	100.0	26.4	22.1	12.8	8.0	3.0	10.6	7.5	4.3	2.8	1.3	1.3
7-9	100.0	20.5	18.4	15.5	10.7	4.4	7.8	10.3	6.1	3.8	2.3	0.2
10-12	100.0	20.5	15.9	14.1	11.0	6.8	11.8	9.7	4.8	4.1	1.2	0.0
13-15	100.0	15.9	16.0	15.6	17.8	4.7	10.2	8.0	6.5	2.9	1.8	1.1
16-18	100.0	15.3	16.4	13.9	13.4	9.1	6.9	9.5	6.1	2.6	1.2	2.2
19-21	100.0	17.7	15.6	13.3	15.6	9.4	11.7	9.4	3.1	1.6	2.2	1.1
22-24	100.0	17.2	15.6	13.3	15.6	12.4	5.6	10.1	3.4	2.2	3.1	0.0
25+	100.0	18.0	16.9	19.1	10.1	15.3	8.5	6.8	6.2	1.7	1.1	1.1
Others	100.0	28.7	15.1	15.4	9.4	5.1	10.8	9.2	3.7	1.7	0.7	0.2

Difference in Reporting Ages of Mothers in Relation to Age of the Eldest Child

The ages of the children were obtained from the first full enumeration in 1962. For about 17 percent females no ages of eldest children were reported. Table V shows no regular pattern in reporting the ages of women by the ages of their eldest child but still one can draw a crude inference that ages of higher proportion of women who had younger (under 10 years) eldest children were reported more consistently than the ages of women whose eldest children were older. Over-stating of ages appears to be relatively greater for the women whose eldest children were older (10 years and over). The highest proportion is observed for a woman whose eldest child was 25 and over and the lowest is observed for a woman whose eldest child was 4-6 years of age. It is also noted that lowest proportion of women whose ages were under-stated is observed for women whose eldest son was one year old and the highest proportion of such women is found for women whose eldest child was 10-12 years old. For those whose ages were under-stated the proportion increases from women whose eldest child was one year old upto those whose eldest child was 10-12 years old and then it declines.

Net Effect of Mis-statement of Ages

A question arises that in view of the high magnitude of errors in age reporting what is the net effect of over and understatement of ages on the age distribution. Table VI shows ages as reported in the 1962 enumeration and the ages reported in the expected five years age groups in 1965. Since analysis is confined to those women who were enumerated in both the years, 1962 and 1965, the expected distribution in columns 2 and 3 should be the same as in columns 5 and 6 of this table. It is noticed that there was no woman under 15 years in 1962 but after 3 years in 1965, 1.3 percent of the women have been reported under 18 years. The percentage of women in five year

TABLE VI

Population and Percent Distribution of Females 15-49 in Conventional
Five Year Age Groups in 1962 and Expected Five Year Age Groups in 1965

	1962			1965		1965 - 1962	
Age Group	Population	Percent	Age Group	Population	Percent	Population	Percent
All ages	3270	100.0		3270	100.0		
Under 15	-	-	Under 18	44	1.3	44+	1.3+
15-19	394	12.0	18-22	406	12.4	12+	0.4+
20-24	616	18.8	23-27	628	19.2	12+	0.4+
25-29	729	22.3	28-32	750	22.9	21+	0.6+
30-34	537	16.4	33-37	432	13.2	105-	3.2-
35-39	415	12.7	38-42	461	14.1	46+	1.4+
40-44	365	11.2	43-47	233	7.1	132-	4.1-
45-49	207	6.3	48-52	195	6.0	12-	0.3-
50+	7	0.2	53+	121	3.7	114+	3.5+

age groups in the age span 15-29 in 1962 has somewhat
increased in the expected age groups in 1965, indicating a
net understatement of ages. In age group 30-34, percentage
has considerably declined in the expected age group and in
age group 35-39 the percentage has again increased. There is ɩ
a decline in the percentage of women 40 years and over from
17.7 percent in 1962 to 16.8 percent in the expected ages
in 1965.

The deviation from the expected pattern is simply due
to mis-allocation of women. It is again noted that there is
a tendency of understating the ages of younger women and over-
stating the ages of older women and the female population in
1965 looks younger than it was in 1962. In general, it is
noted that inspite of cancellation of errors due to over and
understatement of ages, the difference between the expected
and observed age structure in 1965 is statistically
significant.

Summary

The analysis shows that there is a great deal of
variation in the age reporting of a panel of females from
oneyearly enumeration to another. Although errors are not
very systematic some results of this study do conform to
an expected pattern. It is noted that ages were under-stated
as well as over-stated and the extent of both under-statement
and over-statement is of a high degree of magnitude. Ages of
younger women were reported somewhat more accurately than the
ages of middle aged and older women. Ages of older women were
more overstated than the ages of middle aged and younger
women. The margin of error is greater for older woman. There
is some evidence that ages reported at odd digits are less
inaccurate than ages reported at zero, even and five. Ages
of women of low parity were reported less inaccurately, and
were less overstated than women of high parity. Ages of
women who had younger children were less inaccurate, less

overstated and less understated than ages of women who had
older children. The net effect of over and under statement
of ages on the age structure of females inspite of cancellation
of errors is not negligible.

It appears that if the errors in the age reporting in
the censuses and other surveys in Pakistan are at least of
the same magnitude as presented in this paper, it poses a
serious problem for the analysts of demographic data. The errors
seem to be so wide spread that some of the smoothing techniques
when applied do not improve the age data. Similarly demographic
parameters inferred from the reported age data are also likely
to be affected by errors.

REFERENCES

1. Aslam, M.,
    Sultan S. Hashmi
    and
    William Seltzer.

    "Abridged Life Tables of Pakistan and Provinces by Sex, 1962", Pakistan Development Review, Vol. VII, No.1, Spring 1967.

2. Bose, S.R.

    "Labour Force and Employment in Pakistan 1961-86 - A Preliminary Analysis", Pakistan Development Review, Vol. III, No.3, Autumn 1963.

3. Government of Pakistan.

    Census of Pakistan, 1951, Vol. 1, Pakistan Report and Tables, Office of the Census Commissioner, Government of Pakistan, Karachi, no date.

4. Government of Pakistan.

    Census of Pakistan, Vol. 1, Pakistan, Population, 1961, Office of the Census Commissioner, Ministry of Home and Kashmir Affairs, Karachi, no date.

5. Hashmi, Sultan S.

    Main Features of Demographic Conditions in Pakistan, a country background paper. Central Statistical Office, Government of Pakistan, Karachi, 1963.

6. Hashmi, Sultan S.

    The People of Karachi: Demographic Characteristics, Pakistan Institute of Development Economics, Karachi, 1964.

7. Khan, M.K.H.

    "Population Projection of Pakistan 1960-85", (Mimeographed), Social Sciences Research Centre, University of Panjab, Lahore 1965.

8. Krotki, Karol J.

    "Population Size, Growth and Age Distribution Fourth Release from the 1961 Census of Pakistan", Pakistan Development Review, Vol. III, No. 2, Summer 1963.

9. Mauldin, W.P.
    and
    Sultan S. Hashmi

    "Illustrative Estimates and Projections of the Population of Pakistan, 1951 to 1961", in M.L. Qureshi (ed.), <u>Population Growth and Economic Development with special reference to Pakistan</u>, Pakistan Institute of Development Economics, Karachi, 1960.

10. Robinson, W.C.,
    William Seltzer
    and
    Sultan S. Hashmi.

    "Quasi Stable Estimates of the Vital Rates of Pakistan", <u>Pakistan Development Review</u>, Vol. V, No. 4, Winter 1965.

11. Population Growth
    Estimation.

    <u>Report of the Population Growth Estimation Experiment Description and Some Results for 1962 and 1963</u>, Pakistan Institute of Development Economics, Karachi, 1968.

12. Yusuf, Farhat.

    "On the Extent of Digital Preference in Reporting of Ages in Pakistan", <u>Pakistan Development Review</u>, Vol. VII, No. 4, Winter 1967.

13. Yusuf, Farhat
    and
    Anna Turowicz.

    "Proposal Regarding Some Analysis of the Reproductive Performance of Women in the PGE Sample", (Mimeographed), Pakistan Institute of Development Economics, Karachi, no date.