Canadian Studies in Population, Vol. 22(2), 1995, pp. 107-119.

Method for Estimating the Under-Reporting of Migrants in Census Data: Illustrated with Indian Data

M. Sivamurthy*, A.S. Kadi, and S.B. Madagi
Karnatak University
Dharwad, Karnataka, India.

*Formerly Professor of Statistics, Bangalore University, Bangalore and U.N. Expert, Demographic Centre, Cairo.

Abstract

In census enumeration's migrants who will have moved back to their place of birth are often not reported as migrants. This phenomenon is noticed in the census enumeration's of developed as well as of developing countries. The present paper suggests that an examination of the sex ratios of migrants by duration of stay at the place of enumeration, could reveal the occurrence of such under-reporting of migrants of one sex in comparison with the reporting of the migrants of the other sex.

A method is presented for estimating the extent of the under-reporting of migrants of the concerned sex, when it is possible to assume that the migrants of the other sex have been better reported.

Migration data from the Indian censuses are analyzed and it is shown that male migrants have been under-reported in relation to the reporting of female migrants, and the under-reporting increases with increase in the duration of stay of migrants. The extent of the under-reporting of male migrants is estimated assuming that the reporting of the female migrants is complete. However, it must be noted that female migrants themselves might have been under-reported to some extent and hence the estimated under-reporting should be taken as an estimate of relative under-reporting of male migrants.

Résumé

Pour le recensement, les personnes qui retournent à leur lieu de naissance ne sont souvent pas rapportées comme migrants. Ce phénomène est noté dans les recensements effectués par les pays développés et en voie de développement. Le présent article suggère qu'un examen de la répartition par sexe des migrants par durée de séjour au lieu de recensement pourrait révéler la présence d'une telle omission d'enregistrement des migrants d'un des sexes par rapport aux migrants de l'autre sexe.

Les auteurs présentent une méthode visant à estimer l'étendue des omissions concernant les migrants du sexe concerné, quand il est possible de supposer que les migrants de l'autre sexe sont mieux représentés.

Une analyse des données issues de recensements indiens démontre que les hommes sont sous-déclarés par rapport aux femmes, et que cette situation s'accentue à mesure que le séjour des migrants se prolonge. Pour estimer l'ampleur du phénomène, les auteurs ont supposé que les déclarations des femmes étaient complètes. Cependant, si les femmes font elles-mêmes l'objet d'omissions, la sous-déclaration estimée doit être vue comme une estimation de la sous-déclaration relative des migrants masculins.

Key Words: migration, under-reporting, census data, method of estimation.

Introduction

It is by now well-known that a number of migrants are left out in census enumeration's due to several reasons [Goldstein,1964,1972; Hugo,1989; Nicholson, 1990; Skeledon, 1987; Sivamurthy and Kadi, 1983; United Nations, 1970; Willekens and Nair, 1982; Zachariah, 1977]. The main reasons for this phenomenon are that, apart from errors of willful mis-reporting, only migrants surviving at the time of census who are residing at the place of destination which is different from the place of birth, are commonly reported in a census as migrant. Thus, those migrants who returned to their place of birth during the inter-censal period, are generally missed out in the subsequent census enumeration, although the instructions to the enumerators specify that those whose previous place of residence was different from the birth place should also be enumerated as migrants. Moreover, as the duration of stay in the destination increases, migrants tend to report the place of destination itself as their own place. Further, in the census enumeration, reporting is generally done by one person for all the persons in a household and this causes mis-reporting of persons as non-migrants.

This phenomenon is found to exist even in the census enumeration of economically developed countries. For instance, Nicholson (1990) has concluded after analyzing Narvagian census data in comparison with registration data, that the migration data from censuses of developed countries are not free from such deficiencies; and the volume and pattern of population movement represented by the census data may be misleading.

In the case of India also, some studies have shown the existence of considerable extent of return migration which normally results in the not reporting of persons as migrants in the census enumeration's. Zachariah (1967) found the existence of considerable return migration from Greater Bombay, and concluded that without the proper understanding of the return migration, the study of internal migration will be incomplete. Sivamurthy and Kadi (1984) estimated that 2.65 million male migrants (forming 38% of male migrants enumerated in the 1961 census) and 1.71 million female migrants (forming 28% of female migrants enumerated in the 1961 census) had returned from their place of destination during the decade 1961-71, and were not reported as migrants in the 1971 census. In order to understand properly the pattern and impact of population movement in India (or in any other country), it would therefore be necessary to take into account these migrants who will not have been reported as migrants in the census enumeration.

With the recent availability of more detailed data on migrants including the duration of stay at the place of enumeration, there seems to emerge further evidence of this possibility of the existence of return migrants who are not reported in the census enumeration as migrants. The purpose of this paper is to present a method which seems to be useful in certain situations, to estimate the relative extent of this under-reporting of migrants in census enumeration, utilizing the data on duration of stay of migrants at the place of enumeration. Indian census data are used to illustrate the application of the method.

Method for Estimating the Relative Extent of Under-Reporting

The method suggested here, is based on an examination of the trend in the sex ratio among migrants (i.e., number of male migrants per thousand female migrants) by duration of stay, and shall yield the relative extent of under-reporting of migrants of one sex (male or female) in situations where it may be possible to assert that the possibility of under-reporting of migrants of the other sex is smaller. If the migrants of the other sex are completely enumerated, then the estimate of the under-reporting of the migrants of the concerned sex, will obviously be more accurate. Otherwise, the estimate will be an underestimate.

The use of sex ratios for demonstrating and even for effecting corrections for age reporting errors, is well-known in demographic literature. But, the sex ratios among migrants seem to have not been used so far, for demonstrating the relative under-reporting of migrants and for correcting the same. An

attempt is made in this paper to use the sex ratios of migrants by duration of stay, for this purpose. If the sex ratio among migrants with duration of stay of 'd' years is defined as:

$$SR(d)=Nm(d)/Nf(d); for, d=0,1,2,3,...w \quad (1)$$

where Nm(d) and Nf(d) are male and female migrants and w is the recorded maximum number of years of stay; then SR(d) can change as 'd' increases, as a result of differences in survival ratios between males and females, and as a consequence of persons being not reported as migrants due to any reason including that of return migration. Therefore, if the trend in SR(d) as 'd' increases, can not be explained by the relative changes in the survival ratios of males and females in the population, then it will have to be attributed to the not reporting of persons of one sex relative to the other sex, as migrants.

It may be noted that the changes in SR(d) involve only the ratios of over-all survival ratios (i.e., not survival ratios by age groups) at different ages for males to those for females, and the relative improvements in these for males compared to those for females. If proper life tables are available over the past periods of time, it will be possible to adjust the SR(d) for the relative changes in the over-all survival ratios for males and females. But, in practice this adjustment may not be very significant in view of the fact that SR(d) would also be affected by relative duration reporting errors. Hence, it may be assumed for practical purposes that SR(d)=SR(0), where SR(0) is the sex ratio at the time of migration for that migration cohort, which has the duration of stay d (d=0,1,2,...,w). This is the basis of the present method. It must be noted here that the census data will involve many migration cohorts which are enumerated as migrants with different durations of stay. Further, the number of the migrants of the sex, which is assumed to have been enumerated completely, will most likely be under-reported itself and therefore the method represented here will only provide an under-estimate of the number of migrants not reported in the census enumeration.

Now, if it is possible to assert that the reporting of migrants of one sex (say females, in the case of India) is likely to be more complete than the reporting of migrants of the other sex, then the relative number of migrants of the other sex (i.e., males in the case of India) not reported in the census enumeration can be estimated as follows: (a) Male migrants not reported [NRNm(d)], if we assume female migrants are completely reported, is given by :

$$NRNm(d) = \{SR(0).Nf(d)\}-Nm(d), \text{ for duration } d \quad (2)$$

and the total number of male migrants not reported in the census is :

$$NRNm(0+) = \quad NRNm(d) \tag{3}$$

(b) Female migrants not reported [NRNf(d)], if we assume male migrants are completely enumerated, is given by:

$$NRNf(d) = \{Nm(d)/SR(0)\}-Nf(d), \text{ for duration } d, \tag{4}$$

and the total number female migrants not reported in the census is

$$NRNf(0+) = NRNf(d) \tag{5}$$

The question that still remains, is: How to get SR(0) for the different migration cohorts which are enumerated in the census as migrants with different duration of stay of 'd' years?. If there are data from several censuses for the population, then it may be possible to study the trend in SR(0) by computing the sex ratio of migrants with 'zero' duration in the different censuses; and adopt an appropriate value for the SR(0) for a particular migration cohort. Otherwise, the value of SR(0) may have to be fixed on practical considerations, or else the method can not be applied in such a situations.

Trends in SR(d) in the Indian Case

In order to apply the above method to the Indian census data, we shall examine in this section, the trend in SR(d) using the 1961, 1971 and 1981 census data of India. Although India has a long history of census taking, the question on the duration of stay of migrants has been included in census questionnaires only since the 1961 census. Further, in 1961 census, migrants were identified only on the basis of Place of Birth (POB), and it is only since the 1971 census information on POB as well as on the Place of Last Residence (PLR), has been asked with the question on duration of stay of migrants at the place of enumeration.

However, the analysis of the number of migrants using POB and PLR definitions, has shown that in the case of India the two definitions yield more or less the same number of migrants (See Sivamurthy and Kadi, 1984; Kadi and Sivamurthy, 1988; and also the ratio shown in Table 1). Hence the data from the three censuses are used here for comparison. The data from the recent 1991 census are not yet available.

From the value of the sex ratios by duration of stay presented in Table 1, it is evident that the SR(d) decline rather sharply with the increase in 'd' in all the three census data. Also, the sex ratios in all durations of stay have changed only to a small extent from census to census. Although, the errors in reporting of duration of stay will have affected the sex ratios to some extent, the trend is unmistakably clear. This kind of sharp decline in sex ratio can not be explained by the differences in the survival ratios for males and females in India. In fact, in the case of the Indian population, the survival ratios for males have been higher than females, especially during the distant past years [Sivamurthy, 1981]. The sex ratios among migrants should have therefore, increased with the increase in duration of stay 'd', where as the opposite trend is shown in Table 1.

Further, the sex ratio in the (<1) year of duration of stay shows decrease and increase from census to census which may reflect the effect of differential reporting errors in reporting the duration of stay in the three censuses. It seems therefore, that it may not be unrealistic to assume SR(0) as constant in the case of India. Since the sex ratio among migrants in duration of stay (<1) year can be assumed to be the one least affected by the effect of return migration as well as the effect of differentials in male-female mortality, it seems reasonable to use the sex ratio in this duration group as the SR(0) for estimating the number of non-reported migrants.

Also, in the case of India there is a reason to believe that female migrants are more completely reported. The data on reasons for migration collected in the 1981 census, show that as much as 87% of the female migrants had moved because of "marriage" or because "family moved" (Kadi and Sivamurthy, 1988). In the Indian socio-cultural setup, it is common practice for a woman to move to her husband's place as soon as she became a migrant, based on her marital status irrespective of duration of stay. Hence, the chance of a woman being not reported as a migrant is much less than that of a man being not reported as migrant. It may be therefore, assumed that the female migrants will have been more completely enumerated in the Indian censuses. With this assumption, estimates of relative under-reporting of male migrants are obtained for the 1971 and 1981 censuses, as explained in the following section. We have used only the 1971 and 1981 census data in the estimation, because the PLR definition is used in these two censuses.

Table 1: Population and Number of Migrants: 1961, 1971 and 1981 Census of India

	1961			1971			1981			Sex Ratio		
Population	M	F	T	M	F	T	M	F	T	1961	1971	1981
(in millions)	226	213	439	284	264	548	344	321	665	1061	1075	1012
Life Time Migrants (in millions)												
(a) P.O.B.Defn	42	93	135	49	109	158	57	139	196	452	452	410
(b) P.L.R Defn	-	-	-	49.6	110	159.6	59.2	142.3	201.5	-	451	416
Ratio (a)/(b)	-	-	-	0.99	0.98	0.99	0.96	0.98	0.97	-	-	-
Migrants by Duration of Stay (years)												
(P.L.R Defn)												
<1	6.4	6.8	13.2	5.5	6.3	11.8	5.1	5.3	10.4	940	873	962
1-4	12.5	18.4	30.9	13.3	18.9	32.2	15.5	24.2	39.7	676	700	640
5-9	6.6	15.2	21.8	8.0	16.2	24.2	9.8	21.0	30.8	438	494	467
0-19	4.8	12.5	17.3	9.5	26.2	35.7	12.4	33.6	46.0	363	363	369
20+	9.7	37.6	47.3	9.7	36.5	46.2	12.4	53.0	65.4	259	266	232
Period not specified	1.0	2.4	3.4	3.5	5.8	9.3	4.0	5.2	9.2	-	-	-
% of Lifetime Migration												
(a) P.O.B Defn	18.6	43.7	30.7	17.2	40.9	28.6	16.6	43.3	29.5	-	-	-
(b) P.L.R Defn	-	-	-	17.5	41.6	29.1	17.2	44.3	30.3	-	-	-
% of Intercensal Migration												
(0-9) Years	607	43.4	48.8	54.0	37.6	42.7	51.3	35.4	40.1	-	-	-
% of Recent Migration												
<1 Years	15.2	7.3	9.7	11.0	5.7	7.3	8.6	3.7	5.1	-	-	-

Sources:

(1) Part II-c (iii) Migration tables of census of India 1961. A. Mitra. Registrar General and Ex-officio Census Commissioner for India.

(2) Part V-A &B (i) Migration table (Table D-1 & D-2), Part II-D (i) Migration tables 1971. Registrar General and census Commissioner India, New Delhi.

(3) Part V A & B (i) Migration table (Table D-1 & D-2), Part D-2, Migration tables 1981. Registrar General and Census Commissioner India, New Delhi.

Estimation of the Relative Under-Reporting Migrants in Indian Censuses

For estimating the male migrants not reported in the census enumerations, assuming that the female migrants have been fully enumerated, the distribution of female migrants by duration of stay is needed. Since, the reporting of 'duration of stay' involves the recalling of time periods, it is likely to be mis-reported. Unfortunately, it seems to be difficult to evaluate and judge the extent of the reporting error in this variable. However, as it can be seen from the graphs of the cumulative number of migrants by duration of stay presented in Figure 1, it may be inferred that the cumulation of the data in duration groups seems to have reduced the effect of mis-reporting of the duration of stay to a large extent, although the effect is more evident in the case of males than in the case of females. However, in order to reduce the effect of mis-reporting of duration of stay, the cumulated data were graphically smoothed and the results are presented for males and for females in Table 2. We have also left out the 'duration not specified' group from the estimation process.

Figure 1. Cumulative Distribution of Male-Female Migrants by Duration of Stay at the Place of Enumeration: 1971, 1981 Censuses

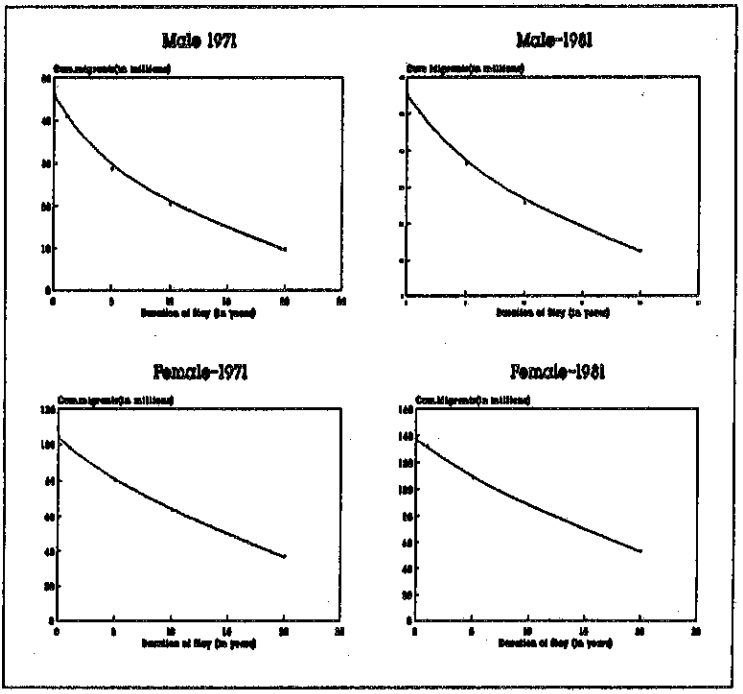

Table 2: Smoothed Number of Migrants by Sex and Duration of Stay, 1971 and 1981 Census of India

Duration of Stay (in years)	Smoothed number of migrants (from Graphical Method) (in millions)				Sex Ratio (per 1000)			
	1971		1981					
	Male	Female	Male	Female	1971	1981	Avg *	
<1	5.0	6.3	4.8	5.5	800	881	841	
1-4	12.3	17.8	14.2	23.3	691	610	650	
5-9	8.3	16.7	10.4	20.8	500	499	500	
10-19	10.6	26.8	13.4	34.5	395	388	392	
20+	9.7	36.5	12.4	53.0	266	233	250	

* Average = (1971+1981)/2

Table 3: Enumerated and Estimated Number of Male Migrants by Duration of Stay: 1971 and 1981 Census of India

Duration of Stay (in years) d	1971 Enumerated Migrants M	1971 Estimated Migrants M	1971 M/Mx100	1981 Enumerated Migrants M	1981 Estimated Migrants M	1981 M/Mx100
0-1	5.00	5.28	98.69	4.80	4.44	108.00
1-4	12.30	15.93	77.20	14.20	20.36	69.74
5-9	8.30	13.58	61.11	10.40	17.66	58.90
Intercensal						
Total	25.60	34.79	73.6	29.40	42.50	69.20
10-19	10.60	22.40	47.3	13.40	28.27	47.40
20+	9.70	30.71	31.6	12.40	44.60	27.80

Source: See Table 1.

Note

a) (i) M=Census Volume of Male Migrants
(ii) M=Nf(d). SR(0), (i.e. estimated volume of male migrants).
(iii) d=Duration of Stay at the Place of Enumeration.

b) The group " period not specified is not taken into account".

It may be observed from the sex ratios obtained from the smoothed data presented in Table 2, that there seems to be some shifting of female migrants from duration group (1-4) years to the group (<1) years in the 1971 census, and / or from 'less than 1 year' group to the group (1-4) years in the 1981 census data. In order to reduce this kind of the error, the average value of the sex ratios (see Table 2) have been used in estimating the male migrants. Thus, the smoothed values of female migrants by duration are used along with these average sex ratio values to estimate the number of male migrants and the number of male migrants who have not been reported in the census enumerations. The results are presented in Table 3, and are discussed in Section 5.

Discussion and Conclusions

It may be observed from Table 3, that out of the estimated 34.8 million male migrants, assuming that female migrants were completely enumerated, who had migrated during the decade 1961-71, only 25.6 million (i.e., 73.6%) were enumerated in the 1971 census; while, out of 42.5 million estimated number of male migrants during the decade 1971-81, about 29.4 million migrants (i.e., 69.2%) were enumerated in the 1981 census. Further, the estimates show that only about 47% of the male migrants who should have been enumerated in the duration group (10-19) years had been reported both in the 1971 and 1981 censuses. On the other hand, the proportion of male migrants enumerated in the duration group (20+) years is seen to be about 32% in 1971 and about 28% in 1981. Thus, the results clearly demonstrate that the non-reporting increases substantially with the increase in duration of stay. But, when we examine the estimated size of the migration cohort which migrated during the decade 1961-71 (i.e., 34.8 million as estimated from the 1971 census data) in comparison with the estimated size of the same migration cohort based on the 1981 census data (i.e., 28.27 million in the duration group (10-19) years), it may be inferred that the estimate for the duration group (0-9) years and that for the duration group (20+) years might be slightly overestimated. This reflects the fact that the reporting errors in reporting the duration of stay in the case of females, seem to have caused shifting of migrants from the duration group (10-19) years to (0-9) years group and also to (20+) years group. If we arbitrarily shift 2 million from the estimated migrants of (0-9) years duration group and 4.6 million from the estimate of (20+) years duration group to the (10-19) years duration group in the 1981 figures in Table 3, it may be seen that the estimated size of the migrants cohort of 1961-71 based on 1981 census data will be 34.87 million male migrants, which becomes almost exactly equal to the size estimated from the 1971 census data of (0-9) years duration group. Then the

proportion of the enumerated to the estimated number of male migrants in the different duration groups in 1981 census data becomes: 72.7% in (0-9); 38.4% in (10-19); and 31.0% in (20+) years duration groups. Instead of this, if we shift 4 million from (0-9) years group and 2.6 million from (20+) years group to the (10-19) years group, the above proportion will be 76.4%, 38.4%, and 29.5% respectively. These proportions when compared with the proportions for 1971 census shown in Table 3, seem to suggest that the smoothing has not been able to adjust fully for the shifting of migrants from (10-19) years duration group to the other groups. The error will have affected the 1971 and 1981 census data for males as well as for females.

Further, examining the proportion of enumerated male migrants to the estimated migrants in duration groups (<1) year, (1-4) years and (5-9) years, it may be observed that nearly 25% of the migrants seems to have died or returned to their place of birth in the duration group (1-4) years itself and will have not been reported in the census.

It appears, therefore, that in the Indian situation, the study of migration may be more fruitful if it is based on the (<1) year duration of stay data. Although it seems to be difficult to correct this group for any possible effect of errors in reporting the duration of stay, the error is not likely to be serious.

From the analysis presented, it may be concluded that the method suggested here, appears to be quite useful as it can provide estimates of relative non-reporting of migrants in a census, even for the concerned inter-censal period. Of course, it must be realized that there will be non-reporting of migrants of the sex assumed to be fully reported in the census (namely the female migrants in the Indian case). This is the main limitation of the method.

Thus, the results will provide indications rather than accurate estimates of the extent of the original sizes of migration cohort and the extent not reported in the census enumerations. It is hoped that this information will be of value to migration analysts. The main usefulness of the method is that it can be applied even when we have data from only one census, provided we can assume that the sex ratio of migrants with duration of stay (<1) year could be used as the SR(0) for all the migration cohorts.

References

Goldstein, S. 1964. The extent of repeated migration: An analysis based on the Danish *Population Register's Journal of the American Statistical Association* 59:1121-1132.

Internal Migration. Indian Census Centenary Seminar organized by the office of the Registrar General, New Delhi, October 23-29.

Hugo, G.J. (ed.). 1988. Internal and international flows: Some recent developments in Asia. *International Union for the Scientific Study of Population* 2:239-260.

Kadi, A.S. and M. Sivamurthy. 1988. Inter-state migration in India: 1971-1981. *Canadian Studies in Population* 51(1):37-50.

Nicholson, B. 1990. The hidden components in census derived migration data: Assessing its size and distribution. *Demography* 27(1):111-119.

Sivamurthy, M. 1981. The Deceleration of Mortality Decline in Asian Countries. International Union for the Scientific Study of Population, Manila, Conference, Manila 2: 51-76.

Sivamurthy, M. and A.S. Kadi. 1983. Inter-state migration in India: 1961-1971. *Demography India* 12: 239-258.

_____ 1984. Inter-State return migration in India: 1961-71. *Janasamkhya* 2:91-104.

Skeldon, R. 1987. Migration and the population census in Asia and the Pacific. *International Migration Review:* 21(4).

United Nations. 1970. Manual VI: Methods of Measuring Internal Migration. ST/SOA/Series A/47: United Nations, New York.

Willekens, F. and P.S. Nair. 1982. The data base for migration analysis in developing countries with emphasis on India. Working Paper No. 28. Netherlands Inter-University Demographic Institute, Voorburg, Netherlands.

Zachariah, K.C. 1967. Estimation of return migration from place of birth and duration of residence data. *International Union for the Scientific Study of Population* Sydney Conference, Sydney. 615-622.

_____.1977. Measurement of internal migration from census data. In A.A. Brown and E. Newberger (eds.) *Internal Migration.* Academic Press, New York. 121-134.

Acknowledgements

We thank the referees for their useful comments which have helped us in revising the paper.

Received, September 1994; revised, February 1996.