Tim Dyson

Analysis and Adjustment of the 1971 Indian Age Distribution, and a Reappraisal of Mortality and Fertility Estimates

THIS paper analyses and adjusts the 1971 Indian age-distribution, and attempts to gauge the prevailing levels of fertility and mortality. It uses techniques that have not hitherto been employed on data at the All India level. As with all such work, the results are not sacrosanct and must be viewed tentatively.

Unless otherwise stated, the base data is that given in the Government of India, Pocket Book of Population Statistics (India, 1972), and both the age distribution and population totals are quoted in hundreds. In the generation of intercensal life tables it is best to secure comparability of populations at the two censuses. Since only the total (final) population data for the 1961 census were available the corresponding figure for the 1971 census (total= 5,479,498) was used rather than the one per cent data (sample) which in fact led to a slight underestimate of the total population finally enumerated. Since at the time of analysis the 1971 'total' figure was not available in five-year groupings, it was proportioned by age and sex on the one per cent (provisional) data. For the remainder of the paper the one per cent sample total is used.[1]

1. The persent work was done prior to availability of the full count data. There can be no doubt that the 1% data suffers from some sampling fluctuations. However Ragavachari and Natarajan also use the one per cent distribution. They show the close correspondence between the one per cent and full count distributions, and write that "the percentage distribution by broad age groups of the 1% sample data follows closely the corresponding distribution obtained from the full count data" (1974 : 1).

Several assumptions had, inevitably, to be made in the analysis. Thus, in the generation of intercensal life tables, male and female infant mortality rates were taken as 134 and 143 per thousand live births respectively. These figures are chosen as (i) they do not diverge greatly from those assumed by several demographers working on recent infant mortality and (ii) they are the best figures available for which there is direct justification : consequently the published Sample Registration figures for 1968 and 1969 for both sexes were averaged. These figures are unlikely to be overestimates. They are indeed likely to be underestimates given the under reporting of events which occurs even in the best survey situations (Brass, 1973).

The Uncorrected Age Distribution

If we look at raw data (see Table 1) it is possible to detect several interesting irregularities.

Females		
Age	*Numbers*	
0	80,087	385,482
1	51,356	
2	84,453	
3	88,132	
4	81,454	
5	87,669	396,483
6	86,678	
7	73,214	
8	90,881	
9	58,041	
10	95,347	

(/) *Under-enumeration in the 0-4 Age Group*

It would have required a drastic fall in fertility for us to accept the enumerated size of the 0-4 age group. Whatever may be the evidence as to some recent fertility reduction it has certainly not been of the magnitude required to attach credibility to the published figures. Assuming that there has neither been a sizeable recent increase in infant mortality, nor such a fall in fertility, there are two main explanations for this deficiency. Either there has been age-shifting from the 0-4 age-group to the 5-9 age-group, or we are dealing with a case of underenumeration. A closer look, employing the single year data, will allow us to draw some conclusions as to these hypotheses.

It would seem reasonable that part of the explanation, why the 0-4 age-group is deficient lies in the argument that some children aged 3 and, particularly, 4 have been entered on the census schedule as five years old. Can we therefore assume, that the 5-9 age-group is 'too big' and that the remedy is to re-allocate persons from 5-9 to 0-4? Unfortunately we cannot, for although the 5-9 age-group has probably gained members from the preceding ages, it has also probably lost at least as many to the 10-14 age-group due to digital preference for ten. We are forced to accept the rough order of magnitude of the 5-9 figures.[2] Any 'smoothing' that effectively just re-allocates persons from 5-9 to 0-4 will be correcting the 0-4 group at the expense of the 5-9 group. Yet, the 5-9 figures are probably approximately right. The problem now becomes that of gauging how many children in the 0-4 group were missed out altogether.

Additional support for this interpretation can be gained by examining the exact numbers enumerated at ages 0, 1, 2, 3 and 4 years. It can be seen that the major deficiency of this (0-4) age-group lies at age one, and as such, is largely unconnected with shifting from 0-4 to 5-9. Why should this cohort (51,356) be so much smaller than those enumerated in the adjacent 0, 2, 3, and 4 years (average size = 83,531)? There is no reason to suggest that the cohort at age one has experienced especially heavy infant mortality. A likely explanation is that children aged one have been enumerated at age zero (digital preference), and at age zero they have been omitted altogether.[3]

Estimating the numbers under-enumerated in this first age-group (0-4), is an important step. However at this stage we might usefully ask what kind of figure might be expected. A very simple indication of the size of the 0-4 group can be gained if we raise the numbers at age one to the 'average' level of the contiguous ages 0, 2, 3, and 4 (i.e. 83,531). Then the figure for the total five year group will be : 5 X 83,531 = 417,655. This figure is probably a minimum estimate for : (i) the 'average' level is itself almost certainly an underestimate-ages 2, 3 and 4 having been open to mortality attrition for a longer period than the cohort at age one; (ii) the numbers at age zero are probably deficient due to under-enumeration; (iii) the numbers at ages 2, 3, and 4 also probably suffered through underenumeration and digital preference for the five. Finally, (iv) to

2. For support for this view, based on a detailed survivorship analysis following the numbers at 5-9 through five successive censuses, see ORG (1972 : 3).

3. For a plausible but compatible alternative explanation, see Ragavachari and Natarajan (1974 : 5).

use the 'average' level of temporally previous cohorts (ages 2, 3, and 4) to estimate the size of a subsequent cohort, is to ignore the fact that we could expect younger cohorts to be bigger than older cohorts in this population.

(*II*) *Erratic Sex Ratio Progression by Age*

The erratic progression of the ratio shown in Table 1 is probably as much due to differential sex reporting by age, and age-shifting, as to any 'real' sex ratio change. Perhaps the most interesting feature is the pronounced 'dip' at ages 20-29.[4]

TABLE 1—INDIA : AGE AND SEX CLASSIFICATION 1971 (IN HUNDREDS) (ESTIMATED FROM ONE PER CENT SAMPLE DATA)

Age Group	*Males*	*Females*	*Sex Ratio m\f*
0-4	397,889	385,482	103
5-9	424,193	396,483	107
10-14	358,411	317,777	113
15-19	252,531	222,949	113
20-24	214,362	216,141	99
25-29	200,081	205,394	97
30-34	181,158	179,463	101
35-39	171,161	156,783	109
40-44	150,351	132,546	113
45-49	123,585	103,740	119
50-54	110,743	93,883	118
55-59	67,043	58,125	115
60-64	74,084	68,374	108
65-69	35,377	32,415	109
70+	58,475	56,180	104
Total	2,819,471	2,625,735	107

4. This has been a feature of most Indian age-distributions since 1901 *(India* 1963 : 39, Table 2).

(*III*) *Female Shifting into the Reproductive Age-Span*

The low sex ratios at ages 20-34 are similar to an irregularity common in other data (Brass *et al.,* 1968 : 43). Basically women aged 15-19 overstate age, whilst those 35-49 tend to understate it. One effect is to produce abnormally low sex ratios at ages 20-34, and compensating high ratios in contiguous age-groups.

Additional support is given to this age-shifting hypothesis when we consider that the data give 421,535 females (16.1%) between 20 and 29 and only 414,443 males (14.7%). This is in spite of the fact that the total male population is considerably bigger than its female counterpart.

Given that we are probably dealing with an enumerated 'surplus' of women in the reproductive ages, this has important consequences for any attempt to correct for deficiency of numbers at ages 0-4.[5] An approach to such correction is to take the numbers of women in the reproductive ages, project them backwards five years on the basis of a relevant mortality schedule, and finally, assuming some Total Fertility Rate, (TFR), fertility distribution, and infant/childhood mortality rate $({}_5P_b)$ to generate the number aged 0-4. However, if we were to estimate this first (0-4) cohort using the enumerated numbers of women in the fertile age-range, the result would be an overestimate because the number of fertile women in the raw data is inflated. Therefore, before any attempt at gauging the size of the 0-4 group can be made, the female age data must be smoothed.

Before examining the problem of the defective 0-4 cohort we must compute abridged Life Tables for the 1961-71 period. These are interesting in themselves, and will play a part in the adjustment of the over-all age-distribution. Here we will concentrate on the production of the female Life Table, although the male results given in Table 3 are arrived at by the same procedure.

The Intercensal Abridged Life Tables

The technique used in this section was developed by Brass (1969 : 187). The object is to produce a two parameter Life Table from intercensal survivorship

5. Support for the age-shifting hypothesis can also be derived from an abnormally high ten year intercensal survivorship ratio for females. ($10^{P}15\text{-}19$ = 1.1926), for the period 1961-1971.

ratios. The production of an intercensal life table for the period 1961-1971 is probably more relevant for the present work than LT's published for earlier periods. At the same time it is probably little removed from the mortality level pertaining to the latter half of the period.

The basic method is as follows : the aim is to compute a stationary population with reference to the intercensal survivorship ratios. From the age-distribution at the two censuses ${}_{10}P_{x+10}$ values are computed i.e., ten year intercensal survivorship ratios (see Table 2). Likely values of ${}_5L_0$ and ${}_5L_5$ are then assumed (in this case ${}_5L_0$ = 41451 and ${}_5L_5$ = 37823, ∴ L_{0-9} = 79274). These values were taken from a Brass one-parameter model 45 (e_0 — 42.5) (Carrier and Hobcraft, 1971). Although overall Indian female mortality is certainly lower than that of the model, the infant mortality rate of the model (127‰) is considerably lower than the rates usually estimated for Indian females (we have taken 143‰), and these balancing considerations indicate that the figures are reasonable. Computaion of the stationary population is dictated solely by the intercensal survivorship ratios, through a chaining technique: we have an estimate of ${}_{10}P_{0-9}$ (-841134), if we multiply this by L_{0-9} we get L_{10-19} from this (using a separation factor taken from relevant one parameter models) we get L_{10-14} and L_{15-19}[6] We also have ${}_{10}P_{10-14}$(.938452), if we multiply this by L_{10-14} (obtained) we get L_{20-24} (31701). Continuing the process we get:

$${}_{10}P_{15-19} = 1.192606 = \frac{L_{25-29}}{L_{15-19}}. \quad \text{But } L_{15-19} = 32900$$

$$\therefore \quad L_{25-29} = 39237.$$

The process is repeated right down the age scale; at each age we get an estimate L_x i.e. the stationary population, and from this a set of l_x values are obtained using the following approximations :

1/5 ${}_5l_0$ = $l_{1.6}$, 1/5 ${}_5L_5$ = $l_{7.5}$, 1/5 ${}_5L_{10}$ = $l_{12.5}$. 1/5 ${}_5L_{15}$ — $l_{17.5}$ • • etcetera.[7]

As can be seen from column 7 (Table 2) the lx values are not smooth (much of the disturbance, emanating from the survivorship ratio greater than one),

6. The use of unsmoothed survivorship ratios is deliberate (see Brass 1969 : 187).

7. 1/5 ${}_6/x$ + a is a close approximation to the value of $\frac{lx = a}{2}$. e. the mid-point of the age-group. This holds for all age-groups except the first, where, due to the preponderance of deaths in the early part of the period 0-4, the approximation is to age 1.5 years.

TABLE 2 — COMPUTATION OF ABRIDGED LIFE TABLE, 1961-1971 (FEMALES) FROM INTERCENSAL SURVIVORSHIP RATIOS

Age group 1961	*Population 1961*	*Age Group 1971*	*Population 1971*	*Ten Year Inter-censal Survivorship Ratios($_{10}P_x$)*	*The Stationary Population $_5L_X$*	$1/5\ _5L_x \simeq l_{x+2\cdot5}$	*Standard One Parameter level 45 115 $_5L_x \simeq l_{x+2\cdot5}$*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0-9	*64,514,987*	*10-19*	54,265,760	0.841134	41451 and 37823	8290 and 7565	
10-14	23,031,631	20-24	21,614,085	0.938452	33780*	6756	7305
15-19	17,283,838	25-29	20,612,815	1.192606	32900	6580	7114
20-24	19,133,698	30-34	18,010,439	0.941294	31701	6340	6839
25-29	18,049,551	35-39	15,734,383	0.871733	39237	7847	6529
30-34	14,854,603	40-44	13,302,031	0.895482	29840	5968	6222
35-39	11,860,362	45-49	10,411,089	0.877805	34204	6840	5904
40-44	10,773,683	50-54	9,421,832	0.874523	26721	5344	5557
45-49	8,322,495	55-59	5,531,072	0.664598	30024	6004	5158
50-54	7,979,772	60-64	6,861.698	0.859886	23368	4673	4683
55-59	4,549,072	65-69	3,253,168	0.715122	19954	3991	4117
60+	12,355,422	70+	5,637,998	0.456318	20094	4019	3440

L_{10-14} was assumed *to* equal 0.5066 L_{10-19}.

but even so comparison of our obtained data, with the similarly translated "standard" values (columns 7 and 8) indicates that at younger ages mortality is higher than standard, while at older ages it is less than standard. Since we have already accepted a specific level of infant mortality, (IMR = 143), we are only interested in employing some average of adult female mortality. For this purpose those l_x values ($l_{32.6}$, $l_{37.5}$, $l_{42.6}$, $l_{47.5}$) from age-groups (30-34, 35-39, 40-44, and 45-49) were averaged. These particular points are preferred as they not only represent the bulk of the adult female population, but also neatly straddle age-groups into and from which age-shifting takes place.

In order to derive a Life Table from these adult *Ix* values the logit system of model life-tables is used (Brass *et al.*, 1968). The system is defined by the following equation:

$$Y(x) = a + \hat{a}\ Ys\ (x),$$

where $Y(x)$ is the logit of the observed values and Ys (x) is the logit of the standard values, a is an approximate indication of mortality level; p is a measure of the 'slope' of the mortality configuration relative to a standard. Given that we have two groupings of l_x values for both observed and standard life-tables, we can obtain their logits and subsequently solve for alpha and beta.

The average of the observed adult l_x logits is .2133 and the average of the standard adult l_x logits is .1439. The logit of the observed l_i (8570) (assumed) is $y/_1$= .8953. The corresponding standard value is $Y_3\ /_x$ = .9646. Solving for alpha and beta we get alpha = 0.09373 and beta = 0.8301. This Life Table could be generated from these values and the logits of the one-parameter level 45.

By harnessing one of the advantages of this system—that any two l_x values can be 'joined up', two values l_1 and l_{10} and their corresponding logits, were used to 'shift' alpha and beta relative to the Afro-Asian standard (Brass *et al.* 1968 :133). The new values derived were alpha = .0666 and beta = .8312. These values relate the Indian female mortality pattern and level to the Afro-Asian standard. Using this new standard, a complete life table was computed (Table 3).

The results seem plausible and this is perhaps their main claim to validity, given the nature of the work. The expectation of life at birth for females is

TABLE 3—MALE AND FEMALE ABRIDGED LIFE TABLES (1961-1971), l_x $L_x T_X$ AND e_x FUNCTIONS

Age	Males l_x	Males $_nL_x$	Males T_x	Males e_x	Females l_x	Females $_nL_x$	Females T_x	Females e_x
0	10000	9062	494643	49.46	10000	8999	457611	45.76
1	8660	32926	485581	56.07	8570	32298	448612	5235
5	7945	38997	452655	56.97	7744	37867	416314	53.76
10	7653	38007	413658	54.05	7403	36717	378447	51.12
15	7550	37332	375651	49.75	7284	35932	341730	46.92
20	7382	36367	338319	45.83	7089	34810	305798	43.14
25	7163	35280	301952	42.15	6835	33552	270988	39.64
30	6948	34200	266672	38.38	6586	32307	237436	36.05
35	6731	33077	232472	34.54	6337	31022	205129	32.37
40	6499	31837	199395	30.68	6072	29612	174107	28.68
45	6235	30395	167558	26.87	5773	27985	144495	25.03
50	5922	28632	137163	23.16	5421	26022	116510	21.49
55	5531	26447	108531	19.62	4988	23627	90488	18.14
60	5049	23672	82084	16.26	4463	20660	66861	14.98
65	4421	20202	58412	13.21	3801	17072	46201	12.16
70	3660	38210	38210	10.44	3028	29129	29129	9.62

estimated at 45.8 years, and that for males 49.5 years. Although these figures, based on adult mortality at age-groups 30-49, are preferred and used in subsequent analysis, they are given extra weight by their close correspondence to those obtained if instead adult age-groups 25-44 and 34-54 are averaged (Table 4).

TABLE 4-LIFE TABLE PARAMETERS RESULTING FROM ALTERNATIVE AVERAGING OF AGE-GROUPS ($l_{x+2.5}$ VALUES) FROM TABLE 2

Age Groups Averaged	*Males*				*Females*			
		â	e_0	$_5P_b$	a	â	e_0	$_5P_b$
25-44	0.1975	0.7377	50.3	0.8400	0.1360	0.7616	48.1	0.8285
30-49*	0.1703	0.7650	49.4	0.8397	0.0666	0.8312	45.7	0.8259
35-54	0.1915	0.7437	50 1	0.8397	0.0666	0.8312	45.7	0.8259

•Used in the production of the Life Tables of Table 3.

As with all intercensal survival methods, equal levels of enumeration at the two censuses are assumed. In this instance the assumption pertains only to the relevant adult age-groups. The official post-enumeration checks for 1961 and 1971 suggested undercounts at all ages of 0.7% and 1.7% respectively *(India,* 1964; *India,* 1972a). If, for illustrative purposes we assume these percentage undercounts at adult ages (30-49) and the same infant mortality rates, and follow the same procedure as above, then male and female life expectations are raised to 51.7 years (a = .2320, â = .7032) and 47.6 years (a = 0.1220 and â = .7757) respectively.

But in the absence of data on the age-distribution of omissions, both common experience, and argument below suggest that a large proportion of omissions occur at young ages. The age-groups employed here, should be relatively omission free. Again, it is difficult to know exactly how much weight can be placed on the PEC results. For these reasons the earlier life expectations are to be preferred with the proviso that, if anything, they may be slight underestimates. However, the issue does serve to underline the tentative nature of results,[8]

8. The age distribution of omissions is not relevant when dealing with figures of total population. If we accept 0.7% and 1.7% undercounts of total population, then the annual intercensal growth rate is raised from 2.2% to 2.3% per annum, and this affects any estimates (e.g. Adlakha and Kirk 1974 : 386) based on the equation: intercensal growth rate = intercensal birth rate—intercensal death rate.

One other point is worth making. Almost all works on Indian mortality, unless based directly on empirical data (e.g. SRS) have used one parameter life-tables. Yet as recent work (Adlakha 1972 : 581) has shown, one parameter life tables cannot adequately represent the relatively high levels of infant mortality (relative to mortality in adulthood) found in such data. This analysis has shown India to be no different—even given that the levels of infant mortality assumed here are if anything underestimates. In fact, all the derived life tables exhibit low beta values, implying that at least two parameters are required to incorporate high infant/childhood mortality with the relatively lower mortality of adult ages.

Smoothing the Female Age Distribution

The method used to smooth the female age distribution was developed by Brass (1969 :184-6). First a reference-distribution 'model' is chosen as standard: in this case a three parameter stable population model, fitted to the uncorrected age-distribution for both sexes combined. (See Carrier and Hobcraft, 1971 pp 114-203). The model was fitted by the following characteristics:

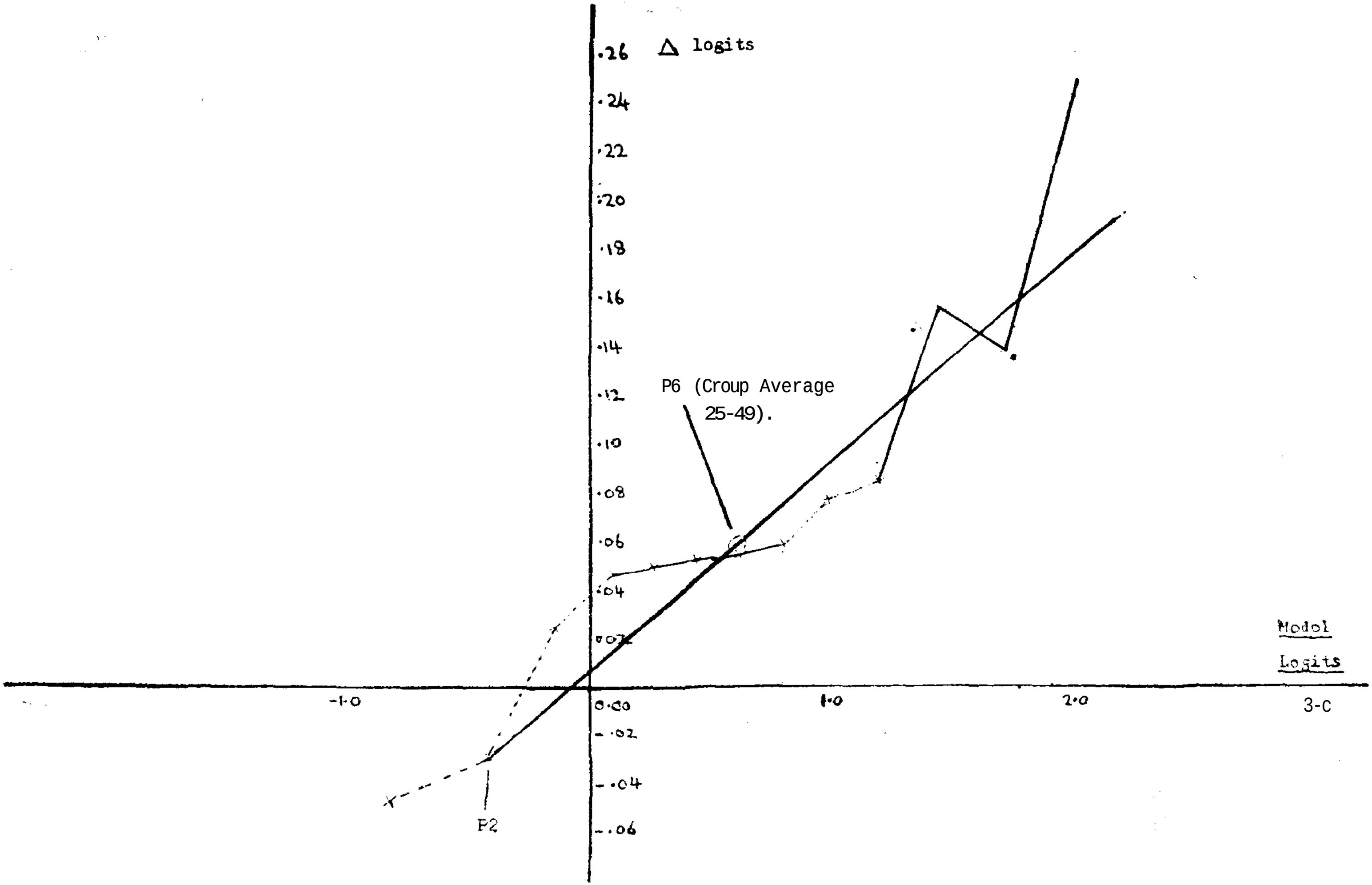
Percentage of population 0-14 *(P)* = 42,

Percentage of population 15-44 (0 = 14,

Intercensal (1961-1971) growth rate V = 2.2% per annum.[9]

Secondly, both the observed female and model (standard) distributions were percentaged on their respective populations aged five and over (Table 5 columns 2 and 5). (This last step was taken as any percentaging of the observed female data on its 'total' inclusive of 0-4, would be incorrect due to the fact that the 0-4 group is under-enumerated). The percentages for both model and observed data were then cumulated and their respective logits calculated (columns 4 and 6 of Table 5). Brass terms introduction of logits a 'linearizing transformation' because it brings the relationship plotted (between standard and observed values) closer to a straight line, and thus accentuates deviations from the general trend. Taking this process of magnification one step further, the female logits were subtracted from the model logits to give 'model minus female' logits (i.e. A logits column (7)) and it is these that were smoothed against the

9. For the purpose the Indian age-distribution can be taken as effectively closed. In an application of the method Brass has written that "It is important to note that although the reference standard is a stable population, the age-distribution corrected by these means is, in this example, far from stable. The requirements of smoothness in the differences is much less rigid than a constraint to the stable model form" (Brass 1969 : 185).



Graph 1. Model logits against 'Model minus female' logits (A logits).

TABLE 5—SMOOTHING OF THE FEMALE AGE DISTRIBUTION AGED FIVE AND **OVER**

Age Group	*Model % Distri-bution*	*Model % Cumu-lated*	*Model logits*	*Female % Distri-bution*	*Female logits*	*A logits (model minus female)*	*"New" A logits*	*New Female logits*	*Female % Distri-bution corrected*	*Female Distri-bution (hundreds)*
(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(9)*	*(10)*	*(11)*
5-9	16.36	16.36	-.8158	17.70	-.7684	-.0474	-.0474	-.7684	17.70	396,483
10-14	14.22	30.58	-.4099	14.18	-.3797	-.0302	-.0302	-. 397	14.18	317,777
15-19	12.44	43.02	-.1405	9.95	-.1649	.0244	-.0060	-.1345	11.44	256,218
20-24	10.75	53.77	.0755	9.65	.0296	.0459	.0123	.0632	9.83	220,217
25-29	9.22	62.99	.2659	9.17	.2163	.0496	.0285	.2374	8.50	190,421
30-34	7.89	70.88	.4448	8.01	.3921	.0527	.0440	.4008	7.38	165,331
35-39	6.71	77.59	.6210	7.00	.5671	.0539	.0584	.5626	6.47	144,944
40-44	5.67	83.26	.8021	5.92	.7440	.0581	.0756	.7271	5.56	124,558
45-49	4.70	87.96	.9943	4.63	.9164	.0779	.0917	.9026	4.82	107,980
50-54	3.82	91.78	1.2064	4.19	1.1212	.0852	.1100	1.0964	4.08	91,402
55-59	2.99	94.77	1.4486	2.59	1.2925	.1561	.1300	1.3186	3.36	75,272
60-64	2.20	96.97	1.7329	3.05	1.5943	.1386	.1556	1.5773	2.59	58,022
65-69	1.49	98.46	2.0792	1.45	1.8298	.2494	.1790	1.9002	1.90	42,565
70+	1.54	100.00		2.51					2.19	49,061
Total	100.00			100.00					100.00	2,240,251

model logits (Graph 1). "The points follow a gentle curve at childhood ages and are close to linearity in adulthood." (Brass, 1969 : 186).

However, the need for curvilinearity is lessened by the fact that we have neglected the first childhood age group. Somewhat ruthlessly perhaps, the "group average" method suggested by Hobcraft and Carrier (1971 : 15) was utilized at first. However, this led to a substantial reduction of the numbers aged 10-14 from the 317,668 enumerated to 307,139. Although it is true that there are strong pressures to understate age at this time of life, there are perhaps even stronger pressures to overstate it. With this in mind the final graduating was done using a 5 age group (25-49) average at higher ages, but at younger ages passing the line through point *P2*, i.e. maintaining the same proportion of population between ages 5 and 14. Therefore a straight line was drawn through *P2* and P6 (25-49). Even at such a level of abstraction the graphing of points shows a curving of points 3, 4, *5*, *6* and 7 which is suggestive of a glut of fertile females, and the zig zag of points at later ages is indicative of digital preference for age-groups containing a zero. Knowing the standard logits, and having now obtained 'new' values for logit A differences, new female logit values can be calculated and by reversing the process the smoothed percentage female distribution was obtained (column 11). It is important to realise that the corrected distribution has a considerably different form from that of the standard model stable population.

The results are certainly an improvement on the raw data, although the method of darwing a straight line by group average cannot hope to reflect every small peculiarity of the 'true' age-structure—whatever that might be. However, we can note that generally the structure now falls off more smoothly with age. At higher ages this is especially so : the data no longer exhibit the tendency to 'bunch' in the age-groups containing a zero (e.g. 60-64). Not only have the total numbers 15-49 been reduced slightly but more important is the shifting of women, back from the prime reproductive years 25-39, to the peripheral age-groups from which they came, i.e. 15-19 and 45-49. One final feature is the isolation of a possible defect hitherto unmentioned : age overstatement at higher ages. The raw data give 56,180 females aged 70 or more. The adjusted distribution implies only 49,061—a 13% reduction.

Under-Enumeration at Zero-to-Four

We are now faced with the problem of the first infant and childhood age-group. One approach, would be to project the age-corrected fertile females through reverse five year survivorship ratios derived from the life-tables: averaging these females with the numbers enumerated at the end of the period would give an approximation of the number of women who, over the five years 1966-71 were to produce the 0-4 group. Assuming a fertility level, the five year's births (1966-71) could be generated and "survived" through the probability of surviving from birth to ages 0-4 (i.e, $_5P_b$) itself derived from the life table. The question now becomes which fertility level to choose? Lee Jay Cho (1964 : 363) estimates Indian total fertility at 5.4 for the period 1955-1960. The census Actuary estimated it at 5.5 for 1951-61 *(India,* 1972 *c :* 7). A recent projection (Zachariah and Cuca, 1972) used TFR = 5.4 (1970). Tomas Frejka (1973) for the period 1965-70 uses 5.6, while Adlakha and Kirk (1974 : 396) for the period 1961-71 estimate total fertility at 5.7-5.8. Somewhat out on a limb are the Coale-Demeny estimates (UN, 1967 : 28) of 6.4-6.6 live births for 1961.

There are reasons to suppose that most of these fertility estimates are too low. The Indian sample registration system, utilizing a multicheck survey approach, itself gives estimates of 5.76 for 1968 and 5.62 for 1969.[10] *(India, 1972c :* 95). Turkish experience with this survey approach gave birth under-enumeration of 13%. Several writers have highlighted possibilities for event-underenumeration : perhaps most important is the problem of infants dying within a few days of birth. So despite the fact that India's experience with this kind of survey technique has been relatively good, it is unlikely to have been complete in coverage. There have been several limited checks. Agrawal (1969) refers to two resurveys in 1965-66, one in Gujarat, the other in Kashmir, which detected 4% and 13% of live births 'missed' respectively. More recently (1973) Lingner and Wells write "an intensive resurvey conducted in a random sample of units in Kerala State as part of field verification procedures, uncovered sizeable number of vital events, amounting to 8% of total births and 5% of total deaths, missed by both recorders and interviewers". What little checking there has been then, implies fertility underestimates of 4%, 8% and 13% (average 8.3%). This would loosely imply a TFR somewhere between 5.9 and 6.4.

We can also note the recent work of Soni (1975). Working with the national survey data of the Operations Research Group Baroda, she analysed the data on current and completed fertility using the Brass technique (Brass *et al.,* 1968 : 92) and the Coale Trussell method (1974 : 195) to arrive at estimates for 1970 of 6.17 and 6.03 respectively.

10. It should be noted that the Sample Registration Syetem is less complete in urban areas which certainly have lower fertility.

A part-alternative to accepting some compromise of these estimates is afforded by a series of equations developed by Hobcraft and Carrier (1971). Working on computer-based quasi-stable population structures, they attempt to gain a fertility measure from data on age-structure. The lead-in variable is the child-adult ratio (CAR) which is derived solely from age-structure (i):

$$CAR = \frac{\text{Population Aged 0-14}}{\text{Population Aged 15-44}} \qquad (j)$$

Experimenting on transitional populations and assuming an appropriate set of age-specific fertility weights, they related the CAR to a reproduction rate intermediate to the gross-reproduction rate (GRR) and net reproduction rate (NRR) (equation ii) which itself was denned partly in terms of GRR (equation iii). The intermediate index they named the Mean Reproduction Rate (MRR) (iii).

$$MRR = CAR(1.64 + 0.8\,CAR) \qquad (ii)$$

$$MRR = GRR\ \frac{l_2}{l_0} \qquad (iii)$$

Given alpha and beta, and the logit value of the Afro-Asian standard l_t, we can compute the relevant female l_2 value using $Y_x = \alpha + \beta Y_{sx}$. The estimate obtained is l_2 = 8133. From a GRR, and assuming a sex ratio at birth, in this case 104, we can estimate the relevant TFR. Using the enumerated age-group at zero to four, the above series of equations was utilized and an estimate of total fertility achieved : TFR = 6.1.

Now if this estimate is taken, and a reasonable set of age-specific fertility weights assumed it can be applied to the distribution of women who would generate the births of the first five years of the age-distribution. These births can be 'survived' using ${}_5P_b$. This new estimate of the zero to four group can then be used in the computations of a new CAR, which, in turn, will lead to a new TFR. If this iterative procedure is followed to equilibrium the TFR achieved is 7.0.

However, it must be remembered that estimates from these series of equations, *qua* age-structurally derived, using age-groups 0-14 and 15-44, in fact are implied by the age-structure as a whole. They do not account for any recent fertility reduction. A TFR of 7.0 probably represents an approximate level of the past, from which fertility has subsequently declined.

With all these considerations in mind, a TFR of 6.1 was assumed with the caveat that this exact figure must be viewed with some caution. It was distributed on the average of the two sets of available age-specific fertility weights.[11] Table 6 shows the estimation of the number of women alive, by age-group, in the reproductive years 15-49 in 1966. These populations are averaged with the smoothed 1971 populations at these ages (see columns 5, 6, and 7). The weighted TFR of 6.1, is then related to those women in the reproductive years to give an estimate of the five year's female births, (523,808), which finally are survived through ${}_5P_b$ (bottom of Table 6) to give an estimate of 432,602 for those aged 0-4. The estimate for the first male age-group is 457,438.

Further Adjustments

In Section I it was argued that the main reason for the deficiency in the 0-4 age-group was complete omissions. However, we also noted that there is support for a relatively small part of the deficiency, especially at age four, being attributable to preference for the '5'. It follows that the numbers enumerated at 5-9, which until now we have accepted, should be slightly reduced to take this into consideration.

Fitting a quadratic or polynomial curve to selected age-groups can lead to a slight reduction of numbers at 5-9, (e.g. fitting quadratic multipliers (Carrier and Hobcraft, 1971: 204) to age-groups 5-14, 15-24, 25-39 reduces the number of females at 5-9 from 396,483 to 394,162). However, such procedures were rejected because (i) if consistently applied they would alter age-groups already adjusted, and (ii) because it is generally accepted that such fitting procedures are hazardous when dealing with young age-groups.

Faced with this difficulty a simpler, if cruder adjustment procedure, of the type used previously (Section I) was adopted. Our new estimate of the 0-4 age-group shows that it was underenumerated by 385,482/432,602 i.e. 12%. If we assume that the underenumeration estimate for the whole of this age-group applies to the numbers enumerated at exact age four (81,454 see Section

11. These in fact refer to the rural population (see *India,* 1972*b* : 98).

TABLE 6-ESTIMATION (*i*) OF THE NUMBER OF FEMALES THAT, OVER THE PERIOD 1966-1971, PRODUCED THE 0-4 AGE GROUP, AND (*ii* OF THE 0-4 AGE GROUP IN 1971

Age in 1971	*Numbers of Females 1971*	*Reverse 5 Year SRS*	*Age in 1966*	*Projected Female Population 1966*	*Population (female 1971) (corrected figures)*	*Average Female Population 1966-1971*	*Age Specific* Fertility Weights*	*Age Specific Fertility rates (TFR=6.l)*	*Births 1966-1971*
V)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
20-24	220,217	1.032176	15-19	227,303	256,285	241,794	.091	0.555	134,196
25-29	190,421	1.037496	20-24	197,561	220,217	208,889	.226	1.379	288,058
30-34	165,331	1.038538	25-29	171,702	190,421	181,061	.237	1.446	261,814
35-39	144,944	1.041424	30-34	150,948	165,331	158,139	.198	1.208	191,032
40-44	124,553	1.047619	35-39	130,489	144,944	137,716	.145	0.884	121,741
45-49	107,980	1.058142	40-44	114,558	124,558	119,408	.070	0.427	50,987
50-54	91,402	1.075441	45-49	98,297	107,980	103,138	.033	0.201	20,731
									1.068,559

On a Sex Ratio at birth of 104 the total births consist of 523,808 females and 544,751 males.
Of the females : 523,808 x 0.82538 = 432,602 survive to 0-4 in 1971.
Of the males : 544,751 x 0.83972 = 457,438 survive to 0-4 in 1971.
*Taken from Pocket Book of Population Statistics *(India,* 19726, p. 98).

I), then the numbers enumerated at age four should have been 91,411. Alternatively we can say that the estimated deficiency at exact age four is 91,411 — — 81,454 = 9,957. Now while it is probably true that a small part of this deficiency was not due to preference for '5', and in fact represents complete omissions, it is also probably true that a few children aged three (in addition to those aged four) also showed preference for the five. Given these balancing considerations it seems reasonable to accept 9,957 as a rough estimate of those females who were transferred from 0-4 to 5-9. On this basis the numbers

TABLE 7—ADJUSTED AGE-DISTRIBUTION (1971) AND ASSOCIATED CHARACTERISTICS

Age Group	*Males*	*Females*	*Sex Ratio m\f*
0-4	457,438	432,602	106
5-9	411,174	386,546	106
10-14	340,821	317,777	107
15-19	268,270	256,285	105
20-24	231,185	220,217	105
25-29	201,819	190,421	106
30-34	176,484	165,331	107
35-39	159,210	144,944	110
40-44	139,633	124,558	112
45-49	121,207	107,980	112
50-54	103,069	91,402	113
55-59	84,931	75,272	113
60-64	65,642	58,022	113
65-69	47,504	42,564	112
70+	54,126	49,061	110
Total	2,862,513	2,662,983	107

TFR = 6.1 (1966-71)
Male e_o = 49.43 (1961-71)
Female e_a = 45.75 (1961-71)
Total Population = 5,525,496

enumerated at 5-9 should be reduced by 9,957 to 386,546 in the case of females. The corresponding figure for males aged 5-9 is 411,174 implying that, both sexes combined, just over two million persons aged 0-4 were enumerated at 5-9. The complete adjusted female age-distribution for 1971 is given in Table 7.

The graduated male distribution can now be obtained (Table 7) using almost exactly the same procedure as was used for females (Section 4), only now the whole population, inclusive of 0-4, can be used. Moreover, the smoothing procedure is simplified in that we must accept 457,438 as the size of the 0-4 male age-group and 411,174 as the numbers aged 5-9.

Discussion and Conclusions

Although there is nothing sacrosanct about the exact figures estimated or argued for here, several points emerge from the results. Firstly, it will be remembered that in Section I we computed a first estimate of the size of the female age-group 0-4. For several reasons this estimate (417,655) must be regarded as a minimum. Even so, it implies an undercount of females at 0-4, of three and a quarter million and if the same calculations are performed for males, at least 6.5 million persons were omitted in this first age-group. The 'final' estimate of the first female age-group (432,602), based on TFR=6.1 increases the estimate of female underenumeration at 0-4 to four and three-quarter million. While combining the data for both sexes and assuming the estimate to be correct, over ten million children were underenumerated in this age-group alone—a deficiency of over 12%. On the calculations above (Section 6) just over two million of these were enumerated at 5-9. The remaining eight million are 'complete comissions.'

It follows that, adjustment of the Indian age-structure cannot be adequately accomplished without introducing estimates of the numbers completely omitted at 0-4—and thus by implication raising the estimate of total population.[12]

12. Working with the Census Evaluation Study results, Ragavachari and Natarajan make the same point. They write that "there was substantial underenumeration in the 0-4 age-group apart from a sizeable chunk of net transfer of population from it to the age-group 5-9 due to age distortion", and that "the results of the Census Evaluation Study provided for the first time a mechanism for building up an acceptable pattern of the age curve instead of following the practice of extrapolating the curve to the younger ages which was

Footnote 12 *(contd. on page 91)*

Just on the basis of the 0-4 estimate, and allowing for some shifting into the age-group 5-9, the present analysis has raised the total population figure from 5,454,958 to 5,525,496. This alone represents a 1.3% undercount of total population. It supports Visaria's (1971 : 1439) suggestion—based on comparison of population and household estimates—of a substantial undercount of population in 1971.

Finally, if we are right in our analysis that fertility has declined from 7.0 live births per woman to approximately 6.1 live births per woman, then the number of persons aged 5-9 should be consistent with some intermediate level of total fertility. In fact, assuming the intercensal life tables of Section 3 (e_0 = 49.4 years (males) e_0 = 45.7 years (females)) the TFR for 1961-66 implied by the numbers aged 5-9 (Table 7) is 6.57 live births per woman. This is comparable with the highest estimates of Indian fertility—those of Coale and Demeny for 1961 of 6.4 to 6.6 live births per woman. It is difficult to arrive at an exact figure for total fertility for the early sixties. The truth is probably some compromise of these estimates. However, in general we can say that there are indications of a substantial fertility decline between 1961-66 and 1966-71. This might well be connected with the intensification of the nation-wide family planning programme in the latter part of the decade.

Footnote 12 *(contd. from page* 90)

inevitable in the earlier censuses" (1974 : 2). The inflation factors and omission rates they present (see Tables 4 and 5 pp 4-5) imply that there were approximately seven million complete omissions at 0-2—a figure comparable with that in the present work. However they follow the established practice of prorating their adjusted distribution on the enumerated total. In contrast, it is worth noting that one of the main conclusions of the study of the 1961 distribution was that "the enumerated population aged 0-7 is fairly near the true population for the age group" *(India* 1963 : 6). and that "the Post Enumeration Check of 1961 does not show any differential omission of a substantial extent in any age/sex group" *(India* 1963 : 28). There is no particular reason to expect that the enumeration errors estimated for 1961 will be identical to the errors of 1971. On the contrary, if there is more under-enumeration in general in 1971 than in 1961, as the post-enumeration check suggests, then underenumeration in the 0-4 age-group in particular is likely to be proportionately greater. In view of this one should be chary about assuming that age-adjustments for 1961 are adequate when dealing with the 1971 age-distribution (see for example Rele and Sinha 1973 : 269 and Adlakha and Kirk 1974 : 385). *Ceteris paribus,* to the extent that such analyses neglect complete omissions at 0-4 in 1971 they will—whether based on survivorship methods or to a greater extent on age structure (e.g. quasi-stable analysis)—lead to underestimates of the level of fertility. These general arguments are detailed elewhere (see Cassen and Dyson, forthcoming).

In concluding, several statements can be tentatively forwarded. To begin with, the analysis has shown that when applying model life-tables to India, at least two parameters are required so as to combine the relatively high infant/ childhood mortality with that of later ages. Turning to the age-distribution there can be little doubt that there is a considerable undercount at 0-4 and that this cannot be accounted for merely by reallocating persons from 5-9. As a corollary, the population is probably larger than that enumerated. The analysis also implies that fertility in the early 1960s was well above current levels. However, there are reasons to believe that total fertility is higher than is often supposed.

Acknowledgements

The author wishes to thank John Hobcraft, William Brass, Robert Cassen, and the anonymous referee, for helpful suggestions, and Sue Evans for typing the manuscript.

References

1. Adlakha, A., 1972, Model life tables, an Empirical Test of their Applicability to less Developed Countries, *Demography, 9,* 1972.
2. Adlakha, A. and Kirk, D., 1974, Vital rates in India, 1961-71. Estimated from 1971 Census Data, *Population Studies,* 28(3).
3. Agrawal, B. L., 1969, Sample registration in India, *Population Studies,* 33(3).
4. Brass, W., *et al.,* 1968, *The Demography of Tropical Africa,* Princeton : Princeton University Press.
5. Brass, W., 1969, *Disciplining Demographic Data,* International Population Conference; London 1969, Vol. 1, Liege : The international LInion for the Scientific Study of Population.
6. Brass, W., 1973, *A Critique of Methods of Estimating Population Growth in Countries with Limited Data,* Laboratories for Population Statistics, Reprint Series No. 4, The University of North Carolina.
7. Carrier, N. and Hobcraft, J., 1971, *Demographic Estimation for Developing Societies,* Population Investigation Committee, London School of Economics.
8. Cassen, R. H. and Dyson. T., Population projections for India. Forthcoming, in *Population and Development Rev'ew,* The Population Council, New York.
9. Cho, Lee J., 1964, Estimated refined measures of fertility for all major countries of the World, *Demography,* 1(1).
10. Coale, A. J. and Trussell, T. J., 1974, Model fertility schedules : Variations in the age structure of childbearing in human populations, *Population Index,* 40(2).
11. Frejka, T., 1973, *The [Future of Population Growth : Alternative Paths to Equilibrium,* Population Council, New York.

12. India (Census of) 1962, *Final Population Totals,* Paper No 1 of 1962 Office of the Registrar General. New Delhi.
13. India (Census of) 1964, *Age Tables,* Paper No 2 of 1963, Office of the Registrar General, New Delhi.
14. India (Census of) 1972a, *Post Enumeration Check,* (Preliminary Results), Office of the Registrar Generel, (mimeo) New Delhi.
15. India, 19726, *Pocket-Book of Population Statistics.* Office of the Registrar General. New Delhi.
16. India, 1972c, *Measures of Fertility and Mortality in India.* SRS Analytical Series No 2. Office of the Registrar General, New Delhi.
17. Linger,J. W. and Wells, H. B., 1973, *Organisation and Methods of the Dual Record System in India,* Laboratories for Population Statistics, Scientific Report Series, No 9. The University of North Carolina at Chapel Hill.
18. ORG (Operations Research Group, Baroda), 1973, *Population of India-A Projection: 1971-2001,* Baroda, 1973.
19. Ragavachari, S. and Natarajan, K. S., 1974, *Smoothed Age Data of the 1971 Census and Life Tables, 1961-71.* Paper 2 of 1974, Miscellaneous Studies, Registrar General. New Delhi.
20. Rele, J. R. and Sinha, U.P., 1973, Fertility and Mortality in India 1951-1971-A census analysis, *Demography India,* 2(2), New Delhi.
21. Soni, V., 1975, *A Demographic Analysis of the Sterilization Program in the Indian States : 1957-73.* Ph.D. Thesis; University of London, Unpublished.
22. United Nations, 1967, *Methods of Estimating Basic Demographic Measures from Incomplete Data,* Manual IV, New York.
23. Visaria, P., 1971, Provisional population totals of the 1971 Census, *Economic and Political Weekly,* **VI,** 29.
24. Zachariah, K. C. and Cuca, R., 1972, *Population Projections for Bank Member Countries, 1970-2000.* International Bank for Reconstruction and Development.