R. Kumar

Use of Regression Models in Determining Accurate Vital Statistics in Developing Countries—A Methodology

Introduction

NON-AVAILABILITY of accurate information on vital statistics through normal registration source(s) is one of the main problems in determining accurate vital rates in all developing countries. The usual data collection source(s) suffer from a number of problems; under-registration being one of the most prominent. The problems have been aggrevated by several factors, namely, lack of awareness of the importance of this information among the masses, the data collectors themselves are often unable to read or write and the other imperfections in the data collection system itself.

Vital statistics data collection in India also suffers from similar problems. Estimates made by Mehrotra (1968 : 213) show that under-registration of births in different states of India varies from 15 to 85 per cent. A preliminary analysis carried out by the author (1972 : 9) on the basis of data collected in 29 villages in rural Punjab during 1969 to 1971 shows that *Chowkidars* (village watchman responsible for collecting vital statistics in rural Punjab) under-reported approximately eleven per cent of the events collected by "Primary sources"[1].

In demographic literature one comes across a number of techniques which have been developed to overcome this problem. The technique designed by C. Chandrasekaran and Edward W. Deming (1949) is based upon the theory of

1. Details about primary sources discussed on page 218.

probability, the application of which uses two different sources which are supposed to be independent. Ansley J. Coale (1961) also made use of probability theory in designing a procedure for estimating "accurate" numbers of vital events. Karol J. Krotki (1974) used population growth estimation as a method of obtaining death rates. Michael S. Teitelbaum (1974) used 'reverse projection' technique on census data in order to determine birth under-registration in the constituent countries of England and wales. G. K. Mehrotra (1968) calculated under-registration of births on the basis of survival rates. The office of (he Registrar General, India (1961), made use of 'difference method' in official estimates of birth and death rates during 1951-60. The method is based upon the age distribution obtained during 1951 and 1961 censuses. Technique proposed by J. R. Rele and U.P. Sinha (1973) use essentially the stable population theory. Assuming India's population to be quasi-stable, they applied the stable population theory to estimate birth and death rates. Thompson's Index for estimating vital rates is again based on stable population theory. The Thompson's index, sometimes also known as replacement index, is the ratio of children under five to women aged 15-44 divided by a similar ratio from the life table stationary population.

One of the main problems with the Chanderasekaran—Deming technique is that its application requires the use of two sources which are supposed to be independent. In addition, the underlying assumption of the two sources being independent is often difficult to verify. These two factors restrict the technique's practical utility. The Reverse projection technique proposed by Teitelbaum uses age-specific mortality rates, Krotki used population growth estimation as a method of obtaining death rates and Mehrotra used the survival rates to estimate under-registration of births. All these techniques use mortality rates which could be subject to the same problems as fertility data. Hence, in areas where mortality rates are not accurately reported, their application will not provide accurate estimates on fertility.

"With these considerations in mind, the feasibility of developing a technique which would use only one source and/or is not dependent on mortality data was investigated. In these attempts the relationship between data collected by the *Chowkidars* and the 'total' events recorded by all the sources was studied. The main reason for selecting *Chowkidars* in exploring this technique is the practical utility, as the *Chowkidars* are the only regular source of data collection in North India.

The Study

The data used in this analysis are taken from a field trial designed to test under rural conditions, the interactions between different combinations of an integrated MCH/FP programme and to measure their impact on fertility behaviour of the experimental populations. The study-design was a factorial experiment consisting of five treatment modalities, four receiving different combinations of maternal, child and family planning services and the fifth servicing as a pure control. To establish these causal relations with the highest degree of accuracy, a number of sources were employed to gather information on births taking place among experimental populations. For most of the experiment, the author had the overall responsibility for supervision of the data collected and its analysis.

Vital Statistics Data Collection Sources

(A) Primary Sources. Primary sources are those which were specifically designed to collect data on births, deaths and marriages taking place within the experimental villages. There were three sources, two started specifically by the Project employing Project personnel, whereas the third gathered information from vital statistics records generated by the village *Chowkidars.*

These sources were :

(i) VITAL STATISTICS INVESTIGATORS (vsis). One of the primary sources of data collection on vital events was through a team of investigators each covering 8-9 villages. In each village twelve 'key-informants' were selected. These included : A *Nai* (barber), all village *dais* (mid-wife), a teacher, an indigenous medical practitioner, a shopkeeper, a *panchayat* member (elected member to the village council), a *gramsevak* (community development village level worker) and some influential elders of a village-community. There informants were visited every fortnight by the VS Investigators, to minimize chance of error due to "recall-lapse", to be regularly questioned on vital events occurring in their respec tive village. These investigators were specifically instructed not to consult village *Chowkidars* and Family Health Workers.

(ii) FAMILY HEALTH WORKERS (phws). The experimental design provided fora resident worker in each of the action villages. These workers, involved in providing different combinations of maternal and child health (MCH) and Family Planning (FP) services, were instructed to maintain records on births and deaths occurring in their respective villages. The villages *dais*, who conducted almost all deliveries were the main informants for the FHWs and received monetory incentive when reporting any birth, death or still-birth. These workers were responsible for submitting these reports every month to the Project Headquarters.

(iii). VILLAGE CHOWKIDARS (CHOWS). *Village-Chowkidars,* a resident of a village community, is required by the government to report on live-births, still-births and deaths occurring in the village. It is assumed, being a member of the village community, he will hear of or be informed directly by the *dm* or the family affected by the event. The *Chowkidar* takes his report every fortnight to the police station under whose jurisdiction his village falls. These reports are then sent to the Chief Medical Officer of the district every month.

(B) Secondary Sources. The secondary sources were not primarily designed to collect data on vital statistics but some of the instances brought to the attention of Project personnel showed sufficient evidence of their utility in adding to the number of births and deaths already identified by primary sources in the experimental villages. For example, a young child being visited for morbidity and dietary surveys might not have a birth record, or the out-come of a pregnancy identified in the fertility-survey was unknown, or a young child who died did not have a birth record. These discrepencies created a need to look into the ways through which complete coverage of vital events could be assured. For this, certain indirect methods were devised, the sources of information being termed "secondary sources", to identify those events which were not recorded by all the primary sources. Again, there were three secondary sources, two based on surveys carried out for women and child care services and the third was derived from the death records generated by the primary sources.

(i) BIRTH PROSPECT CARD. The basic purpose of this card was for follow-up of the outcome of a conception which had taken place a few months earlier. Information about conception was available from a longitudinal fertility survey through which each eligible woman (a currently married women between 15-49 years of age) in a study village was contacted every two months by the resident worker. A birth prospect card was made as soon as a pregnancy was suspected or confirmed for a woman. After the expected date of delivery, usually nine months from last menstrual period, a search in vital statistics birth forms was made for the anticipated birth. When such a birth was not traced, it was usually referred to the original fertility form of the mother for possible explanation as to what happened to the termination. Depending upon the findings, the appropriate record was generated.

(ii) VITAL STATISTICS FORMS OF YOUNG CHILDREN Vital statistics death forms available for young children formed another source to improve the coverage of birth. Where the matching birth form was not available, a new birth form was generated for the dead child.

(iii) MORBIDITY SURVEY, In villages where morbidity data were available for child under *three* years of age, the birth dates available on the morbidity forms were used to check for birth reported by vital statistics investigators.

Comparing Events Recorded by Different Sources

Village *Chowkidars* were instructed to record all events taking place in the village whereas the VSIs and FHWs were asked to record all events occurring outside, as well as inside, a study village relating to the experimental-population, A practice widely prevalent in North India, particularly for the birth of first child, is for the daughter-in-law to return to her maternal home for delivery. Since, in this parr of the country, brides are always sought from another village, a proportion of women normally resident in one community must go to another (their mother's village) fur confinement. Similarly. confinements of returning daughters take place within the village. Hence, for this analysis, events were grouped into the following categories :

(i) *Local.* Those events which relate to the residents in the study village and take place in the village.
(ii) *Incoming.* Those events which take place in the study village but relate to the non-residents, such as daughter of the village returning for delivery,
(iii) *Outgoing.* Those events which occur outside the study village but relate to the experimental population, such as daughter-in-law of the village leaving to maternal home for delivery.

Regression Mode, for Estimating Number of Births

Collection of vital even-is through multiple sources provided an excellent opportunity to determine accuracy of different sources. This opportunity was also taken todevelopea technique which, when applied to an existing source of data collection *(Chowkidars)* would provide a reasonably accurate estimate of the number of births taking place in a particular region.

The Technique

The technique proposed here to arrive at 'accurate' estimates with respect to number of births, is based upon the theory of regression. It is well recognized that the regression models are limited *in* their scope for application to other regions owing to wide disparities in the extent of under reporting. It is also recognized that these regression models once developed would hold true for that particular time and, hence, their subsequent application may not result in accurate estimates if effectiveness of a source in reporting vital events changes significantly over a period of time.

In order to determine regression equation between the 'total' and the events recorded by *Chowkidars*, an 'experimental phase' is proposed. During this phase use of multiple sources to identify all possible events is suggested. If carefully designed and supervised, two sources may be sufficient to identify almost all events taking place in a village. During this phase, care needs to be taken with respect to the events collected by *Chowkidars*. Introduction of additional data collection system(s) should not alter the effectiveness of *Chowkidars* in reporting vital events. Failure to do so will not provide accurate estimates.

Needless to say, selection of communities for the 'experimental phase' should be representative of the area for which the regression equation is being developed. A one year period for the experimental-phase may be sufficient to obtain a reasonably accurate model. With respect to the number of villages (areas), at present nothing can be said except that the greater the number the better the estimates in general.

Assumptions

In the development of this technique the following two assumptions were made :

(i) All sources put together during the experimental-phase will be able to record all or most events occurring in the area. The additional effect of each source will go on decreasing as the number of sources increases, unless performance of a particular source is extremely poor.

$$Y - a_0 + a_1 x_1 + a_2 x_2 + \dots a_n x_n$$

where, $Y =$ total number of events identified by all the sources; $a_0 =$ common events registered by all the sources; $a_1, a_2 \dots a_n =$ multipliers associated with each source to determine the additional events added by them with the assumption that

$$a_1 > a_2 > a_3 \dots > a_n.$$

(ii) Application of the model rests on the assumption that effectiveness of *Chowkidars* in reporting vital events will not alter significantly over the following few years.

The Model

In an attempt to develop a regression model, events recorded by various sources were matched on certain selected variables in order to identify common events registered by all the primary sources, events registered in combination with another source and events exclusively recorded by a particular source. Events exclusively recorded by the secondary sources were also identified. Events recorded by different sources in combination with one another or exclusively were added to the common events in order to determine the "total" events. Equatienally this can be written as :

$$Y = C + N_{vf} + N_{vc} + N_{cf} + N_v + N_f + N_c + N_s*$$

where,

Y — Total number of events identified by all the sources;
C = Common events identified by all the sources;
N_{vf} = Events jointly identified by VSIs and FHWs;
N_{ve} = Events jointly identified by VSIs and CHOWS;
N_{tf} = Events jointly identified by CHOWS and FHWs;
N_v = Events identified only by VSIs;
Nf = Events identified only by FHWs;
N_e - Events identified only by CHOWS;
Nss= Events identified only by secondary sources.

The 'total' events y, identified by all the sources in a particular calendar month, were plotted on y-axis (considered as dependent variable) and those by C*howkidars*, for the corresponding months, were plotted on AT-axis (considered as independent variable).

In the attempts to identify the best fitted line between data collected by the *Chowkidars* and the total number of events identified by all the sources, a number of transformations were done. These transformations include : square, squareroot and log. A summary the correlation matrix between various transformations for different years is presented in Table 1.

It is realized that inclusion of additional sources in the development of Regression technique will slightly improve the estimates. But, as in North India, *Chowkidars* are the only regular source of collecting information on vital statistics, the technique designed to use additional sources will not be of much practical use. Hence, the technique developed is based upon the data collected by village *Chowkidars.*

With the decision not to use data collected by other sources in the development of the regression model, because of its possible use in other regions, the correlation matrix relates only to the data collected by *Chowkidars* and the total

TABLE 1-COEFFICIENT OF CORRELATION, WITH DIFFERENT TRANSFORMATIONS, BETWEEN "TOTAL" BIRTHS RECORDED BY ALL THE SOURCES AND BIRTHS IDENTIFIED BY VILLAGE *CHOWKIDARS*

Year	*Number of Births Recorded*	*Log (natural) Transformation*	*Square Root Transformation*	*Square Transformation*
1969	.96337	.95436	.96063	.96014
1970	.97964	.98225	.98114	.97578
1971	.99282	.99397	.99345	.99118
1969 + 70	.97009	.96584	.96932	.96643
1969+71	.97939	.97261	.97711	.97987
1970+71	.96688	.97218	.97015	.95750
1969+70+71	.96933	.96862	,97025	.96221

events identified by all the sources, with different transformations (Table I).

Table 1 reveals that for almost all years, individually or in combination, the coefficient of correlation was the highest with log transformation, though the difference between them is negligible. Of all the coefficients of correlations, highest (.99397) was found with log transformation for the year 1971. However, aggregate analysis for all the three years (1969-71) shows the coefficient of correlation to be highest with squareroot transformation (.97025). It is evident from the table that the difference in the number of events recorded by all the sources and the ones recorded by *Chowkidars* is the least for the year 1971.

It appears logical that the relationship observed between data collected by village *Chowkidars* and the 'total' number of births identified by all the sources during 1971 should provide the best estimates. This is based upon the assumption that if for a particular year *chowkidars* could identify events as closely to the'total'as during 1971, under similar conditions it would be possible for them to maintain the same level. If the same relationship was not observed, which is the case, then there must be something wrong either with the sources or *Chowkidars* or both. Whatever the reason may be, application of the regression model observed during 19yi should provide the best estimate owing to highest correlation coefficient.

In practice, it is possible to use the observed relationship between data collected by *Chowkidars* and the 'total' with any transformation, as the difference in the coefficient of correlation is negligible. But as the log transformations have the highest coefficient of correlation, hence, preferred.

Figure 1 shows scattergram depicting relationship between log (total) and log (*Chowkidars*) for 1971 which appears to be linear relationship.

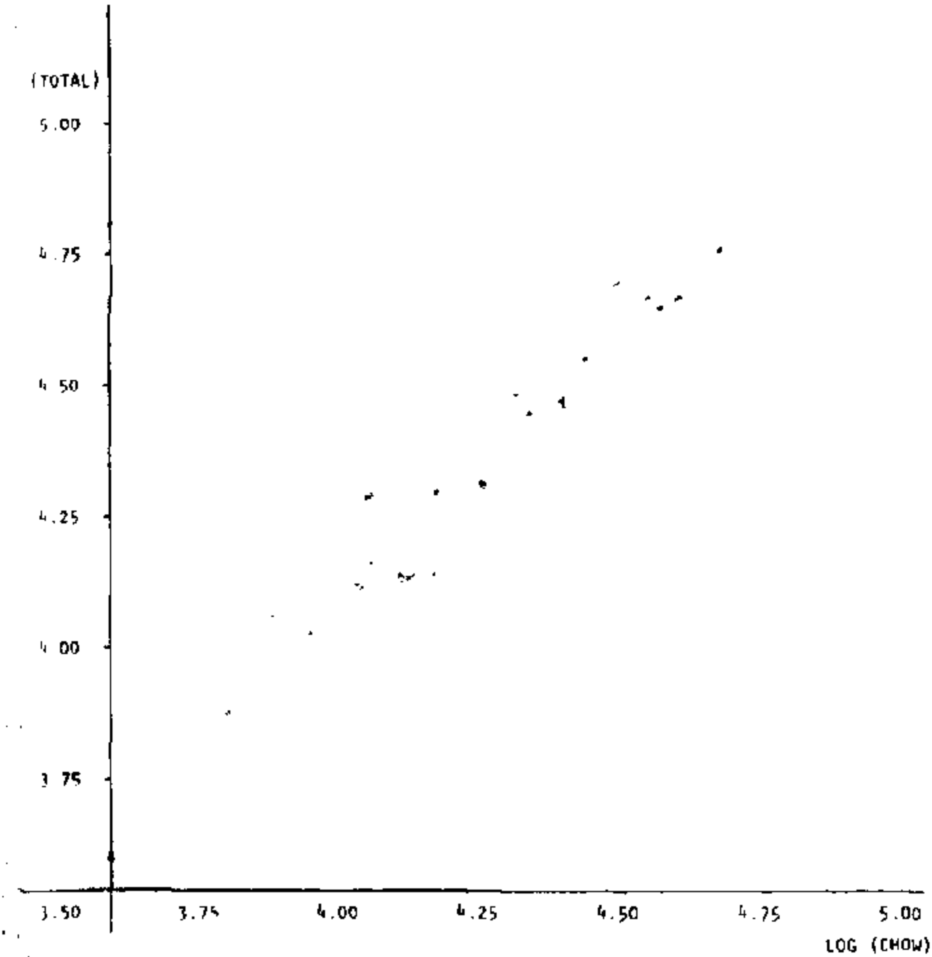

Fig. 1. Scattergram showing relationship between log (total) and log (chow) for the data collected during 1971.

The regression analysis gave a logrithmic least square equation of the form :

$$\log (Y) = \alpha + \beta \log (x).$$

Estimate of the total births $\hat{Y}$, was calculated by taking antilog of log (Y).

Table 2 shows the regression equations calculated on (he basis of log transformation for different years.

TABLE 2 -REGRESSION EQUATION FOR DIFFERENT YEARS ON THE BASIS OF LOG TRANSFORMATION

Year	Constant (a)	Slope (β)	Standard error on (β)	F-value
1969	.5349805	.9179271	.09084	102,119
1970	.3579332	.9734598	.05879	274.137
1971	.17322 4	.9930981	.03464	821.940
1969+70	.3730434	.9641114	.05515	305.614
1969+71	.4859693	.950184	1.04713	385.170
1970+71	.3628055	.9605237	.04934	378.960
1969+70+71	.4297990	.9445347	.04156	516.468

In using the regression equation one needs to keep in mind that this is calculated on data aggregated for one calendar month. In order to use this equa* ,tion to make estimates for a longer period, one needs to make a slight adjustment. For this the regression equation was modified as follows :

Y — antilog [log (Y)] = [log (number of agregated months)] + (3 log (X_i).

Multiplication of a with log of number of months for which estimates arc being aggregated appears to have no theoretical base but as it provided estimates very close to the observed 'total', this was adopted as the technique of estimation.

' For example to estimate the number of births for one year on the basis of regression model for 1971, the modified regression equation is :

$$Y = \text{antilog [log } (Y)\} - .1732214 \text{ [log (12)]} + .9930981 \log(x,)$$
$$= .1732214 \text{ [2.4849066]} + .9930981 \log (,r,)$$
$$= .430439 + .9930981 \log (X_1)$$

where, x_t is the number of births identified by *Chowkidars* during one year.

Suppose that in a particular year *Chowkidars* have identified 126 births.

The expected number of total births using the regression:

$$= .430439 + .9930981 \log (126)$$
$$= .430439 + .9930981(4.8362819)$$
$$= 5.2333414.$$

. Expected number of total events, $\hat{Y}$,

$$= \text{antilog}(5.2333414)$$
$$= 187.418$$
$$= 187 \text{ births (say)}.$$

Therefore, the estimated number of births by the regression model is 187 as compared to 126 identified by the *Chowkidars.*

Validity of a technique can only be ascertained if the estimates made by it compare favourably either with estimates made through another established technique or the data collected by source(s), the accuracy of which can not be questioned. In the absence of such a comparison it becomes difficult to judge the validity of the technique. Hence, to establish accuracy of the proposed model the estimates made by it were compared with 'total' number of births identified by all the sources and the ones estimated by the Chanderasekaran-Deming Technique (CD-technique). In estimating the number of births by CD-technique the data collected through two sources was used : (i) births recorded by *Chowkidars,* and (ii) births identified by Vital Statistics Investigators for corresponding months and villages (independence of both the sources was assured during the data collection phase). The number of births recorded by both the sources were matched on certain selected socio-demographic variables in order to ascertain births jointly and individually recorded by them.

Table 3 shows the number of births estimated by Regression technique (col.3), Chandrasekaran-Deming technique (col. 4) and compares them with the total number of births recorded by all the sources (col. 2). The table also shows number of births recorded by the *Chowkidars* (col. 1) for the same period and villages.

Comparison of Births Estimated by Regression Technique with Those Estimated by Chandrasekaran-Deming Technique and 'Total* Births Identified by All the Sources

Table 3 shows that for the calendar months and villages for which data was analysed, all sources, primary and secondary, identified 2847 births. For the same months and villages, *Chowkidars* registered 2454 (86.20 per cent of the

TABLE 3—NUMBER OF BIRTHS ESTIMATED BY REGRESSION AND CHANDER-SEKARAN-DEMING TECHNIQUES (ON THE BASIS OF DATA COLLECTED BY *CHOWKIDARS*, AND *CHOWKIDARS* AND VSIS RESPECTIVELY; AND THEIR COMPARISON WITH THE TOTAL NUMBER OF BIRTHS RECORDED BY ALL THE SOURCES

Number of births recorded (In a month} by		*Number of births estimated by*		*Difference*		
Chowkidars	*All the Sources*	*Regression technique*	*C-D technique*	*(Cot .2- Col. 3)*	*(Col. 2-Col. 4)*	*(Col. 3-Col. 4)*
1	*3*	*3*	*4*	*5*	*6*	*7*
40	50	46	47	+4	4-3	—1
42	44	49	43	—5	—1	+6
42	48	49	49	—1	—1	0
4*	50	56	50	—6	0	+6
49	56	57	55	—1	+1	+2
49	55	57	54	-2	+1	+3
50	57	58	57	—1	0	+1
50	62	58	58	+4	+4	0
52	60	60	55	0	+5	+5
52	56	60	54	-4	+2	+6
54	64	63	59	+1	+5	+4
60	77	69	70	+8	+7	—I
64	74	74	74	0	0	0
70	81	81	80	0	+1	+1
70	80	81	77	—1	+3	+4
70	80	81	77	-1	+3	+4
73	82	84	78	-2	+4	+6
73	88	84	80	+4	+8	+4
75	86	87	82	—1	+4	+5
77	89	89	83	0	+6	+6
80	95	92	88	+3	+7	+4
81	102	93	93	+9	+9	0
83	101	96	101	+5	0	—5
83	95	96	90	-1	+5	+6
86	102	99	98	+3	+4	+1
86	115	99	98	+ 16	+ 17	+1
89	97	101	95	—4	+2	+6
92	110	106	107	+4	+3	—1
93	109	107	104	+2	+5	+3
97	104	112	101	—8	+3	+ 11
99	107	114	106	-7	+1	+8
105	117	121	116	-4	+1	+5
108	122	124	120	—2	+2	+4
113	132	130	124	+2	+8	+6
2454 (86.20)	2847 (100.00)	2833 (99.51)	2737 (96.14)			

'total' births) births. When the proposed Regression technique was applied to the data collected by *Chowkidars* the number of estimated births works out to be 2833 (99.51 per cent of the 'total' births) births.

The analysis presented in Table 3 clearly reveals the extent of accuracy of the Regression technique. As compared to the 'total' number of births identified by all the sources the Regression technique has missed only 0.49 percent of the births whereas *Chowkidars* missed 13,80 percent. This clearly demonstrates the potentialities of the technique in estimating 'total' number of births based upon the data collected by village *Chowkidars.*

The results obtained through Regression technique were also compared with the ones obtained by Chandrasekaran-Deming technique. Col. 4 of Table 3 shows then umber of births estimated by C-D technique on the basis of data collected by village *Chowkidars* and VSls. As compared to the total events recorded by all the sources, estimated number of births by C-D technique is 2737 (96.14 per cent of the total events).

The analysis shows that, as compared to the C-D technique, better estimates were achieved through the Regression technique. As compared to the 'total* births, where the Regression technique has missed only 0.49 per cent births, C-D technique has missed 3.89 per cent. As compared to the C-D technique, Regression has estimated 3.40 per cent more births.

After having looked at the gross difference (ignoring the over (+) and under (—) reporting), the net difference was determined (taking into consideration the 'over' and 'under' reporting) in comparing accuracy of both the techniques. In evaluation of a technique such as this it is the net difference that appears to be more important.

Cols. 5, 6 and 7 of the Table 3 shows the difference (both 'over' and 'under' estimation) in the number of births estimated by the Regression technique and 'total', estimates made by both the techniques. The extent of this difference has been classified in Table 4 and Table 5.

Table 4 classifies the difference in the number of births identified by all the sources and the ones estimated by the Regression technique for the corresponding months and villages. Of the 34 comparisons, in four (11.77 per cent) instances no difference was found between events recorded by all the sources and ones estimated by the Regression technique. In 17 (50.00 per cent) of the cases, the number of births estimated by the Regression technique was higher than those recorded by different sources whereas in 13 (38.23 per cent) of the cases the situation was the reverse. In approximately 29 per cent of the cases the difference was up to two births only. As compared to the 'total' births identified by all the sources, Regression technique has 'over estimated' 51 births and.

TABLE 4--DIFFERENCE IN THE "TOTAL" NUMBER OF BIRTHS IDENTIFIED BY ALL THE SOURCES AND THE ONES ESTIMATED BY THE REGRESSION TECHNIQUE (TOTAL — REGRESSION)

Difference in the number of births	*Frequency*			*Cumulative total*	*Cumulative percentage*
	+	—	*Total*		
0	4		4	4	11.77
1	1	7	8	12	35.29
2	2	3	5	17	50.C0
3	2	0	2	19	55.89
4	4	3	7	26	76.47
5-9	3	4	7	33	97.06
10+	1	0	1	34	100.00

'Under' estimated by Regression technique = 65 births.
'Over' estimated by Regression Technique = 51 births.
Net difference = (Total -- Regression) = 65 — 51 = 14 births.
(As compared to the total, Regression Technique has under estimated only 14 births.)

TABLE 5—DIFFERENCE IN THE "TOTAL" NUMBER OF BIRTHS IDENTIFIED BY ALL THE SOURCES AND THE ONES ESTIMATED BY THE CHANDRA-SEKARAN-DEMING TECHNIQUE (TOTAL — C-D)

Difference in the number of births	*Frequency*			*Cumulative total*	*Cumulative percentage*
	+	—	*Total*		
0	4		4	4	11.77
1	6	1	7	11	32.35
2	3	0	3	14	41.18
3	5	0	5	19	55.88
4	4	0	4	23	67.64
5-9	10	0	10	33	97.06
10+	1	0	1	34	100.00

'Under' estimated by C-D Technique = 125 births.
'Over' estimated by C-D Technique = 1 birth.
Net difference = (Total -- C-D Technique) = 125 — 1 = 124 births.
(As compared to the total births identified by all the sources, Chandrasekaran-Deming technique has under estimated 124 births.)

TABLE 6—COMPARING NUMBER OF BIRTHS RECORDED BY *CHOWKIDARS*, IDENTIFIED BY ALL THE SOURCES AND ESTIMATED BY THE REGRESSION TECHNIQUE BY EXPERIMENTAL YEARS AND TREATMENT MODALITIES

Treatment Group	*1969*			1970			1971			*1972*			*1973*		
	Chow	*Total*	*Estimated*	*Chow*	*Total*	*Estimated*	*Chow*	*Total*	*Estimated*	*Chow*	*Total*	*Estimated*	*Chow*	*Total*	*Estimated*
FP + WS + CC	111	136	165	126	170	187	112	176	166	120	207	179	115	173	171
FP + WS	96	123	143	89	149	133	87	141	130	114	175	170	74	133	111
FP + CC	75	101	112	86	110	128	101	150	151	89	123	133	81	127	121
FP Edu.	—	—	—	—	—	—	—	—	—	92	109	137	116	144	173

'under estimated* 65 births, the net difference being 14 births. This pattern *of* "over" and 'under' estimation indicates these estimates are unbiased.

Table 5 shows classification of the difference in the number of births estimated by Chandrasekaran-Deming and Regression techniques. Again, in four instances there was complete agreement between the two estimates. Of the remaining comparisons, as compared to Chandrasekaran-Deming technique, Regression technique has 'over' estimated in 20 (86.67 per cent) of the instances. Only in four comparisons, as compared to the Regression, C-D technique has 'over* estimated the number of births. In approximately 41 per cent of the cases the difference was less than three births. As compared to the Chandrasekaran-Deming technique, Regression technique has 'over' estimated 118 births whereas as compared to the Regression, C-D technique has'overestimated only eight births. The net difference between the estimates made by two techniques is 110 births.

The observed Regression model was also applied to the data collected by village *Chowkidars* in different treatment modalities over the experimental period (Table 6). It is evident from the table that for 1969 and 1970 estimates made by the Regression techniques are higher than the observed total and the ones recorded by *Chowkidars.* For 1971-73 the observed total and the estimated births are comparable whereas there is a marked difference between the births recorded by the *Chowkidars* and the ones by other two sources.

Figure 2 compares the number of births estimated by the Regression and 'total* in relation to the births recorded by *Chowkidars.* As expected, the births estimated by the Regression are almost on a straight line whereas the observed total is scattered over.

Conclusions

The findings of this analysis suggest that the proposed Regression technique has great potential of becoming a possible method for estimating correct number of births in developing countries. The technique has estimated almost the same number of births as recorded by all the primary and secondary sources. It has also given better results than the ones estimated by Chandrasekaran-Deming technique. The Regression technique's distinctive advantages are ;

(i) Its application uses data collected through one source only,

(ii) Once developed, the regression equation can be used as long as effectiveness of *Chowkidars* in reporting vital events does not change markedly.

(iii) It is not dependent on any other type of demographic data.

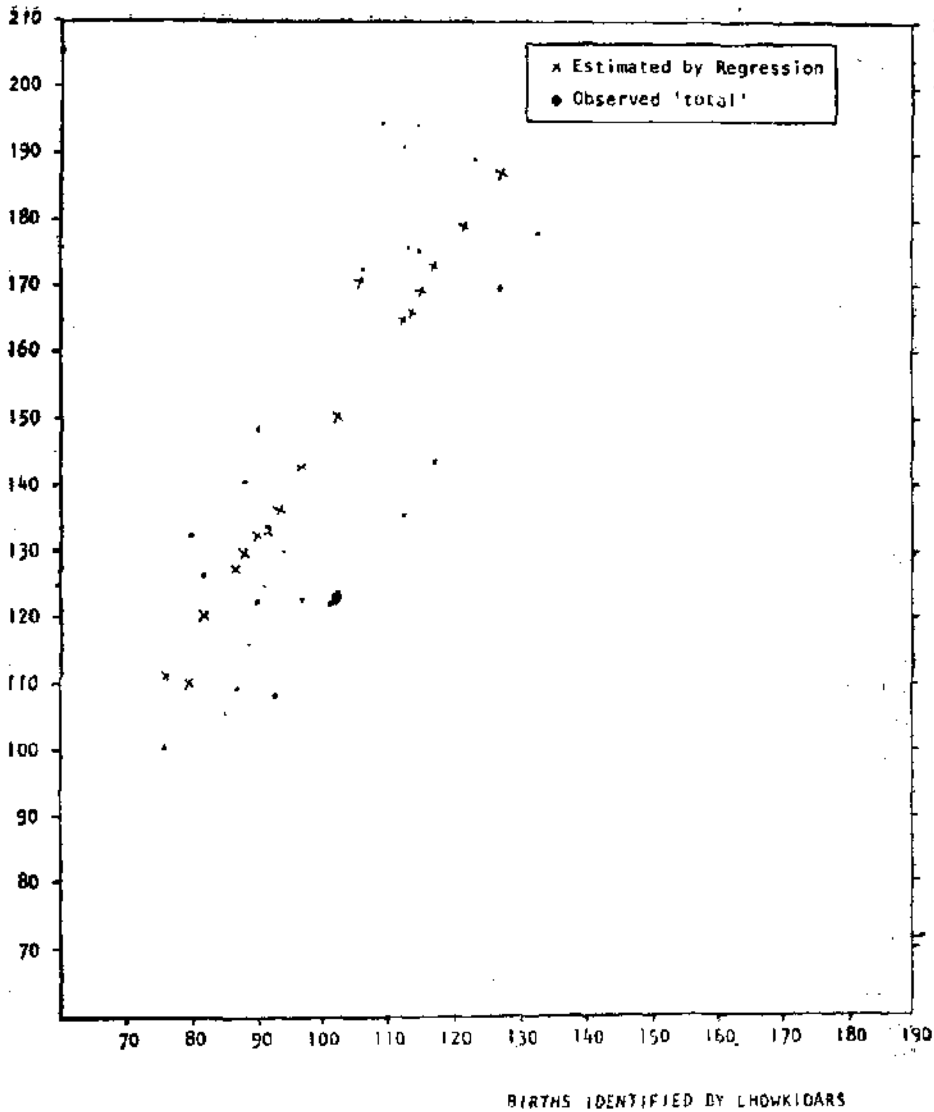

Fig. 2. Comparing number of births estimated by the Regression and 'total' in relation to the births recorded by *Chowkidars.*

(iv) Also, it does not make any assumptions either with respect to stable or quasi-stable population or the two sources being independent which sometimes becomes difficult to validate. This does not mean that the Regression technique is not without any assumptions, but those assumptions are comparatively easy to check, hence, less likely to affect the accuracy of the estimates. And,

(v) Its application is least expensive of all currently available techniques.

References

1. Chandrasekaran, C. and Edward, W. Deming, 1949, On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Society*, 88(249), 101-115.
2. Coate, Ansley J. 1961, Design of an experimental procedure for obtaining accurate vital statistics. *International Population Conference, New York, 1961*, London : Union international pour l'etude Scientifique de la population, pp. *372-375*.
3. **Krotki,** Karol J., 1974, Population growth estimation (PGE/ERADJ method of obtaining death **rates** : **An** overview. *Demography. India, III*(1), 172-183.
4. Kumar, R., 1972, Collection of vital statistics in rural Punjab. *NIHAE Bulletin*. 5(2), 9.
5. **Mehroira, G. K.,** 1968, Under registration of births in India. *Artha Vijnana*, 10(2), 213.
6. Rele. J. R., and Sinha, U. P., 1973, Fertility and mortality in India, 1951-60. In : Bose, A. et *al*. (eds.), *Studies in Demography*. Chapel Hill: The University of North Carolina Press.
7. Teitelbaum. Michael S., 1974, Birth under registration in constituents countries of England and Wales : 1841-1910. *Population Studies*, 28(2), 1974.
8. The Registrar General of India, *Vital Statistics of India for 1961*.