Madhuree A. Talwalkar

Matching of the Vital Events from Two Sources : An Experience in Vadhu Rural Health Project

Introduction

THERF is a growing concern about the availability of up to date and reliable vital statistics all over the country. The importance of accurate vital statistics for planning, implementation and evaluation of the health services cannot be overemphasised. Paradoxically this very information is lacking for rural areas where it is all the more needed.

Two types of methods are commonly used for collecting data on vital events—(1) Registration—continuous recording of the occurrence of vital events as and when they occur within a specific population (prospective type) and (2) Surveys—household surveys, either complete enumeration or on sample basis (mainly retrospective type). In rural areas, the system of Civil Registration of vital events, for which the Grampanchayat is responsible, relies mainly on the former method. With a view to get better estimates, dual recording system, the Sample Registration System (SRS) has been designed using both methods.

Both these methods, registration and survey, have advantages and limitations. In continuous recording, chances of recall lapse are minimised. In the survey there is an in-built system of getting population estimate simultaneously. Matching of data from both the sources can be useful to get a maximum by combining the two sources. Technique of matching has been used by several workers and results have been revealing. (Srinivasan and Muthaih., 1968; Marks *et al., 1978).* The present paper describes our experience in *the* Vadhu Rural Health project.

Vadhu Rural Health Project has been sponsored by the K. E. M. Hospital,

Putie in a group of 19 villages distributed in two Talukas of Pune district viz. Shirur and Haveli.[1] Besides the normal infrastructure, in each of these villages, there is a pair of Community Health Guides (CHG's), one male and one female. Of these, the male worker, maintains continuous record of vital events. For a group of villages, there exists a Multipurpose Worker (MPW) who maintains registers of vital events as a part of his routine health work. These registers *(R) of CHV/MPW is* one source of information. A complete enumeration survey was undertaken in the project area in June-July 1982, covering 5,741 households. Data on vital events which occurred in the year 1981 were collected in this survey.[2] The survey records *(S)* is the second source of information.

Objectives

When the estimates of vital rates of 1981 from the survey and from the Registers were compared, it was observed that the Crude Birth Rate *(CBR-S)* was higher than the *(CBR-R)* while the Crude Death Rate *(CDR-S)* was lower than the *CDR-R.* The main objective of the matching exercise earned *out in* 1984, was thus to examine the origin of discrepancies with a view to resolving them. Further, the focus, while resolving the discrepancies, was on assessing the quality of fertility and mortality data of the survey and thereby to arrive at the adjusted estimates of the survey rates. As made clear later, there are some methodological issues which given slightly different adjusted estimates for *R-* rates and 5-rates. This paper analyses different reasons of discrepancies, measures the extent of under-reporting in the survey in relation to the registers, and estimates the adjusted S-rates. It also describes the methodology evolved for matching and discusses different aspects of data verification. No attempt has however, been made to estimate the number of events missed by both the sources by using Chandrasekhar-Deming formula since the assumptions required are not satisfied.[3] Nevertheless, an attempt has been made at a later stage of the analysis to include the events missed by both sources but netted in a further enquiry.

1. The project area was initially made up of 19 villages. These villages were broken into 22 census villages in 1981 census. However, for the purpose of this study, number of units used is only 19.

2. The survey recorded births occurring among the normal residents irrespective of their place of occurrence. The definition of a 'normal resident' was a person staying or intending to stay in the village at least for six months. The registers included these events as well as births among visitors which occurred within the project area. However for comparison all births occurring *only* among the normal residents are considered. The rates based on these type of events are de jure Rates.

3. The assumption of independence of two sources of information **is not satisfied.**

Methodology

Analysis was done considering village as the unit. Methodology is described mainly in terms of 'births'. Same is followed for the deaths with corresponding modifications mentioned at relevant places. To start with, for every village all the births recorded in the survey were listed. This list is called L_1 for births. The particulars of births included identification number of the household from the survey, full name of the mother (with her husband's name and surname), sex of the baby and date of birth. In the case of deaths, the characteristics considered in L_1 were survey household number name of the deceased person, age at death, sex and date of death. Each of the births in L, was then compared with the entries in the registers and births were classified into the following types:

M : Completely matched i.e. information about all the characteristics considered was same in both the sources.

A : Partially matched i.e. the mother was located in the survey as the one who delivered a live baby in 1981 and some of her characteristics were same as those recorded in the registers.

B : Unmatched, i.e. entered only in the survey but not found in the registers.

C : Unmatched i.e. recorded in the registers but not in the survey. These **four** categories were mutually exclusive.

All births other than those in the category *'M'* required some verification. The next list L_2 was prepared for births of type *A, B* and **C.**[4] **To** facilitate identification of the concerned household in the field enquiry, additional columns such as name of the head of the household, hamlet i.e. wadi or wasti etc. were included in the L_2 for births and deaths. The verification was entrusted to the medical social workers specially trained for the purpose. The verification, of the entries in L_t was done mainly with the help of three types of methods.

Methods of verification:

(1) Recorded information from other sources.
(2) Home visits
(3) If both of the above two are not possible, information sought from the **CHG** or **MPW.**

Based on the findings of the verification, a decision was taken for each birth as to whether it should be included or excluded from the adjusted estimates of the survey births.

4. Actually, in addition to this, type *D* was also included in the list L_3. Type *D* is explained later in the section 2.1-

The criterion for the inclusion of events was designed from the survey point of view fulfilling the following three conditions : (I) the household should be present in the village at the time of survey, in 1982; (2) mother (in the case of birth)/person deceased (in the case of death) should be a normal resident (staying or intending to stay at least for six months) of the household; (3) the event should have taken place some time in 1981.

Some births could not be so verified and hence a decision could not be taken. Thus the following sub-groups were formed :

A(Is) : Part of *A* type of births to be *included in* the adjusted *survey births.*
A(Es); Part of *A* type of births to be *excluded* from the adjusted *survey* births.
A(U) : Part of *A* Type of births which could not be verified and hence remained in the 'undecided' category.

In the same manner, *B(Ig,),B(E3), B(U),* C(E3), *C(E8)* ann *C(U)* were defined. Thus the following identities were established :

$$\left.\begin{aligned} S &= M + A + B \\ R &= M + A + C \end{aligned}\right\} \quad \text{I}$$

$$\left.\begin{aligned} A &= A(I_S) + A(E_S) + A(U) \\ B &= B(I_S) + B(E_S) + B(U) \\ C &= C(I_S) + C(E_S) + C(CU) \end{aligned}\right\} \quad \text{II}$$

Findings and Discussions

For the survey, 14 villages were investigated by the CHGs and 5 by the MPWs. In the case of registers for 15 villages CHG registers were used for the matching and for 3 villages, MPW registers were used. For one village no register, either of the CHG or MPW, was available. In case of 11 villages, the investigator and the person maintaining the register was one and the same. To that extent these two sources cannot be considered as independant. Nevertheless, it was revealing that for some villages even though the same person did both the jobs there was a substantial percentage of unmatched events.

In what follows we first discuss matching of the data on 18 villages for which CHG/MPW registers were available. Here, we discuss the various decisions taken, problems that arose in doing so, methods used for verification, origins of discrepancies, problems of data collection, etc. With certain assumptions adjusted estimate for the survey, viz. S_1^* is worked out and relation between S_1^* and R_1^* (adjusted estimate of the register) is established. Thereafter all the 19 villages of the area are considered. Final estimates S_2^* for the number of events in the survey are worked out including the events missed by both the sources. The rates are calculated after making necessary adjustments in the population. Thus, these rates refer to the total project area.

1. *Analysis of Matching (for 18 villages)*

I.I *Distribution of events by 'matched' status*

Table 1 presents number of events from both the sources, according to their 'matched' status. As already mentioned, total number of (S) births is higher than the *(R)* births while total number of (S) deaths is lower than the *(R)* deaths,

TABLE 1—NUMBER **OF** EVENTS **BY** 'MATCHED' STATUS

Type of Event	*Births*	*Deaths*
Completely Matched (M)	496	126
Partially Matched *(A)*	69	8
Not Matched -		
1. Entered only in Survey *(B)*	348	27
1. Entered only in Registers (C)	203	110
Total for survey (S) (M + *A* + *B)*	913	161
Total for registers (R) *(M + A + C)*	768	244

Obviously the source recording lesser number of events had lower percentage of unmatched events if we accept that it records mainly the easily identifiable events. In other words, since the survey had netted more number of births, 38.1 percent of these births were not entered in the registers. Similarly, for the registers, having collected more deaths as compared to the survey, 45.1 per cent of its deaths were not recorded in the survey. The number of events requiring verification, was *(A + B + C).* Among births 620 needed verification as against 496 completely matched births. For deaths the verification was needed for 145 cases vis-a-vis 126 completely matched deaths.

1.2 *Distribution of events by 'decision' taken*

On the basis of the verification, it was decided to include certain number of events from each category under the adjusted survey figures. Table 2 gives the number of events—(1) To be included *(Is),* (2) To be excluded *(Es)* and (3) Undecided (U).

As expected, the decision to include related to almost 90 per cent of the partially matched births and cent per cent of partially matched deaths, with corrections in the partially matched characteristics. From the births entered

TABLE **2—VERIFIED** EVENTS ACCORDING TO 'DECISION' TAKEN

Type of Decision	*Births*			*Deaths*		
	Is	*Es*	*U*	*Is*	*Es*	*V*
Partially Matched *(A)*	**62**	*4*	**3**	**8**		
	(89.9)	**(5.8)**	**(4.3)**	**(100.0)**	—	—
Entered in the Survey but	**248**	**49**	**51**	**21**	**4**	*2*
not in the register (5)	**(71.3)**	**(14.1)**	**(14.6)**	**(77.8)**	**(14.8)**	**(7.4)**
Entered in the Register but	**114**	**55**	**34**	**78**	**18**	**14**
not in the Survey *(C)*	**(56.2)**	**(27.1)**	**(16.7)**	**(70.9)**	**(16.4)**	**(12.7)**
Total	**424**	**108**	**88**	**107**	**22**	**16**
	(68.4)	**(17.4)**	**(14.2)**	**(73.8)**	**(15.2)**	**(11.0)**

(Figures in the parenthesis indicate percentages adding to 100 for each type of events horizontally.)

in the survey but not in the registers, the inclusion amounted to 71.3 per cent; the corresponding percentage for deaths was 77.8 per cent. Out of births entered in the registers but not in the survey, 56.2 per cent should have been included under the survey; the corresponding figure for deaths was 70.9 per cent.

1.3 *Events for which 'decision' could not be taken*

In the case of 14.2 per cent of births *(A + B +* C) and 11 per cent of deaths it was not possible to take any decision. These cases, along with the reason for failure to decide, are presented in Table 3. For more than 40 per cent of these events where decision could not be taken, it was not possible to contact the person for various reasons even with repeated visits. In the case of 17 per cent of births and of 1 in 16 deaths, the households concerned had already left the village. In the case of one event out of every eight, it was not possible to locate the particular member. Moreover, for 21 cases of birth and 6 cases of death, the members concerned could not give any reliable information. This table thus highlights the difficulties of data-collection arising from limitations of respondents and the difficult field conditions.

1.4. *Methods of verification*

For all the other cases the decision to include or to exclude was taken on the basis of different methods as shown in Table 4. For a small number of cases, decisions could be taken even without going to the field. For example having prepared the *L,,* it was noticed that a duplicate household form was filled in for one case or there was a mistake in the coding of the year of birth, etc. It may be mentioned that the next round of data-collection for the survey

TABLE 3—CLASSIFICATION OF CASES BY REASON FOR NO DECISION

Reason	*Births*				*Deaths*			
	A	*B*	*C*	*All*	*A*	*B*	*C*	*All*
Left the village	—	13	2	15 (17.0)	—	1	—	1 (6.3)
Could not give reliable information	—	12	9	21 (23.0)	—	—	6	6 (37.5)
Could not be located	—	2	9	11 (12.5)	—	1	1	2 (12.5)
Could not be contacted—								
(*a*) Not accessible	—	2	3	5	—	—	1	1
(*b*) House locked	3	10	1	14	—	—	3	3
(*c*) No member at Home	—	12	10	22	—	—	3	3
(*a*) + (*b*) + (*c*)	3	24	14	41 (46.6)	—	—	7	7 (43.7)
All	3	51	34	88 (100.0)	—	2	14	16 (100.0)

(Figures in the parentheses indicate percentage).

was done in 1983 and on the same household forms updated information was recorded in a different ink. Some of the events missed in the first round in 1982, were recorded in this survey of 1983. Some events of 1981 were thus readily available and could be verified from this source. Further, sometimes other records were also utilised such as registers of other CHGs or MPWs, Ante-natal cases (ANC) or death records in the case of birth verification. For death verification, birth records were useful to some extent, particularly in the case of infant deaths. At times, there was a confusion between a still birth and infant death. A special mention may be made that the date of reporting in all these records, especially in the ANC records, was very useful in taking decisions. In a large number of cases actual household was visited. Some decisions were based on the information supplied by the CHG or MPW personally.

In the case of 56.2 per cent births, verification was done by actual home visits. Such a percentage was 73.6 for deaths. Other records were useful in 21 per cent of births but only for 7 per cent of deaths. Confirmation by CHG/MPW was helpful in almost equal per cent of both the events viz. 19 per cent for births and 17.8 per cent for deaths. This shows the extent of efforts involved in making about 400 homes visits scattered in 18 villages, not to mention the unsuccessful and repeated visited.

TABLE 4—METHODS USED FOR VERIFICATION

Type	At L_1 Stage	Survey 1983	Other Records			Home Visits	Confirmation		
			CHV/ MPW	ANC	Deaths		CHV	MPW	Total
					Births				
$A(I_s)$	—	—	—	4	—	50	1	7	62
$A(E_s)$	—	—	—	—	—	3	1	—	4
$B(I_s)$	—	—	5	64	—	143	35	1	248
$B(E_s)$	3	2	—	3	—	35	6	—	49
$C(I_s)$	4	10	1	30	2	47	20	—	114
$C(E_s)$	—	—	1	3	—	21	25	5	55
Total	7	12	7	104	2	299	88	13	532
Percentage	(1.3)	(2.2)	(1.3)	(19.6)	(0.4)	(56.2)	(15.6)	(2.4)	(100.0)

Type	At L_1 Stage	Survey 1983	Other Records		Home Visits	Confirmation		
			CHV/MPW	Births		CHV	MPW	Total
				Deaths				
$A(I_s)$	—	—	3	—	4	1	—	8
$A(E_s)$	—	—	—	—	—	—	—	—
$B(I_s)$	—	—	1	—	16	4	—	21
$B(E_s)$	1	—	—	1	2	—	—	4
$C(I_s)$	—	1	—	4	67	6	—	78
$C(E_s)$	—	—	—	—	6	11	1	18
Total	1	1	4	5	95	22	1	129
Percentage	(0.8)	(0.8)	(3.1)	(3.9)	(73.6)	(17.0)	(0.8)	(100.0)

1.5. *Adjusted estimate of the survey events (for 18 villages)llages)*

Having accomplished the verification, the adjusted estimated S_1* for the survey was worked out. **In** doing this, the events of the 'U' category in each of the *A, B, C* types were assumed to have the same proportions of 'to be included' and 'to be excluded events' as of the events where decision could be taken.

In other words $S_1^* = M + A^1(I_s) + B^1(I_s) + C^1(I_s)$

where $A^1(I_s) = A(I_s) + \frac{A(I_s)}{A(I_s) + A(Es)} A(U)$

$$B^1(I_s) = B(I_s) + \frac{B(I_s)}{B(I_s) + B(E_s)} B(U)$$

$$C^1(I_s) = C(I_s) + \frac{C(I_s)}{C(I_s) + C(E_s)} C(U)$$

On the basis of this figure, extent of under-reporting of the survey was worked out as presented in Table 5.

TABLE 5—ADJUSTED ESTIMATES OF NUMBER OF EVENTS IN SURVEY
(For 18 Villages)

Type of Event	*Births*	*Deaths*
Number recorded in Survey (S)	913	161
Number recorded in Registers (R)	768	244
Adjusted Estimates (S_1^*)	989	246
Percent underreporting of survey events in relation to the register	7.7	34.6

Figures of under-reporting of the survey in relation to registers highlights that the coverage of fertility data in the survey was satisfactory. Moreover, having established good coverage, 88 per cent of survey births $M+ A$ (Is) + B$(I_s)/S$ were verified to be correct entries. This reflects a reasonable quality of the fertility data in survey. The coverage of mortality data was, however, far from satisfactory with more than 34 per cent of under-reporting. The fact that as high as 96 per cent of survey deaths were correct entries is merely a consequence of low coverage rather than a reflection on quality.

1.6. *Origin of discrepancies*

Analysis of verified events further revealed different reasons for discrepancies. They are grouped into categories shown in Table 6. A cursory look at this table reveals that 63 per cent of birth discrepancies were due to the coverage error either of the survey or of the register. This percentage was as high as 76.7 for verified deaths. It is obvious that B(Is) cases i.e. 225 for births and 21 for deaths were the cases of coverage of the register and C(Is) cases i.e. 110 for births and 78 for deaths were the cases of coverage error of the survey. Slightly more than 10 per cent of births and 7 per cent of deaths were incorrectly included as having occurred in 1981 either by the survey or by the

TABLE 6A—ORIGIN OF DISCREPANCIES

Type	*Problem due to Metho-dology*	*Coverage Error*	*Error of Under-standing*	*Reference Period Error*	*Reference Area Error*	*Reporting Error*	*Other Errors*	*Total*
				Births				
$A(I_s)$	—	—	—	--	—	62	—	62
$A(E_s)$	—	—	—	3	1	—	—	4 / 66
$B(I_s)$	20	225	1	—	—	2	—	248
$B(E_s)$	—	—	3	43	1	—	2	49 / 297
$C(I_s)$	—	110	—	—	—	—	4	114
$C(E_s)$	24	—	18	11	—	2	—	55 / 169
Total	44	335	22	57	2	66	6	532
Percentage	(8.3)	(63.0)	(4.1)	(10.7)	(0.4)	(12.4)	(1.1)	(100.0)

TABLE 6B—ORIGIN OF DISCREPANCIES

Type	*Problem due to Metho-dology*	*Coverage Error*	*Error of Under-standing*	*Reference Period Error*	*Referenee Area Error*	*Reporting Error*	*Other Errors*	*Total*
				Deaths				
$A(I_s)$	—	—	—	—	—	8	—	8
$A(E_s)$	—	—	—	—	—	--	—	— / 8
$B(I_s)$	—	21	--	—	—	—	—	21
$B(E_s)$	—	—	—	3	—	—	1	4 / 25
$C(I_s)$	—	78	—	—	—	—	—	78
$C(E_s)$	6	—	5	6	—	1	—	18 / 96
Total	6	99	5	9	—	9	1	129
Percentage	(4.6)	(76.7)	(3.9)	(7.0)		(7.0)	(0.8)	(100.0)

registers. Among these there were 3 births recorded by both the sources for reference period but the home visits revealed the date of birth to be outside it. In the case of a couple of births, households staying near the border of the project area were incorrectly included. It is interesting to know that some of the discrepancies in survey and register arose due to different reports by the respondent. For example, for the same person two different names or surnames or both were entered in the two sources. Obviously most of these were categorised under partially matched events. The extent of this error was more for births than for deaths, the two being 12.4 and 7 per cent respectively. Such occasions related mainly to the change of the name of woman after marriage. At times, the discrepancy arose from the reporting of maiden names. In some cases, there was a problem in the understanding of the investigator of the survey or the person maintaining the register. For example, some **CHGs** were under the impression that they had to enter only the events occurring in their village. Consequently, they missed the events occurring among the normal residents of the village at places outside their own village. For some cases, to avoid the duplication of recording, some CHG's did not enter events occurring to normal residents of their village, somewhere within the project area but outside their own village. The survey investigator for few cases included events occurring to non-residents. The extent of error due to lack of understanding was about 4 per cent each for both births and deaths. About one per cent of discrepancies were due to other errors such as in coding or double counting etc. Incidently, such a low figure attests to the good quality of survey data. These exceptional cases also indicated, the areas calling for more attention.

1.7. *Relation between adjusted estimates of events between survey and registers*

Some of the cases not recorded by a source did not fall within the scope of that source due to the limitation of the methodology. In other words, there were differences between the survey and registration in the criteria of inclusion/exclusion. For example, the survey recorded 1981 events occurring to the normal residents of the village in 1982. The survey investigator could contact only those households which were present in the village at the time of the survey i.e. in 1982. He then recorded events of 1981 in these households. Thus households which were normal residents of the village in 1981 but left the village sometime before the survey were not contacted by the investigators. The events of 1981 in these households not falling in the purview of the survey, did, however, form a part of the entries in the register. Similarly, households inmigrating to the villages during 1982 were included in the survey for recording events of 1981 though they were outside the scope of the registers. The discrepancies arising out of such methodological problem was 8.3 per cent for births and 4.6 per cent for deaths. This portion of the discrepancies is responsible for giving different adjusted estimates for the survey and for the register. For further discussion see the Appendix.

As observed from Table 6 such a difference for births is only in 4 cases (24 — 20). The difference for deaths is for six cases. It may be added that all these 6 cases were of single-member households existing in 1981, and recorded in the register. These persons died by end of 1981 and could not be netted by the survey since these households did not exist at the time of survey. Thus in relation to S_1*, the extent of difference between S_1* and R_1* was only 0.4 per cent for births and 2.4 per cent for deaths.

1.8. Coverage error

It was important to know various reasons why a substantially large number of events was categorised as coverage error.

(a) SURVEY COVERAGE ERROR (BIRTHS—CIs). About errors in survey coverage, there were 110 cases of births as shown in Table 6A. Out of these, for 55 cases, the survey had *included* the *household i.e.* the survey form was *fitted in but the* birth was not recorded while for the other 55 cases the survey missed the households completely. Further analysis of the 55 cases for which household form was available revealed 12 cases where births of 1981 did not survive upto the time of survey (10 dying in 1981 and 2 dying in 1982 before the survey). Thus about 21.8 per cent of missed births were the cases of children not surviving at the time of survey. For the other 29 cases, birth was recorded in the survey but outside the year 1981. The remaining 14 cases were clear-cut error of the data collection, out of which 7 were males and 7 were females. When the location of the other 55 cases for which the complete household was missed from the survey was studied an interesting finding emerged. Almost for every one household missed from the main place of the village i.e. Gavthan, there were 4 households missed from the different hamlets. Percentage of all women giving birth in 1981 and staying in the hamlets was observed to be 57.8 from an earlier study for the project area (Talwalkar M. A., 1984). The number of hamlet births among missed births was found significantly higher.[5] In other words, greater number of hamlet households were missed by the survey.

(b) SURVEY COVERAGE ERROR (DEATHS—*Cls).* There were 78 cases of coverage error of survey deaths. Out of these, the survey form was rilled in 46 cases, but the death was not recorded, while in 32 cases, the complete household was missed. In the missed deaths, 20 were males and 26 were females. Percentage of female deaths among all deaths S_1* was 49.1 when compared to 56.5 per cent of missed deaths. However, sex difference of missed deaths was not statistically significant.

Of the completely omitted households 65.5 per cent cases were from the hamlets. No conclusion could, however, be drawn due to non-availability of the location-wise break-up either of the population or of all deaths in 1981.

5. Statistical significance is tested with the X" test using five per cent level of significance.

(c) REGISTRATION COVERAGE ERROR (BIRTHS—*BIs).* The 225 births undercovered by the registers were classified according to their location. 64.9 per cent of these were from hamlets. This is significantly higher as compared to the percentage of hamlet women giving birth in 1981 for the project area, viz. 57.8 quoted earlier. Thus larger number of hamlet births were missed by the registers. No significant sexwise difference was observed among these 225 missed births, the break-up being 117 males and 108 females.

(d) REGISTRATION COVERAGE ERROR (DRATHS—*BIs).* Among the 21 deaths undercovered by the registers 11 were from hamlets. As already mentioned, for the lack of comparable figure for the total project area, it was not possible to draw any conclusion. The sex-wise breakup of these 21 deaths was 9 males and 12 females, not statistically significant as compared to the sex-wise distribution of all deaths in 1981 in the project area.

To sum up the analysis of coverage error, no sexwise omissions were observed for any of the events in either of the sources. Nevertheless, greater number of hamlet households were missed by the survey and larger number of hamlet births were omitted from the registers. Thus in all at least 87 households were missed by the survey, assuming not more than one event occurred in one household.[6] This amounts to 1.5 per cent of the number of households actually covered under the survey. In other words, for every 200 households covered under the survey, at least three households were missed. This is obviously a lower limit, since there would be some more households missed by the survey which were not identified in this verification.

2. *Final Estimates (For all 19 Villages)*

2.1 *Type 'D' events*

.It may be emphasised that the results mentioned so far assume recording of the event in at least one source, survey or register. Nevertheless, there is a possibility of a certain number of the events missed by both the sources. We call it as type 'D' events. For the two independent sources of data collection, the number of such events can be estimated by using Chandrasekharan-Demmg method with some assumptions. However, as noted earlier, this method cannot be applied in this case. The estimates adjusted so far are thus the lower limits of the real values. We now attempt to calculate the final estimate of number of births and deaths for the total area by using some additional information for all 19 villages of the project area.

6. The most probable case of overlapping was due to infant death for which both birth and death occurred during the year 1981. All such cases, for which survey form was not filled in were checked and it was confirmed that there was no overlapping. However, cases of birth of one person and death of another person in the same household during the same year cannot be ruled out.

For the villages where no register was available, all the events recorded in the survey were physically verified in the field and then the decision for inclusion was taken. Moreover, as already mentioned the second round of the survey in 1983 had netted some more events. They were of types : (i) events from the households missed earlier; (ii) missed events from the households identified earlier. Those among these that did not match with the entries in the registers were also verified in the field. Some births missed by both the sources were identified while matching the infant deaths. Similarly some deaths of children missed by both the sources were located in the process of matching births. All these additional events that were classified as D were verified in the field. Then the final estimate is arrived at by the relation.

$$S^*2 = S^*1 + D(I_s)$$

Percentage of under-reporting of the survey is then worked out in relation to this revised figure. It is 9.1 per cent for births and 33 per cent for deaths as shown in Table 7.

TABLE 7—FINAL ESTIMATES OF NUMBER OF EVENTS AND RATES IN SURVEY : (For 19 Villages)

Type of Event	*Births*	*Deaths*
Numbers recorded in the survey	974	179
Adjusted Estimates S^*_2	1072	267
Percent under-reporting of the survey events	9.1	33.0
Adjusted (De jure) Survey Rates	32.4	8.1
Percent under-reporting of the survey rates	6.4	30.9

2.2 *Population adjustment*

The vital rates are also worked *on* using the final figure of events and adjusted estimate of 1981 midyear population. From the survey, the population count for midyear 1982 was available. To this, the balance of population arising due to inclusion of *C* and *D* type cases where the survey schedule was not filled in and exclusion of *'A'* and 'B' type where survey schedules were filled in, was added. For C Type of cases where the decision of exclusion was taken only because the event had not occurred in the reference period, the household was identified as the resident household. In such cases, if the household was found as omitted from the survey, the population for the household was added.

While counting the population, wherever the actual household size was not available, average household size of 5.7, as observed in the survey for the pro-

ject area, was used. Assuming one event per household, number of events are multiplied with the average household size. The population of households which are missed in the earlier round but were identified in the second round of the survey and which did not record any 1981 event, was also added. From this the midyear 1981 population was estimated by using the 1971-81 inter-census growth rate for the project area.

Conclusions

The study of matching, while giving the extent of under-reporting in the survey also revealed several problems of data-collection and data verification in a rural set-up. An important conclusion was that the fertility data collected through surveys is better than the mortality data. This is mainly because of the difference between the nature of the two types of events—a birth and a death. A birth is comparatively readily identified firstly because of the existence of the new born. Secondly, it is not a sudden and unexpected event, ab it is the end result of the pregnancy experience of the mother. There is no such obvious physical evidence for the event of death which would facilitate its reporting. There does not exist any specific type of person in the household necessarily associated with the event of death. Thus a death leading to the break-up or non-existence of the household may not be reported. In that sense it may not be considered as a household event. Thirdly, there is a human tendency to remember happy events and forget unhappy ones. Thus death-reporting in the household surveys, particularly in the areas where there is no appreciation of the importance of vital events data is likely to give unsatisfactory results. Continuous registration is the better method for collection of death information and has to be reinforced.

This comparison of two types of source has revealed some of the misconceptions of the persons maintaining registers and has indicated the areas that need to be strengthened for improvement of the future registration system. Finally, the study has brought out an interesting relation between the adjusted estimate of the survey and of the registers which needs to be considered while comparing the results from the two methods.

Acknowledgement

This study received encouragement from the authorities of K. E. M. Hospital Pune. I must acknowledge suggestions and comments of Prof. K. S. Srikantan, Chief, Population Research Centre, Gokhale Institute of Politics and Economics, Pune. My sincere thanks are also due to the Research Staff associated with the collection, verification and analysis of data.

References

1. **Chandrasekharan, C. and Deming, W. E., 1949, On a method of estimating birth and** death rates and the extent of registration, *Journal of the American Statistical Association,* 44(245), March, 101-115.
2. Marks *et al.,* 1978, po*pulation Growth Estimation,* Population Council, New York.
3. Srikantan, K. S. and Vaijayanti, Bhate, 1982, *National Fertility and Mortality Survey, Maharashtra 1980,* Population Research Centre, Gokhale Institute of Politics and Economics, Pune, pp. 8.1—8.2.
4. **Srinivasan K. and Muthaih, A., 1968, Problems of matching of births identified from** two independent **sources,** *The Journal of Family Welfare, 14{4),* ***13-22.***
5. Talwalkar, M. A., 1984, *Study of Maternal and Child Health Care,* K. E. M. Hospital, Pune, Research Centre (Mimeographed).

To explain the difference between two adjusted estimates S_1^* and R_1^*, without loss of generality we assume $A(U) = B(U) = C(U) = 0$, then $A^1(I_s) = A(I_s)$, $B^1(I_s) = B(I_s)$, and $C^1(I_s) = C(I_s)$. This simplifies the argument and facilitates clarifying the basic issue.
Thus we have

$$\left.\begin{aligned} S_1^* &= M + A(I_s) + B(I_s) + C(I_s) \\ \text{and } R_1^* &= M + A(I_r) + B(I_r) + C(I_r) \end{aligned}\right\} \quad \text{I}$$

where I_s *indicates decision of inclusion from survey point of view and* I_r indicates decision of inclusion from the registers point of view.

We consider A type of events. Since for this type of events, some recording in both the sources is essential, the discrepancy due to problem of methodology does not arise in it. In other words $A(I_s) = A(I_r)$.

The problem due to methodology can arise only in the categories $B(I_s)$ and $C(E_s)$.

Let us define

$$\left.\begin{aligned} B(I_s) &= B(I_s)^X + B(I_s)^Y \\ C(I_s) &= C(I_s)^X + C(I_s)^Y \\ C(E_s) &= C(E_s)^X + C(E_s)^Y \end{aligned}\right\} \quad \text{II}$$

where 'X' part is the number classified as 'Due to problem of methodology' and 'Y' part is the remaining portion in each category. (Refer to table 6).

Then

$$\left.\begin{aligned} C(I_r) &= C(E_s)^X + C(I_s)^Y \\ \text{and } B(I_r) &= B(I_s) - B(I_s)^X \\ &= B(I_s)^Y \end{aligned}\right\} \quad \text{III}$$

since the decision criteria for inclusion and exclusion are reversed in the cases arising 'due to problem of methodology'.
Therefore

$$\begin{aligned} S_1^* - R_1^* &= B(I_s) + C(I_s) - B(I_r) - C(I_r) && \ldots \text{from I} \\ &= B(I_s)^X + B(I_s)^Y + C(I_s) - B(I_s)^Y - C(E_s)^X - C(I_s)^Y \\ & && \ldots \text{from II \& III.} \\ &= B(I_s)^X + C(I_s) - C(I_s)^Y - C(E_s)^X \\ &= B(I_s)^X + C(E_s)^X + C(I_s)^X \end{aligned}$$

Now, events recorded only in the register i.e. 'C' Type, cannot be included in the survey if they arise due to methodological problems. Thus $C(Is)x$ will always be zero.

Therefore

$$S_1^* - R_1^* = B(I_s)^x - C(E_s)^x$$

= (Number of 1981 events occurring in the households inmigrating in 1982) (Number of 1981 events occurring in the households outmigrating or not existing in 1982).