

## **Experience of India SRS with the use of record linkage studies to assess the completeness of the reporting of vital events**

Kameshwar Ohja (Office of the Registrar General & Census Commissioner, VS (SRS) Division, Min. of Home Affairs, New Delhi, India)

Key points for paper on assessment of completeness of data in the Indian SRS

**Question 1:** Brief Overview of sampling approach in SRS- sample size, sampling design and distribution, reselection procedures of enumeration blocks after each census etc.

### **Response**

#### **Sample size**

Infant Mortality is the decisive indicator for estimation of sample size. The permissible level of error has been taken as within 15 prse (percentage relative standard error) at Natural Division level in respect of major States having population more than 10 million as per Census 2011. For minor States, 15 prse has been fixed at the total State level. By and large the above criteria have been followed. However, there have been a few exceptions on account of operational constraints. Half Yearly survey, 2014 was conducted in 8853 units. The decrease of few units from baseline survey can be attributed to displacement of population, administrative constraints, unviable population like migrants etc. Statement 2 shows the number of sample units and population covered in 2014, separately for rural and urban areas of all the States and Union Territories.

#### **Sample design**

The Sample design adopted for SRS is a uni-stage stratified simple random sample without replacement except in stratum II (larger villages) of rural areas, where two stage stratification has been applied. In rural areas of bigger States (population with ten million or more as per Census 2011), the NSS natural division is the first level of geographical stratification. The overall stratification in rural areas has been done on size of villages with villages having population less than 2,000 forming Stratum I and villages with population 2,000 or more forming Stratum II. Smaller villages with population less than 200 were excluded from the sampling frame in such a manner that the total population of villages so excluded did not exceed 2 per cent of the total population of the State. The number of sample villages in each State was allocated to the substrata proportionally to their size (population). The villages within each size stratum were ordered by the female literacy rate based on the Census 2011 data, and three equal size substrata were established. The sample villages within each substratum were selected at random with equal probability. In the case of villages of Stratum II, each sample village with a population of 2,000 or more was subdivided into two or more segments in a way that none of the segments cut across the Census Enumeration Blocks (CEBs) and the population of each segment formed by

grouping the contiguous CEBs was approximately equal and did not exceed 2000. A frame of segments was then prepared and the selection of segments was done at random at the second sampling stage for the SRS enumeration.

In urban areas, the categories of towns/cities have been divided into four strata based on the size classes in contrast to the six strata in the earlier sampling frame. Towns with population less than one lakh have been placed under stratum I, towns/cities with population one lakh or more but less than 5 lakhs under stratum II, towns/cities with population 5 lakh or more under stratum III and four metro cities of Delhi, Mumbai, Chennai and Kolkata as separate strata viz. stratum IV. The sampling unit in urban area is a Census Enumeration Block. The Census Enumeration Blocks within each size stratum were ordered by the female literacy rate based on the Census 2011 data, and three equal size substrata were established. The sample Census Enumeration Block within each substratum was selected at random with equal probability. A simple random sample of these enumeration blocks has been selected within each sub-strata without replacement from each of the size classes of towns/cities in each State/Union Territory.

The revision of SRS sampling frame is undertaken every ten years based on the results of latest census. New sampling units are drawn as per the sampling design described above for the number of units arrived at using IMR as decisive indicator for size determination.

**Question 2:** Overview of continuous data collection procedures in SRS- personal, forms, data compilation; submission of forms to state level.

**Response**

Each sampling unit (PSU) has a part time enumerator (PTE). PTE is a person belonging to the PSU or well acquainted with the area and its population.

The houselist form i.e. Form 1 remains with the PTE to identify the household particulars in respect of the events being found. He lists the events pertaining to the residents irrespective of place of occurrence of event and events pertaining to visitors occurring in the PSU. The events of birth and death are listed in Form 4 and Form 5 respectively. These forms correspond with the six month HYS Period.

Each month, the PTE submits monthly returns of events in Form 6 and Form 7. Form 6 and Form 7 are the extract of Form 4 and Form 5 respectively meant to be submitted to respective State office (DCO) once a particular month gets over. Form 4 and Form 5 are also returned to the DCO after the HYS Period is over.

PTE also updates the Form 3 which lists the pregnancy status of the married women of the PSU.

For ensuring complete netting, the enumerator uses different sources to get information of the occurrence of vital events in the sample unit. These include the help of the village priest, barber, village headman, midwife and such other functionaries. The enumerators maintain contact with these informants at frequent intervals and collects information about the occurrence of births and deaths. On being informed about the occurrence of an event, the enumerator visits the concerned household and records the prescribed particulars. The enumerator also keeps in touch with other socially important persons and visits local or nearby hospitals, nursing homes, cremation or burial grounds, at frequent intervals to keep

updated about the occurrence of events. Besides, enumerator maintains and updates a list of all women in the reproductive span along with their pregnancy status, which helps in better netting of all the births. Despite all these efforts, the enumerator may miss information about some of the events and is, therefore, required to visit all the households once a month in urban area and once in a quarter in rural areas so as to ensure that all the events have been recorded.

### **Detail of half yearly data collection/enumeration of vital events by independent surveyor**

**Question 3.a:** Forms used in the survey (household enumeration record/vital event forms)

#### **Response**

Household enumeration record: Each half yearly survey updates the list of house/household and members of each household of the PSU. Form 1 is used to update the houselist and Form 2 (Household schedule) is used to list/update the members of each household and their characteristics like marital status, education etc. Form 3 (list of married women) is also updated.

Vital event forms: Form 9 and Form 10 is used for netting of births and deaths, along with associated information, of each household respectively.

### **Data collection procedures**

**Question 3.b.i:** Complete re enumeration of all households with recording of vital events in preceding period (is the entire household re enumerated or are only details of vital events recorded in the resurvey)

#### **Response**

Entire household is enumerated and respective forms are updated/ filled.

**Question 3.b.ii:** Is there any involvement of primary enumerator in resurvey- or is it totally independent

#### **Response**

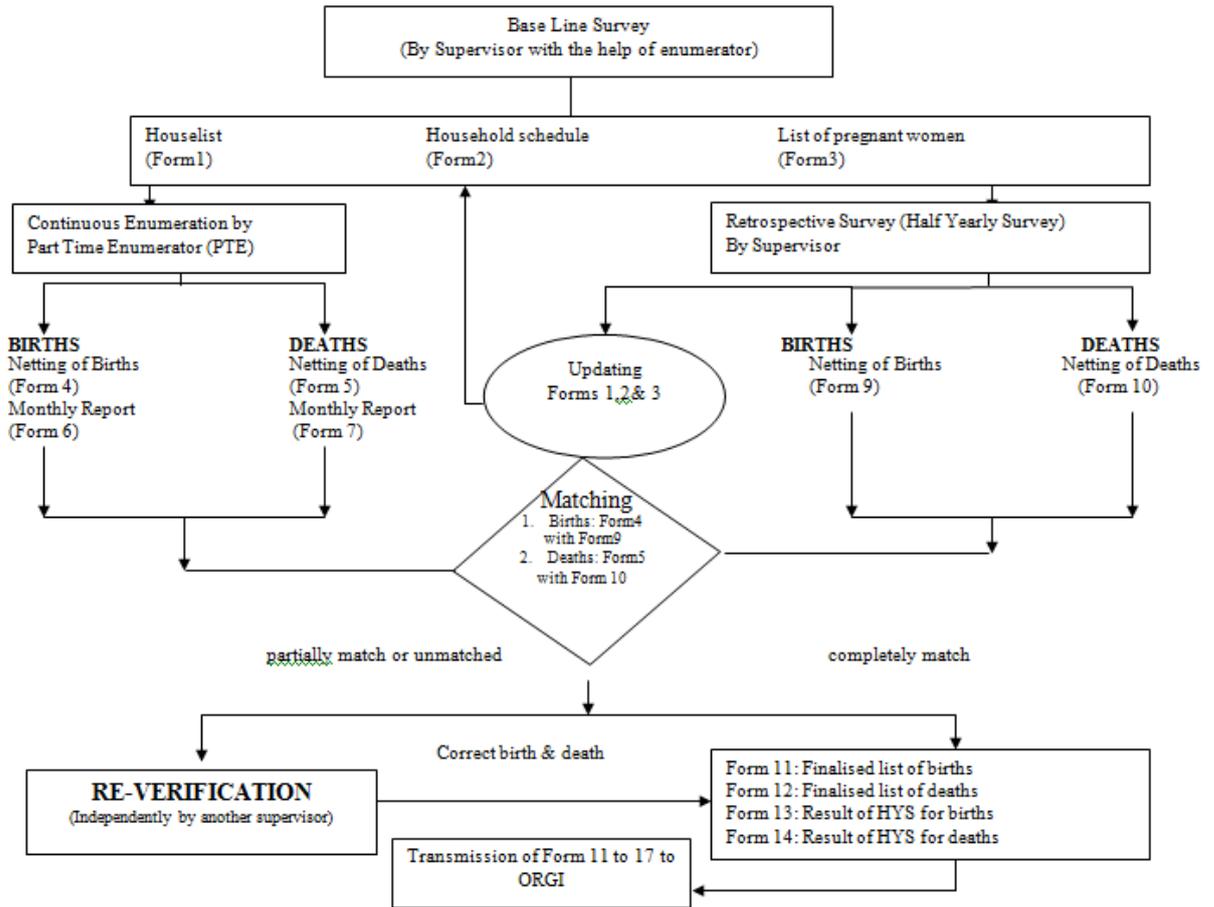
The retrospective survey is completely independent without any involvement of PTE.

**Question 3.b.iii:** Process of submission of resurvey forms to the state level- is there any sharing of details between the primary enumerator and resurvey staff

#### **Response**

Yes, Form 1 and Form 3 are shared to PTE by the resurvey staff known as supervisor. Following diagram may be referred for the data collection procedure

**Chart 1: Flow Chart**



## Process of matching of independent survey with continuous recording data

**Question 3.c.i:** Variables used

### Response

The variables used for matching are listed below:

	<b>Birth</b>	<b>Death</b>
1	House Number	House Number
2	Household Number	Household Number
3	Head of the household ID	Head of the household ID
4	Pregnant women ID	ID of mother (in case of Infant death)
5	Date of outcome	Date of death
6	Sex of outcome	Sex of deceased
7	Relationship of pregnant women to head of the household	Relationship of deceased to head of the household

**Question 3.c.ii:** Any relaxation of age, date of death, address criteria

### Response

No.

**Question 3.c.iii:** Definition of matched/partially matched events

### Response

Matched events: If all variables listed above are same in continuous recording data and independent survey then the event will be treated as matched.

Partially matched events

Birth: At least three variables, which must include House Number, Household Number and Pregnant women ID, are same in continuous recording data and independent survey then the event will be treated as partially matched.

Infant death: At least three variables, which must include House Number, Household Number and ID of mother, are same in continuous recording data and independent survey then the event will be treated as partially matched.

Other deaths: At least three variables, which must include House Number, Household Number and Head of the household ID, are same in continuous recording data and independent survey then the event will be treated as partially matched.

**Question 4:** Summary tables presenting the detailed findings from the matching process aggregated at state level for each state, before field verification. These could be presented according to the following table; for any half year/annual period over the past 1-2 years.

State	Number of deaths from primary enumerator	Number of deaths from independent survey	Fully matched number	Number partially matched	Number unmatched	Remarks

**Response**

Matching process is being carried out at DCO level in paper format till now and accordingly verification list is being prepared for each PSU. This information is not being aggregated and transferred to the central level.

**Procedure for verification of unmatched/ partially matched events.**

**Question 5.i:** What is the time period after half yearly survey when verification is done

**Response**

Within 1 month

**Question 5.ii:** Form/process for sharing the list for verification from state level to the individual SRS unit

**Response**

File containing the list of unmatched and partially matched events is provided by DCO to the supervisor to carry out verification process. For each listed event the supervisor will fill the form 9 or Form 10 depending upon the nature of event.

**Question 5.iii:** Describe the personnel involved in the verification.

**Response**

Verification is carried out by the same set of persons of the DCO who carried out the half yearly survey. However, a PSU canvassed by a supervisor in half yearly survey will be assigned to another supervisor for verification.

**Question 5.iv:** Is it joint visit by primary enumerator/ resurvey or third person?

**Response**

It is a visit by third person to ascertain the correctness of the partially matched and unmatched events.

**Question 5.v:** Forms/records completed at verification; and resubmitted to state

**Response**

The details are updated in Form 9 and Form 10.

**Question 6:** Process for compilation of final list of events after receipt of completed verification form from all field units, and compilation of aggregated list at state/national level for final analysis

**Response**

Post verification the finalized data file of birth and death are prepared on DCO server. This process merge the matched events of the PSU with the events found correct during verification. Based on this, Form 11 and Form 12 are generated which contains respectively the finalized births and deaths of the PSU.

Form 13 and 14 are prepared by aggregating the PSU wise number of birth/death for each state and half yearly survey. The number of events recorded by both PTE & Supervisor, by PTE & missed by supervisor, by supervisor & missed by PTE, missed by both for each PSU is also generated in this form.

Form 15 is generated to aggregate usual residents of a particular PSU by age, sex and marital status. Form 16 is generated to aggregate female usual residents of a particular PSU by broad age group, sex and level of education. Form 17 is generated to aggregate number of females who got married by age at effective marriage of a particular PSU. Forms 15-17 are being generated from the data captured in the Form 2 of the half yearly survey.

**Question 7:** Include example table of number of deaths from final aggregation for each state/ natural division for any recent year.

**Response**

Aggregated table used to estimate the CBR, CDR and IMR is enclosed for the state of Mizoram.

**Question 8:** Statistical method for calculating standard error/ 95% CI of mortality rates.

**Response**

Formula of standard error of ratio estimate is used for mortality rate. For illustration, Infant Mortality Rate (IMR) formula is shown below:

$$\widehat{IMR} = \frac{\text{No. of estimated Infant Death in a year}(\widehat{ID})}{\text{No. of estimated Live Birth in that year}(\widehat{B})}$$

$$Std\ Error(IMR) = \frac{[Var(\widehat{IMR}) + \widehat{IMR}^2 * Var(\widehat{B}) - 2 * \widehat{IMR} * Cov(\widehat{B}, \widehat{IMR})]}{\widehat{B}^2}$$

**Question 9:** Why is capture-recapture method not used to estimate completeness of SRS, although two data sources with record linkage are employed? Is there any issue with the assumptions of the method?

**Response**

The estimates are based on the physically found and verified events only. As estimates based on only physically found events may have better resonance with the departments of the government looking towards SRS estimates for guidance. Therefore there has not been

any attempt to introduce correction in the estimates by looking at the completeness of the events netted.