

6. Data quality and indices of reliability

Data quality is a measure of the condition of the data based on its accuracy, completeness, consistency and reliability. Measuring data quality can help identify data errors that need to be resolved and assess whether the data is fit to serve its intended purpose. In this chapter, data quality in terms of missing data, data completeness with respect to mortality information, unknown information and other indices of data quality and reliability are presented.

The five main areas of consideration of data quality in a PBSR are

1. Completeness of coverage: to obtain information on all first-ever stroke cases diagnosed in the population covered by the registry so as to ensure a high degree of case ascertainment.
2. Adequacy of data: to have certain mandatory core and critical items of patient information on all cases.
3. Accuracy of data: to ensure that data are free from erroneous abstraction, coding, data entry etc.
4. Comparability of data: to adopt standard definitions and classification like the WHO definition of stroke; type of stroke and International Classification of Diseases (ICD-10) coding respectively, that would assure comparability internally and externally.
5. Reliability of data: to ensure reliability of data through periodic audits.

In PBSRs, following exercises were conducted to ensure data quality:

1. Comprehensive details of sources of registration: It is an index to assess coverage. Table on registry wise sources of registration is presented in Section II Chapter 2 -Snapshot of registries.
2. Mandatory variables: Certain variables are mandatory at the time of data entry.
3. Data abstracted in Core form and later entered in software were verified:
 - a. Abstraction and re-abstraction by field investigators and review by PI PBSR and NCDIR team.
 - b. Verification of variables in Core form vs data entry in software by
 - i. Date-range checks [Date of onset vs Date of CT/MRI]
 - ii. Consistency check between variables [Sex v/s Title of Registered Case; Date of Onset v/s Date of Death]
 - c. Description of Final Diagnosis, ICD-10 coding, imaging findings (CT/MRI) were verified at the Coordinating unit. The reports on Quality checks were generated frequently and sent to the registries. The PBSRs verified and provided feedback to the Coordinating unit and the corrections were updated online in the PBSR software. Audit trail of changes was recorded.

- d. Duplicate Checks: It was implemented using the “Duplicate check module” and “Phonetics Module” for identifying duplicates in the PBSR data. The checks were executed on a three - monthly basis and at the end of the year to generate the probable list of duplicate case records. The registries verified and provided feedback on case record to be retained (for example first-ever stroke cases) and records to be deleted (for example, recurrent stroke cases, duplicate cases). With “Tag and delete” application, two or more identical records were deleted with tagging to correct record/ variables. The software application allowed to delete the duplicate record and tag it with the retained record.
- e. Pin code verification: Records with pin codes of geographical areas that did not belong to defined geographical area of PBSR were removed after due verification (Table 6.1).

Table 6.1. Number n of records verified as duplicates and out of area pin codes

Registry	Data received	Cases deleted from data		Final data for analysis
		Duplicates	Pin code	
Cuttack	3283	57	0	3226
Cachar	2524	28	3	2493
Tirunelveli	3746	14	2	3730
Kota	2349	2	0	2347
Varanasi	2088	60	4	2024

Quality of data:

- 1. **Missing data** : The number and proportion of stroke cases with missing data in all PBSRs is given in Table 6.2.

Table 6.2. Number n and proportion (%) of first-ever type of stroke cases with missing (blank) fields

Variables	Cuttack (N=3226)			Cachar (N=2493)			Tirunelveli (N=3730)			Kota (N=2347)			Varanasi (N=2024)		
	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)
Incidence registration number	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Name of source of registration	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Hospital registration number	3226	0	0.0	2493	192	7.7	3730	0	0.0	2347	0	0.0	2024	0	0.0
Date of registration at source/Date of reporting at this Hospital	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0

Variables	Cuttack (N=3226)			Cachar (N=2493)			Tirunelveli (N=3730)			Kota (N=2347)			Varanasi (N=2024)		
	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)	Total N	Missing n	% (n/N)
Date of onset of stroke event	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Date of diagnosis of first ever stroke	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Place of residence (Urban/Rural)	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Duration of stay	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Age	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Hypertension	3226	0	0.0	2493	667	26.8	3730	1	0.0	2347	1	0.0	2024	2	0.1
Diabetes	3226	5	0.2	2493	679	27.2	3730	4	0.1	2347	1	0.0	2024	8	0.4
Current Tobacco Use	3226	5	0.2	2493	690	27.7	3730	4	0.1	2347	0	0.0	2024	33	1.6
Clinical findings at onset (Critical clinical findings/ No critical clinical findings/Others)	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
CT Done/Not done	3226	51	1.6	2493	676	27.1	3730	592	15.9	2347	19	0.8	2024	176	8.7
If CT done, date of CT	2681	1	0.0	1811	22	1.2	2415	2	0.1	1486	0	0.0	1724	1	0.1
If CT done, CT impression	2681	1	0.0	1811	0	0.0	2415	13	0.5	1486	30	2.0	1724	0	0.0
MRI Done/Not done	3226	3071	95.2	2493	2489	99.8	3730	2012	53.9	2347	33	1.4	2024	1267	62.5
If MRI done, Date of MRI	84	0	0.0	3	0	0.0	935	1	0.1	1011	0	0.0	137	0	0.0
If MRI done, MRI impression	84	0	0.0	3	0	0.0	935	0	0.0	1011	0	0.0	137	0	0.0
Date of actual follow-up	3226	16	0.5	2493	771	30.9	3730	135	3.6	2347	2	0.1	2024	95	4.7
Method of follow-up	3226	17	0.5	2493	688	27.6	3730	20	0.5	2347	11	0.5	2024	133	6.6
Vital Status	3226	0	0.0	2493	0	0.0	3730	0	0.0	2347	0	0.0	2024	0	0.0
Place of Death	657	15	2.3	1063	44	4.1	753	4	0.5	302	13	4.3	804	0	0.0
Cause of Death	657	12	1.8	1063	0	0.0	753	1	0.1	302	13	4.3	804	0	0.0

Additional indices of data quality are presented

1. Completeness of data with respect to mortality information (Table 6.3)

Table 6.3. Number (n) and proportion (%) of death cases with completed fields for mortality information

Field/Variable	Cuttack		Cachar		Tirunelveli		Kota		Varanasi	
	(n=657)		(n=1063)		(n=753)		(n=302)		(n=804)	
	n	%	n	%	n	%	n	%	n	%
Date of Death	657	100.0	1063	100.0	753	100.0	302	100.0	804	100.0
Place of Death	642	97.7	1019	95.9	749	99.5	289	95.7	804	100.0
Death Certified by	644	98.0	657	61.8	751	99.7	287	95.0	768	95.5
Cause of death	645	98.2	1063	100.0	752	99.9	289	95.7	804	100.0

2. Follow up status is unknown (Table 6.4)
3. Actual date of follow is unknown (Table 6.4)
4. Vital status is unknown (Table 6.4)

Table 6.4. Number (n) and proportion (%) of stroke cases with unknown information

Field/Variable	Cuttack		Cachar		Tirunelveli		Kota		Varanasi	
	(N=3226)		(N=2493)		(N=3730)		(N=2347)		(N=2024)	
	n	(n/N)	n	(n/N)	n	(n/N)	n	(n/N)	n	(n/N)
		%		%		%		%		%
Follow up (unknown)	16	0.5	1	0.0	0	0.0	0	0.0	1	0.0
Date of actual follow up (unknown)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Vital Status (unknown)	58	1.8	272	10.9	1	0.0	1	0.0	2	0.1

5. Incidence Mortality Matching: All-cause death data of 2018 and 2019 from the CRS were verified for deaths due to stroke. Varanasi PBSR provided only all-cause mortality death data of 2019 from CRS. The death records were abstracted on key variables like name, age, address, date of death and matched with Incident Stroke Registration cases. The records that were matched were verified by the registry to confirm death of the registered stroke cases (Table 6.5.).

Table 6.5. Number (n) of records matched using Incidence mortality matching application

Incidence mortality matching	Total
Cuttack	20
Cachar	1
Tirunelveli	1
Kota	15
Varanasi	0

6. Death Certificate only (DCO): DCO are cases for which there is no other information other than a death certificate with stroke mentioned as a cause of death. These cases were obtained from the CRS and could not be matched with any of the registered cases. The proportion of these cases reflected the proportion of stroke cases that may not have accessed health care or missed by the registry. Number and proportion of DCO cases in all PBSRs is reported in Table 6.6.

Table 6.6. Number (n) and proportion (%) of Death Certificate Only (DCO) cases

Registry	Registered incidence	Registered mortality	DCO Cases	
			n	%
Cuttack	3226	657	86	2.7
Cachar	2493	1063	645	25.9
Tirunelveli	3730	753	199	5.3
Kota	2347	302	0	0.0
Varanasi	2024	804	55	2.7