SESSION -IV
DATA QUALITY

NATIONAL RURAL HEALTH MISSION
राष्ट्रीय ग्रामीण स्वास्थ्य मिशन
(2005-2012)

SESSION OBJECTIVES

After this session you should be able to:

1. Describe what is data quality and why it is important.
2. Define methods of data quality check.
3. List reasons of poor data quality and explain how to address them.
4. Address process to improve data quality.

DATA QUALITY

Data quality refers to the extent to which data measures what they intend to measure.

Dimensions of data quality-

- Completeness
- Timeliness
- Accuracy

COMPLETENESS

Reports are a reflection of services provision and utilization thus an incomplete report will indicate partial service delivery/utilization.

Data completeness is assessed for the following:

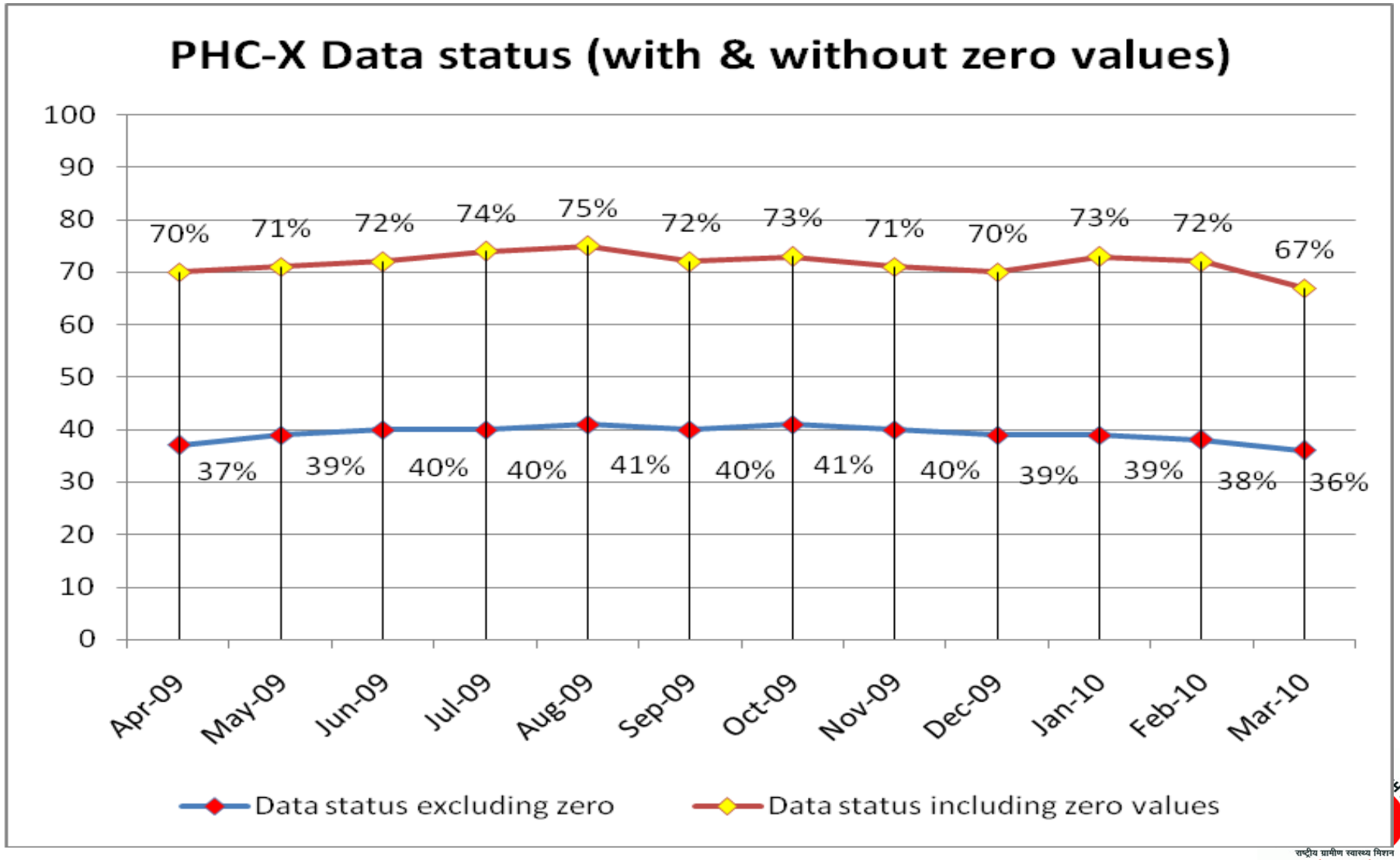
1. Number of facilities reported against total facilities
2. Number of data elements reported against total data elements in a reporting form.

Reporting from "Private Facilities"?

ACTIVITY-1
ASSESSMENT OF DATA/REPORT COMPLETENESS

- While assessing completeness remember zero and blank values.
- Generate data completeness status report by including as well as excluding zero values.

PHC-X data status for 12 months is given in the graph below.



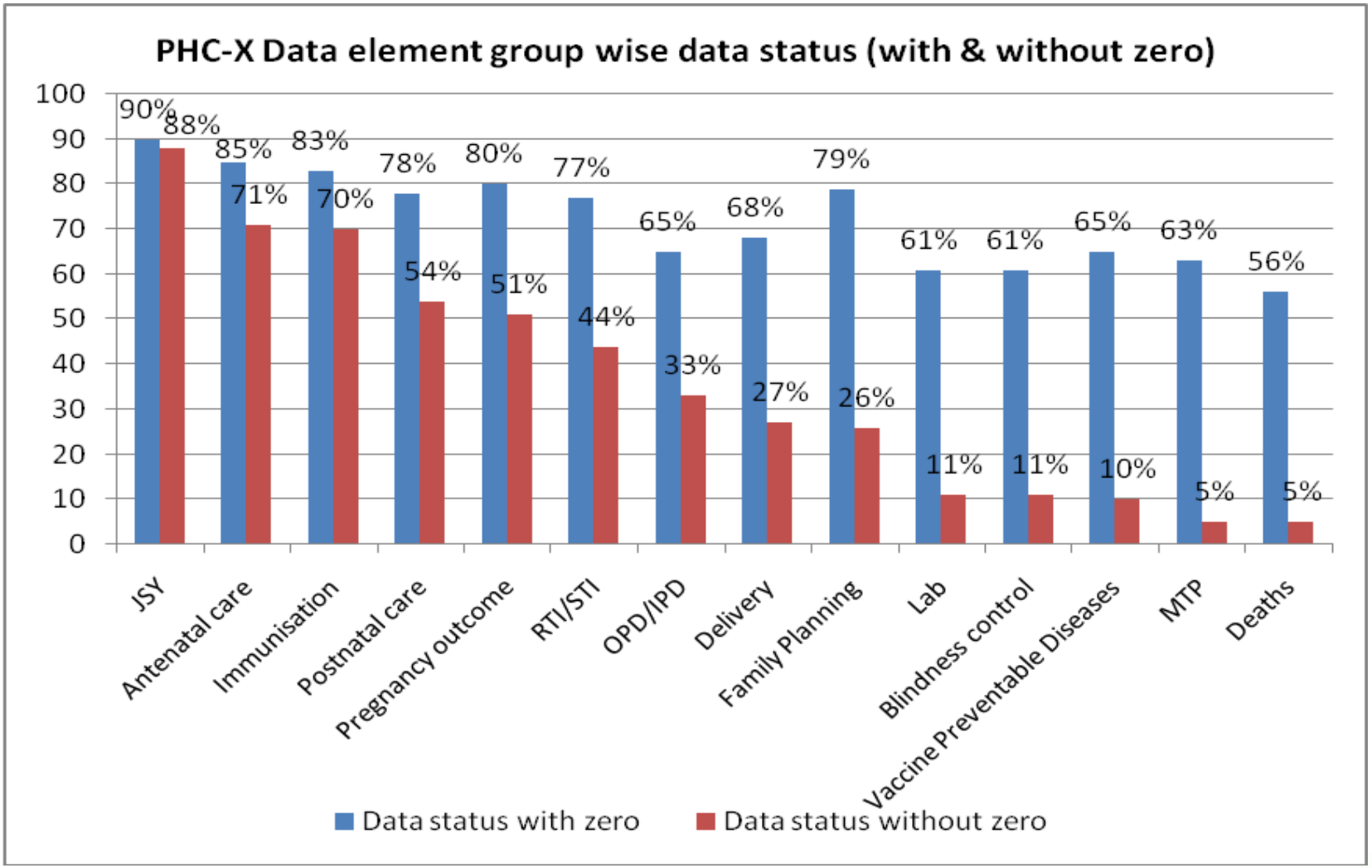
Observations

- Data status is consistent for across months
- If zero values are included the data status is more than 70%; which means out of total sections in the report more than 70% are filled.
- When zero is excluded data status reduces by 30% or more; which means out of total sections filled 30% or more had zero values.

Observations contd..

- What could be reasons for such reporting.
  - unavailability of services in these facilities,
  - unavailability of recording registers for these events,
  - or simple ignorance
- To drill down further we can look data status in data element groups and find out which sections had very less data.

PHC-X Data status for data element groups



Observations

- Sections which had very less data-
  - Lab
  - Blindness Control Program
  - MTP
  - Family Planning
  - Delivery
    - *Vaccine Preventable diseases*
    - *Deaths*

Common rule to report zero/blank

- If service is available but not provided due to any reason put zero e.g., IFA
- If service is available but no beneficiary came put zero e.g., condom
- If service is not available left blank e.g., C-section

TIMELINESS

- Timeliness is very important component of data quality. Timely processing and reporting of data facilitates timely availability of data for decision making.

Example: During monthly review meetings, if out of 10 sub-Centers 5 do not submit report on time it will be difficult for the MO to assess the performance and develop a plan for PHC in particular and of sub-Centers in general.

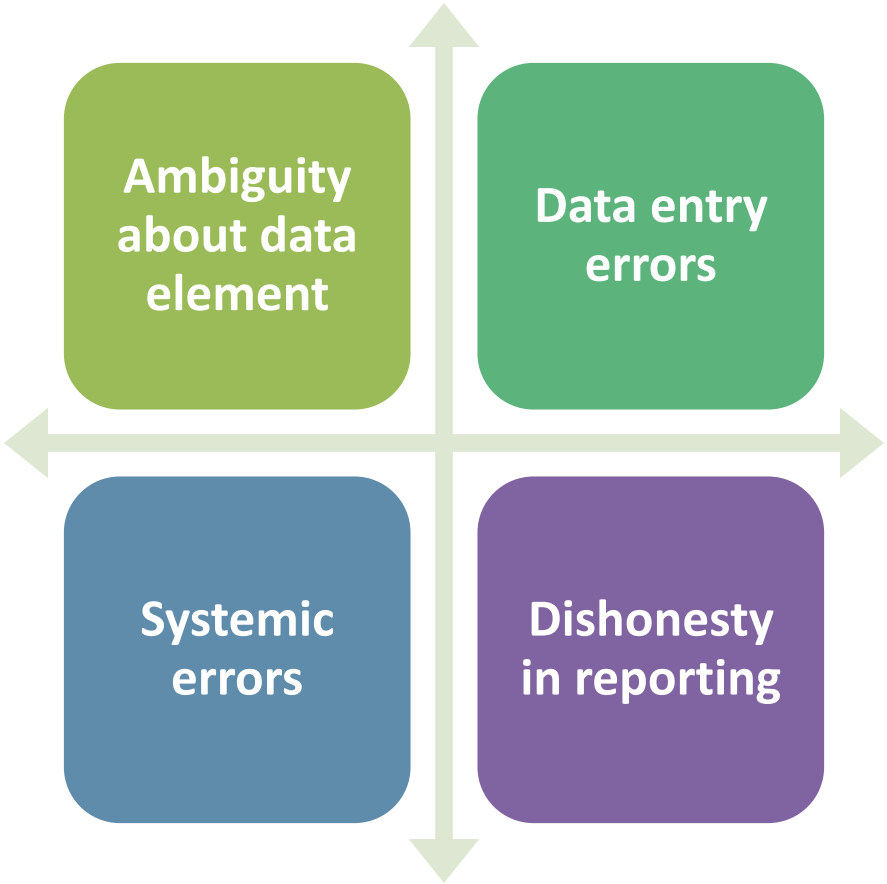
Check for the date of reporting for every facility and find out when all facilities report in your district.

ACCURACY

- Accuracy refers to the correctness of data collected in terms of actual number of services provided or health events organized.
- Inaccurate data will yield incorrect conclusions during analyses and interpretation.
- Small errors at facility level will cumulate into bigger mistakes since data from various providers/facilities are aggregated.

ACCURACY CONTD..

Poor data accuracy could be due to following four factors



Example: Examine ANC data reported by all the blocks of District X and check for accuracy in data.

Data elements	Block A	Block B	Block C	Block D	Block E	Total
Total ANC registrations	1230	1367	2359	1667	991	7614
100 IFA tablets given	1008	1300	235999999	166700	784953	236953960
ANC 100 IFA coverage rate	82.0%	95.1%	10004239%	10000%	79208%	3112082%

Observations

- Block A & B have reported correct figures and no problem was found while processing/analyzing data.
- Block C reported high number of IFA beneficiaries but looking at the figure, one can easily identify typing mistake rather than any systemic problem in reporting.
- Probably Block D reported number of tablets given rather than number of pregnant women.
- Data from Block E is intriguing; probably the Block had high number of actual beneficiaries or lactating women and adolescents were also reported or pregnant women were not given IFA in past months because Block was out of stock and now back log was being cleared. Further probing in required to identify the error.

DATA ENTRY ERRORS

- **Typing errors**: wrong numbers entered in computer
- **Wrong box entry**: data entered in wrong box e.g., 'ANC registration' data entered in 'Registration in first trimester'.
- **Calculation errors**: during data entry basic computation happens if formulae are incorrect than errors can happen.

Data entry errors can be corrected through:

- **Visual scanning:**

	PHC A	PHC B	PHC C	PHC D
Total ANC registration	281	328	491	267
Early ANC registration	90	100	214	95
ANC Third visits	211	309	425	186
ANC given TT1	247	295	424	250
ANC given TT2 or Booster	277	305	425	231
ANC given 100 IFA	276	296	438	253
ANC moderately anemic < 11 gm	68	67	114	51
ANC having Hypertension –New cases	20	76	15	4711

Performing validation checks

- Validation is performed by comparing values of 2 (or more) data elements that are comparable.

Validation rule	Left side	Operator	Right side
Early ANC registration is less than or equal to total ANC registration	Early ANC registration	≤ (less than or equal to)	Total ANC registration

Common Validation Rules

Data Validation Rules	
1	**ANTENATAL CARE**
I	ANC registration should be equal or greater than TT1
II	Early ANC registration must be ≤ to ANC registration
2	**BLINDNESS CONTROL**
I	Eyes collected should be more or equal to eyes utilized
II	Patients operated for cataract should be more than or equal to number of IOL implanted
3	**DELIVERIES**
I	Deliveries caesarean must be ≤ to deliveries institution
II	Deliveries discharged under 48 hours ≤ deliveries at facility
III	Institutional deliveries should be ≤ BCG given
IV	Institutional deliveries should be ≤ OPV0 given
V	Total deliveries should be equal to live births + still births
4	**IMMUNISATION**
I	BCG should be ≤ to live births
II	Immunisation sessions planned should be greater than or equal to sessions held
III	Measles dose given should be greater than or equal to full immunization
IV	OPV Booster should be equal to DPT Booster
V	OPV1 should be equal to DPT1
VI	OPV2 should be equal to DPT2
VII	OPV3 should be equal to DPT3
VII	Vitamin A dose should be equal to measles dose

Common Validation Rules

5	**JSY**
I	ASHAs and ANMs/AWWs paid JSY incentive for institutional deliveries is ≤ to mothers paid JSY incentive for institutional deliveries
II	JSY incentive for home delivery must be ≤ to home deliveries at sub-Centre
III	JSY incentive to mother should be ≤ to deliveries
IV	JSY registration must be ≤ to new ANC registrations
6	**NEWBORNS**
I	Newborns breastfed within 1 hour are less than total live births
II	Newborns weighed at birth ≤ total live births
III	Newborns weighing less than 2.5 kgs ≤ total newborns weighed
7	**POST NATAL CARE**
I	Women receiving first (within 48 hour) post-partum checkup ≤ to total live births plus still births

NATIONAL RURAL HEALTH MISSION
राष्ट्रीय ग्रामीण स्वास्थ्य मिशन
(2005-2012)

Does Validation always indicates an error?

- It is important to note that violation of a validation rule does not always indicate error. Violations can be due to-
  - Management issues like availability of vaccines or medicines in stock,
  - Disease outbreak
  - Actual improvement due to a good BCC program.
- Violation of validation rule prompts you to enquire and check/verify data until satisfactory answer is not found.

ACTIVITY 2

1. Check your last month data using any of the five validation rules.
2. Make group of 4-5 participants.
3. Pick any of the last month's report of your district.
4. Apply any five validation rules given in the table.
5. Identify validation queries and find out what reasons could be for these queries.

IDENTIFICATION OF STATISTICAL OUTLIERS

- Statistical outliers are numbers that do not conform to the trend or are unexpected values.
- In statistical terms, if the value lies 1.5 Standard Deviations away from the range (can also be viewed on stem and leaf plot) it is identified as an outlier.
- This often helps to identify data entry errors or large computation mistakes.

SYSTEMIC ERRORS

Systemic errors are those which are embedded in the system and due to these data quality always remains poor.

Problem 1: Errors due to multiple registers

Commonly Missing Data Elements in Recording Registers

1. Breast feeding within first hour
2. New cases of hypertension
3. Failure/complications and death due to sterilization
4. Adverse event following immunization
5. IUD removals
6. Hb test for ANC
7. Midnight head count
8. Total number of times ambulance used for transporting patients
9. Adolescent counseling services
10. JSY registration at time of ANC
11. Total number of 9-11 months old fully immunized children

NATIONAL RURAL HEALTH MISSION
राष्ट्रीय ग्रामीण स्वास्थ्य मिशन
(2005-2012)

Possible Solutions

- Create a compact ‘Service Delivery Recording Register’ for ANM to carry to the field. This register should have all relevant data elements related to ANC, PNC, Immunization, Family Planning, and OPD. (See Chapter 2) Then when she comes back to her office she transfers the data onto each specific child health, maternal health, eligible couples register.
- Discourage recording in ‘rough diaries’.

Problem 2: Misinterpretation of Data Elements

Data Element	District A	District B
Number of pregnant women given 100 IFA tablets	25	3500

Solution

Each data element needs to be clearly defined and interpreted not only in English language but also in local language.

Data dictionary **must** be available with every service provider recording or reporting data.

Problem 3: Consistency of terms used

- Alignment between the recording and the reporting registers.

Example: Recording format of a sub-Centre in one State does not have data element, 'ANC registration in first trimester', whereas the reporting format has it.

- Consequently, data element either gets reported as blank or as zero, implying that no women at that sub-center were registered for ANC in first trimester.

Solution

Step 1: Review and compare recording and reporting formats at each level to identify data elements that are missing.

Step 2: List data elements that are duplicated in two or more of her reporting registers, e.g., Births or Deaths.

Step 3: Add rows and columns in recording register to accommodate missing data elements.

Step 4: Make notes against data elements that are duplicated, in order to ensure consistency in reporting.

Problem 4: Computation problem

Child Immunization	No. of children
BCG	10
DPT1	12
DPT2	12
DPT3	9
OPV0	9
OPV1	10
OPV2	10
OPV3	9
Hep B1	5
Hep B1	5
Hep B1	8
Measles	10

X Incorrect data compilation	√ Correct data compilation
Add all the numbers and report that 109 children aged 9-11 months were fully immunized in a month.	Only those children who have received BCG, all doses of DPT, OPV and during this month have received Measles dose will be counted as fully immunized. Note- All children who have received Measles dose during the month may or may not be fully immunized.

Problem 5: Problem in data aggregation

REPRODUCTIVE AND CHILD HEALTH							
Ante Natal Care Services	**Block A**	**Block B**	**Block C**	**Block D**	**Block E**	**Block Total**	**District Report**
Total number of pregnant women registered for ANC	387	457	2114	2076	2586	7620	11110
Of which number registered within first trimester	20	288	2142	1636	1202	5288	5288
New women registered under JSY	0	401	169	1765	1588	3923	5445
Number of pregnant women received 3 ANC check ups	2984	239	1357	1679	124	6383	6383
TT1	3446	697	1966	1974	2974	11057	11057
TT2 or Booster	3306	520	1633	1668	2882	10009	10009
Total number of pregnant women given 100 IFA tablets	141	284	41893	235	3349	45902	52022
New cases of pregnancy hypertension detected at institution	0	255	0	5	370	630	630
Number of eclampsia cases managed during delivery	0	0	0	17	2	19	19

Solution-

- Facility-wise data entry in HMIS application
- Data aggregation using MS Excel sheets

Problem 6: Lack of written guidelines & procedures

Example: If data are entered at Block as 'Block consolidated report' and few facilities have not reported, what actions Data Manager should take?

- Make block report based on available data and exclude data for facilities that did not report.
- Impute previous month's data
- Impute data of same month but of previous year
- Estimate data/values based on numbers reported in neighboring locality.

Solution

- In absence of consistent protocols for missing/incomplete data it is very difficult to procure good quality data.
- Blocks should report all data they have and explain which facilities did not report and why.
- Also, if backlog data are entered then an explanatory note should be appended.

Problem 7: Logistical Problems

Non-reporting/inconsistent reporting can be due to –

- Shortage of pre printed forms
- Traveling time to submit report.
- Quality of reporting forms due to repeated photocopy remains very poor.
- Possible solutions for health worker- create their own forms
  - "which are different from the prescribed formats
  - "consistency of data change frequently."

Solution

- Facilities should be have adequate reporting forms on annual or six monthly basis as desired by the state.
- Clear instructions should be given to field staff that report data on printed forms and no other forms will be accepted for reporting.

Problem 8: Duplication

- Data duplication leads to false higher coverage of services and inaccurate decision making.
- For example if a pregnant women delivers in the CHC, ANM should not report this delivery.
- She can record this delivery in her register because the pregnant woman is registered with her but she should not report it. If ANM reports this delivery and CHC also reports, this leads to duplication.

Solution: Follow data collection & reporting guidelines.

Problem 9: Data reported for nonexistent services

- Haemoglobinometer is not available
- SC report says there are ‘pregnancy anemia’ cases
- ANM reports ANC anemia based on clinical examination.
- What problem you can face by this?
  - it adversely affects data accuracy because ANM may overestimate or underestimate anemia cases.

Solution: Follow data collection & reporting guidelines.

Problem 10: Reporting of missing values or lower figures

- This refers to a common problem of "compensatory low figure reporting" because high values were reported in previous months.

Solution:

- As explained above HMIS Managers should check and verify statistical outliers or unusual numbers against the field records and if values are genuine then instead of asking field staff to compensate the data an explanatory note should be send to the higher level.

Problem 11: Wrong choice of indicators /denominators

- This refers to a common problem where data element itself is correct but denominator chosen is inappropriate.
- Example- When estimating the population of a district one has to extrapolate the population from 2001 census data to the mid-year population of the corresponding year then from this number derive expected population for different age groups and categories.
- Failure to extrapolate will lead to higher rates or we may be counting the numerator only from public health facilities whereas the denominator may included all patients seen by both public and private facilities e.g. while calculating C-section rate against expected pregnancies this too could lead to misinterpretation. In some districts migration could affect denominator and so on.

WHAT TO DO IF THE ERROR HAS BEEN FOUND IN REPORT?

- If the error was found in the facility report then go back to the registers and check the value, correct it, and also mark a note about the change made.
- Make sure that your staff understands meaning of this data element.
- Ensure that registers have space to record these data.
- In the forthcoming month check the value to ensure that they have understood the importance of this procedure that you followed.
- It is important to have data reporting guidelines are strict adherence to the same. There can be unique situations where strict adherence to guidelines might not be feasible, these should be made note of and Government Orders should be requested to resolve such errors.

The key to data quality is the use of information. the more regularly it is used, the more the seriousness with which data is entered and problems in flow and analysis are sorted out!! And for this feedback at every level is a must

Thanks


NATIONAL RURAL HEALTH MISSION
राष्ट्रीय ग्रामीण स्वास्थ्य मिशन
(2005-2012)