

# INTRODUCTION TO QUALITY ISSUES IN HMIS DATA



# NRHM- Objectives

- National Rural Health Mission (NRHM)- Improve healthcare delivery by making architectural corrections in existing public health systems
- NRHM plans objectives ...
  - Improve health status of the people
  - Ensure universal access to quality health care services
  - Increase public expenditure on health
  - Reduce health inequities
  - Ensure greater accountability and responsiveness

# NRHM

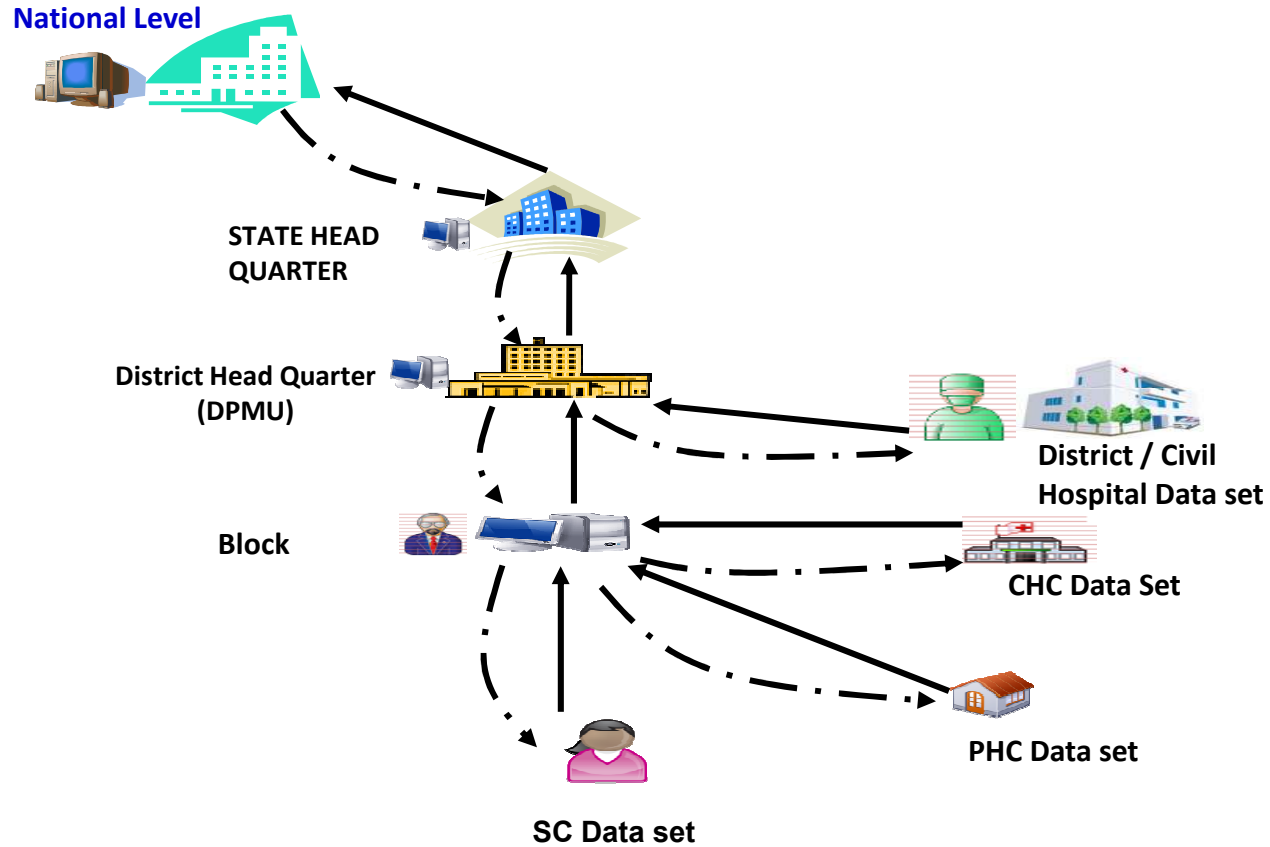
- Ensure decentralization and district level planning and management of health programmes
- Ensure community participation in planning and management
- Ensure all public health facilities attain Indian Public Health Standards- in terms of services delivered, human resource deployed , infrastructure, equipment and supplies.
- Induct multi-disciplinary skills into district health systems- management skills, financial skills, IT and data analysis skills, epidemiological skills etc.

***A health management information system is an essential management tool to achieve these goals- to be able to monitor how far we have travelled, and where the gaps are...***

# Health Management Information Systems(HMIS)

- **Definition:** ‘Health Management Information Systems (HMIS)’ is a tool which helps in gathering, aggregating, analyzing and using information for taking actions to improve performance of health systems.
- **The Mandate of HMIS:** To ensure that there is a continuous flow of good quality disaggregated data on health of populations and health care services to assist in local planning , programme implementation, management, monitoring and evaluation.

# Flow of Data



# BASIC CONCEPTS/TERMS

## Data Element and Data

Data Element is a recorded event. Data is an aggregation of data elements - in the form of numbers, characters, images -that gives information after being analyzed



## Information

is data organized with reference to a context.- which gives data a meaning



## Knowledge

when information is analyzed, communicated and acted upon, it becomes knowledge.

**Data:** No. of pregnant women in an area who received skilled birth assistance



**Information:** % of pregnant women received skilled birth assistance & % of pregnant women who were left out



**Knowledge:** Why are some pregnant women able to receive skilled birth assistance? Why some pregnant women who were left out? Who were left out? What are the issues related to access to service?

# DATA ELEMENT

A data element is a record of health event or health related event.

- Data Elements are recorded in a **primary register (recording formats)** by the **service provider**.
- Similar events for the month are **aggregated and** reported in specified **reporting formats**.

Example:

Number of pregnant women who received an antenatal check-up	Record of a service delivered
Number of children below five years who were affected with measles.	Health event
Number of female children born in the preceding month.	Health related event



# TYPES OF DATA ELEMENT

- Simple data element
- Disaggregated data elements
- Calculated data elements

Number of BCG immunization given to children

Age , sex, community of the child immunized ( not added urban/rural or BPL/not BPL- if we add it becomes 52)

Calculated data elements (information derived from adding data elements)

	Number of BCG immunization given to:	Category	Examples
1.	Male child below 1 year	SC	1
2.	Male child below 1 year	ST	0
3.	Male child below 1 year	Others	2
4.	Female child below 1 year	SC	2
5.	Female child below 1 year	ST	0
6.	Female child below 1 year	Others	1
7.	Male child above 1 year	SC	2
8.	Male child above 1 year	ST	1
9.	Male child above 1 year	Others	1
10.	Female child above 1 year	SC	1
11.	Female child above 1 year	ST	0
12.	Female child above 1 year	Others	1
13.	No. of BCG Immunization given to children	Total	12

# INDICATOR

- An indicator is a data element placed in a given context so that it becomes information that can be used for program monitoring, management, and action.
- Indicators help us assess our performance/progress across time and across places.
- Indicators also serve as a yardstick for comparison with external sources.
- Steps to convert data element into indicator
  - ▣ Identify a data element as the numerator
  - ▣ Divide it by another data element which represents the context- the denominator
  - ▣ Multiply it by a factor to make it easily readable

# INDICATOR

## EXAMPLE

- **Data element:** Total number of children in the 12 to 23 month age group who have been given Measles vaccine=360
- **Numerator:** Total number of children given measles=360
- **Denominator:** Total number of children in the age group 12 to 23 months=450
- **Multiplying factor:** 100
- **Calculation:**  $360/450*100=80\%$
- **Indicator:** Percentage of children in the 12 to 23 month age group given Measles vaccine=80%.

# INDICATOR

## EXAMPLE

- **Data element:** Total number of ASHA received incentive during last month = 11060
- **Numerator:** Total number of ASHA received incentive during last month = 11060
- **Denominator:** Total number of ASHA in the district = 24500
- **Multiplying factor:** 100
- **Calculation:**  $11060/24500*100=45.14\%$
- **Indicator:** Percentage of ASHA received incentive during last month = 45.14%

# INDICATORS FOR COMPARISONS

## — across different places and times

- Indicators help us assess our performance/progress across time and across places. Indicators also serve as a yardstick for comparison with external sources.

### Example:

- **Indicator:** Percentage of registered pregnant women who had an institutional delivery=50%
- **Data elements:** Number of institutional deliveries conducted last year in the PHC=234 and Total number of registered pregnant women=468
- Medical Officer/Supervisor can assess...
  - Did all deliveries happen in the institution?
  - Are delivery services in the PHC utilized well?
  - How is ASHA program working in the District? Is ASHA working properly to motivate mothers to come for institutional delivery? Where are the gaps in the program?

# INFORMATION TO KNOWLEDGE & ACTION

- Communicable disease data for July for PHC 'A' and data related to disease prevention and control are given below:

Number of villages reported fever cases	12
Number of villages surveyed this month for fever cases	8
Number of fever cases this month for whom blood smear examination (SME) done	100
Blood smear malaria positive cases	74
Number of positive cases seen and given Chloroquine tablets	42
Number of vector breeding sites identified	102
Number of vector breeding sites destroyed	78

This is the data related to a malaria control programme in a village?  
What is the action we need to take?

## Converting data into information/indicators

Indicators	Numerators	Denominators	Multiplying factors	Values
Percentage of villages reporting fever cases surveyed this month	Number of villages surveyed this month for fever cases=8	Number of villages having reported fever cases=12	X 100	66.67%
Percentage of fever cases positive for malaria	Blood smear malaria positive cases=74	Number of fever cases this month for which blood smear tested=100	X 100	74.00%
Percentage of malaria positive cases given Chloroquine	Number of positive cases seen and given Chloroquine tablets=42	Blood smear malaria positive cases=74	X 100	56.75 %
Percentage of vector breeding sites destroyed	Number of vector breeding sites destroyed=78	Number of vector breeding sites=102	X 100	76.00%



# DATA QUALITY

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

## Dimensions of data quality-

- Completeness
- Timeliness
- Reliability/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”?

# 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.

# Causes of decreased completeness and timeliness

- ❑ Poor internet connectivity
- ❑ Lack of hardware in some facilities
- ❑ Lack of staff
- ❑ Lack of supervision

# ACCURACY AND RELIABILITY

- 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 AND RELIABILITY..

Poor data accuracy/reliability 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 is 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:

## 1. 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

## 2. 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	$\leq$ (less than or equal to)	Total ANC registration

# Common Validation Rules

Data Validation Rules	
<b>1</b>	<b>ANTENATAL CARE</b>
I	ANC registration should be equal or greater than TT1
II	Early ANC registration must be $\leq$ to ANC registration
<b>2</b>	<b>BLINDNESS CONTROL</b>
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
<b>3</b>	<b>DELIVERIES</b>
I	Deliveries caesarean must be $\leq$ to deliveries institution
II	Deliveries discharged under 48 hours $\leq$ deliveries at facility
III	Institutional deliveries should be $\leq$ BCG given
IV	Institutional deliveries should be $\leq$ OPV0 given
V	Total deliveries should be equal to live births + still births
<b>4</b>	<b>IMMUNISATION</b>
I	BCG should be $\leq$ 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

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

# 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.

### 3. 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 poorly designed primary registers

1. Data element required to report in the form are not there and gets missed-out while reporting.
  2. Data element present- but cannot be computed easily or prone to recording errors.
  3. Multiple registers.
- Solution : Rationalization of Primary Registers. – keeping the service delivery recording function, the tracking function and the computing function, distinct and visible- checking to see all data required is present in the record and lends itself to computation

# Problem :

## b. Computation problem in register

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

<b>X Incorrect data compilation</b>	<b>√ Correct data compilation</b>
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 2: Data Definitions:

### a. Misinterpretation of Data Elements

Data Element	District A	District B
Number of pregnant women given 100 IFA tablets	25	2500

#### 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 in their own language:

# Data definitions

## b. Consistency of terms used

- Alignment between the recording and the reporting registers.

### **Example:**

- a. 'ANC registration in first trimester', 'Early ANC registration'
- b. What is Pregnancy registration, what is JSY registration.

## Problem 3: Problems 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

# 3.Data Aggregation problems

- 1. Data difficult to add up across hundreds of facilities- especially manually, disaggregated....need applications**  
Facility-Wise data entry in off line application computes a block and district aggregation sheet which can then be uploaded. MS Excel sheets if nothing else available.
- 2. Clarity on which facilities get added up where.**  
Denominators relate to “catchment areas” for facilities.
- 3. Facilities reporting late, or not at all..... Needs rules to cope with this.**
- 4. Providing feedbacks as block aggregated forms / sector aggregated forms as well as comparisons of facilities in a block**

# Problem 4: Confirmation and Error Management Procedures

**No clear delegation of powers for approving or confirming data.**

**Especially needed for late reporting facilities, non reporting facilities, cumulative data coming in, error management**

**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-** Written guidelines should be in place

# Problem 5: Logistical Problems

Non-reporting/inconsistent reporting can be due to –

- Form Problems: Shortage of pre printed forms , lack of standardization of forms, poor Quality photocopy etc.
- Traveling time to submit report.
- Lack of staff or hardware for data entry

## Solutions

- Forms adequate for six monthly basis
- Attend to hardware/staff problems or relax schedules accordingly.
- Mobile communication to save travel and staff time- but requires more applications management.



# Problem 6: Duplication

- Data duplication leads to false higher coverage of services and inaccurate decision making. It covers up for lack of private sector data.
- For example if a pregnant women delivers in the CHC, ANM should not report this delivery. But if she delivers in a private sector also she should not report it!!
- She can record this delivery in her register because the pregnant woman is registered with her but she should not report it. Could be entered in tracking page- but not in service page.
- **Central Decision: Area Reporting or Service Reporting.**

## Problem 7: The Zero Problem: How to reported non-existent vs Non utilized services

- Example: Haemoglobinometer is not available ; HSC 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.- suggest these are reported as zero- and no difference be made between zeros and blanks.

## Problem 8: 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.

## Problem 9. Inability to create indicators- or too many data elements for one indicator.

- Not per se a problem of data quality- but because of failure to use data- there is no scrutiny of data element.
- **Each data element must contribute to 1 to 1.5 indicators.**
- Need to identify and remove data elements that are not used.
- Some like couple protection rate- need far too many data elements to compute- high degree of inaccuracy results.

# Problem 10: Death Reporting issues

- Line lists ill understood- and ill adhered to.
- If applications/data entry operators cannot handle line lists- manual conversion to tables must be made available at the facility level.
- This is area reporting- but here also duplication avoidance rules need to be created.
- Areas which are under-reporting deaths need to be identified and worked upon.
- There are many 4 Categories of “cause not known”- and rules regarding aggregation of these poorly appreciated.

# WHAT TO DO IF THE ERROR IS FOUND?

1. Check denominator and how indicator was calculated.. Ok..
2. Triangulate with other sources of information within the same format and with DLHS: is it error at all? Or is it a surprise but true finding to be acted on?.. Then....
3. Disaggregate to next level- see if over- under reporting is uniform or whether it comes from one block/ one facility.

*If the error was found in one facility report then :*

- A. Make sure that it was not a data entry error- which could be systemic or random.
- B. Go back to the registers and check the value, correct it, and also mark a note about the change made.
- C. Ensure that registers have space to record these data.
- D. Make sure that your staff understands meaning of this data element.
- E. Check if there is a data collection problem
- E. In the forthcoming month check the value to ensure that they have understood the importance of this procedure that you followed.

# If error is coming across blocks

- Check what category of error it belongs to- in the Problem 1 to 10 category outlined. See if there is a guidelines in place to which we need to promote strict adherence or whether a guidelines need to be issued. There can be unique situations where strict adherence to guidelines might not be feasible, these should be made note of.
- Get Government Orders issued from district level to resolve the errors.
- Ensure that every HMIS manager/data entry operator keeps a file containing all such orders.
- Ensure Training Manuals 1 and 2 are loaded on to computer in soft version.

***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***