

# STRENGTHENING ANALYSIS AND USE OF ROUTINE FACILITY DATA FOR MATERNAL, NEWBORN, CHILD AND ADOLESCENT HEALTH



## Data quality considerations for MNCAH managers

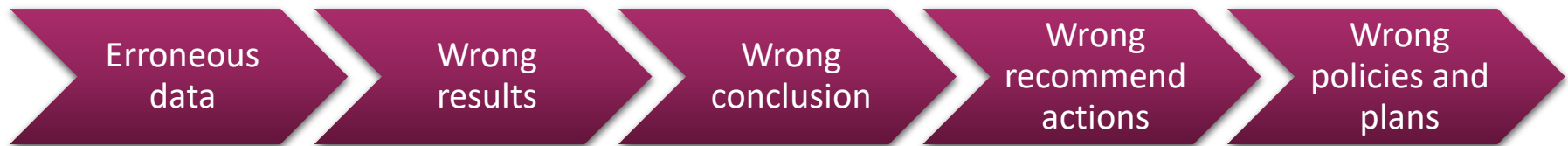
## Objectives of the session

---

- Describe common data quality problems with routine health information system (RHIS) data.
- Explain the importance of data quality with respect to using RHIS data on maternal, newborn, child and adolescent health (MNCAH) for decision-making.

With an increased demand for and need to use routine data, issues of data quality must be addressed.

- High quality data are reliable and accurate.
  - Data should accurately describe the progress and performance of the MNCAH programme or health sector.
- Poor quality data lead to lack of trust in data and thus lack of use of information.



While data quality issues should ideally be addressed before the data are analysed and used, increased review and a culture of regular data use can lead to improved data quality!

Image from:

### FOR DISCUSSION

- What are the main data quality issues that exist within the MNCAH data coming from the routine health information system (RHIS)? How do these impact the ability to understand changes in MNCAH service delivery and use?
- What processes currently exist to investigate quality issues for RHIS data on MNCAH services?
- What are the main challenges that would need to be overcome to improve the quality of the RHIS data on MNCAH services?



The WHO data quality assurance (DQA) toolkit provides was developed to support countries in assessing and improving the quality of RHIS data.

**Module 1:** Framework and metrics

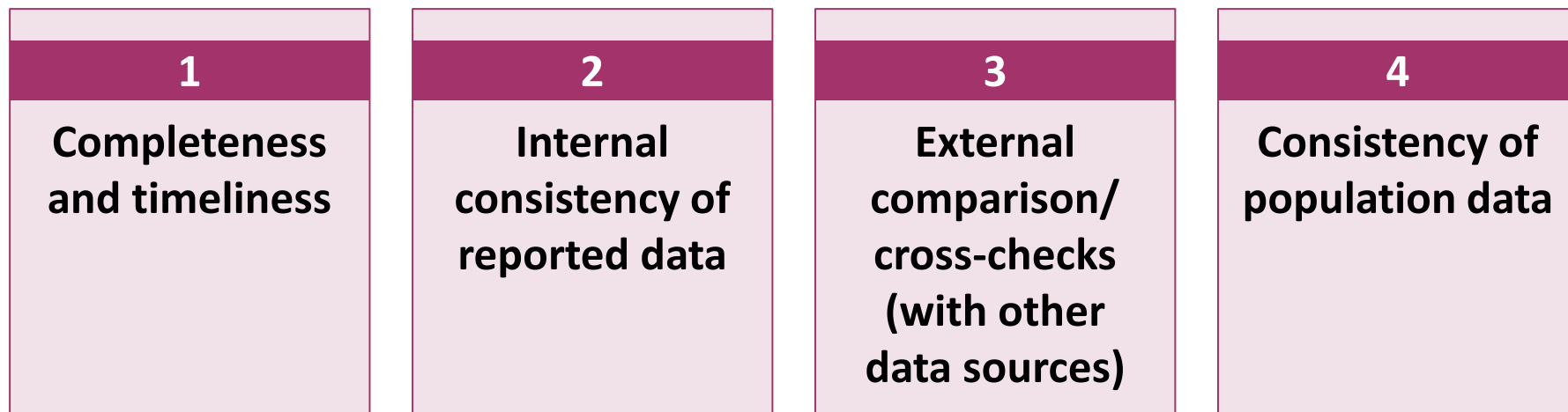
**Module 2:** Desk review

**Module 3:** Data verification and system assessment

**Module 4:** Data quality review for community health data – *Working document*



## What are the data quality problems associated with RHIS data?



# Four dimensions of data quality

## 1. Completeness and timeliness

“The **completeness of the data** is assessed by measuring whether **all the entities that are supposed to report actually do so**. This applies to health-facility reporting to districts and to district reporting to the regional or provincial levels. **Timeliness of data** is assessed by measuring whether the entities which **submitted reports did so before a predefined deadline**.”<sup>1</sup>

### Data quality metrics for completeness and timeliness

Completeness and timeliness of district reporting

Completeness and timeliness of facility reporting

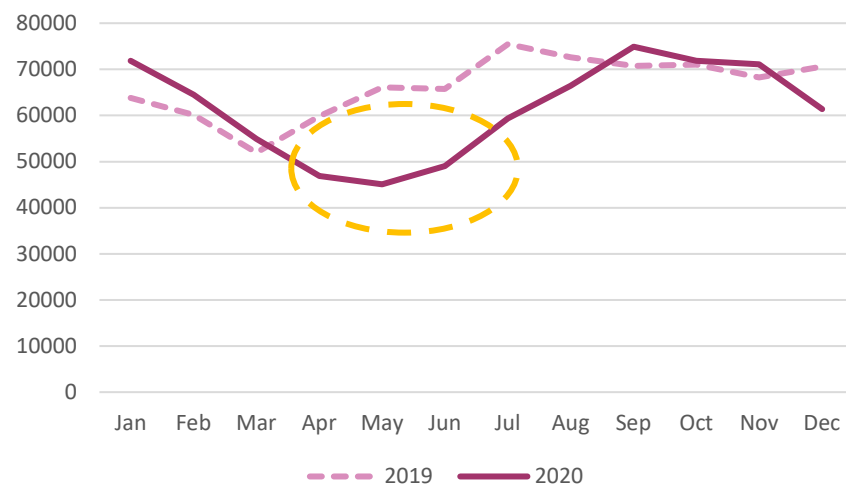
Completeness of indicator data (data element)

Consistency of reporting completeness

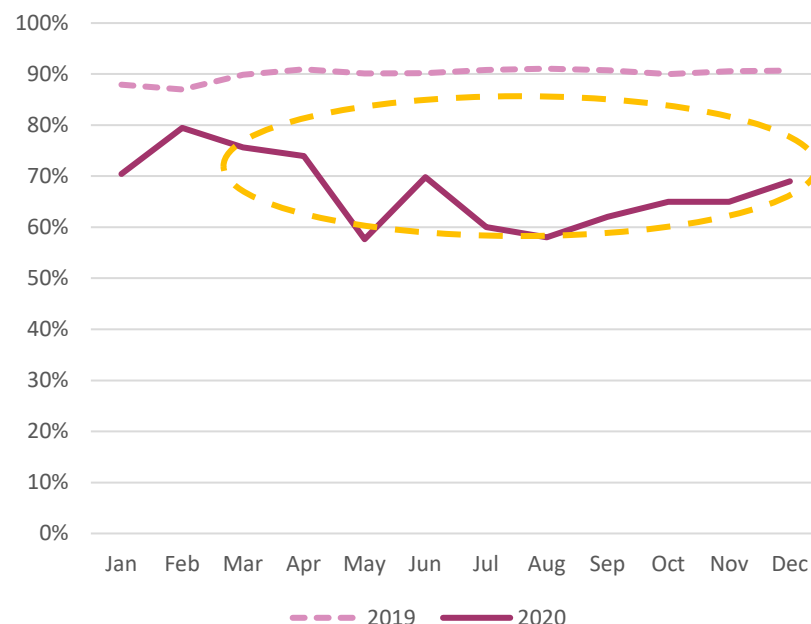


# It is important to consider reporting completeness when reviewing MNCAH indicators

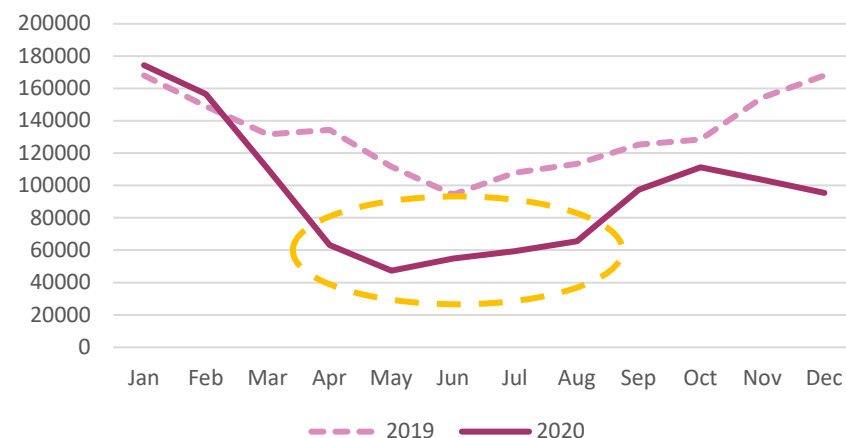
## Number of facility births



## Completeness of HMIS reporting (%)



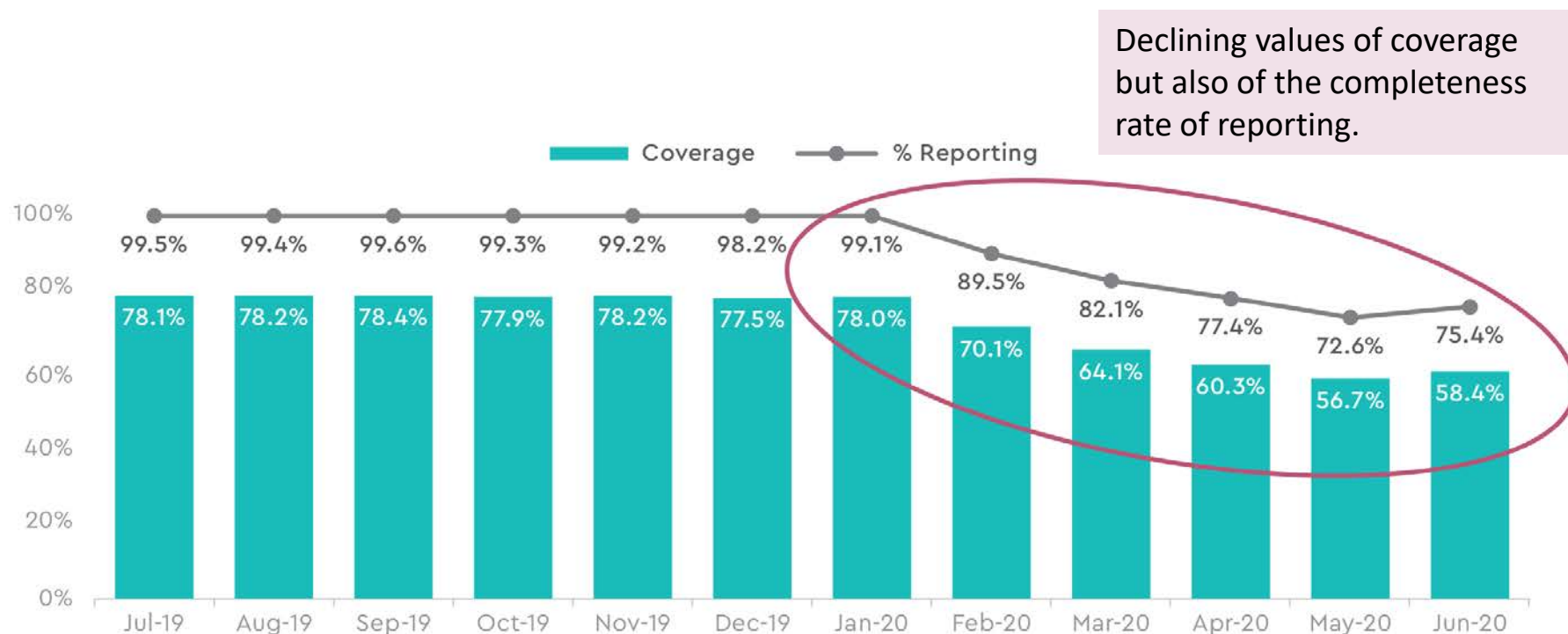
## Number of children under-five years with pneumonia treated in facilities



Note that the reported decrease in MNCAH interventions coincides with a decline in reporting completeness.

***Was the reported change in these indicators due to a change in service or due to incomplete reporting?***

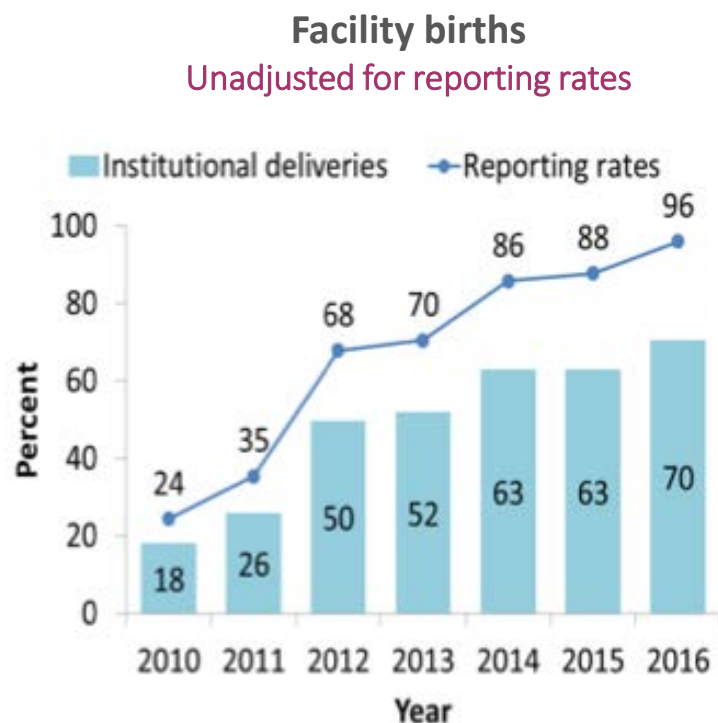
## Example of the effect of completeness rate on interpretation of coverage



In this case, if we don't realize that health facilities and districts are underreporting, this would most likely be misinterpreted as a decline in coverage.

## Example: Data quality analysis – Adjusting for reporting completeness

Incomplete data means that **not all data required for analysis are available**.

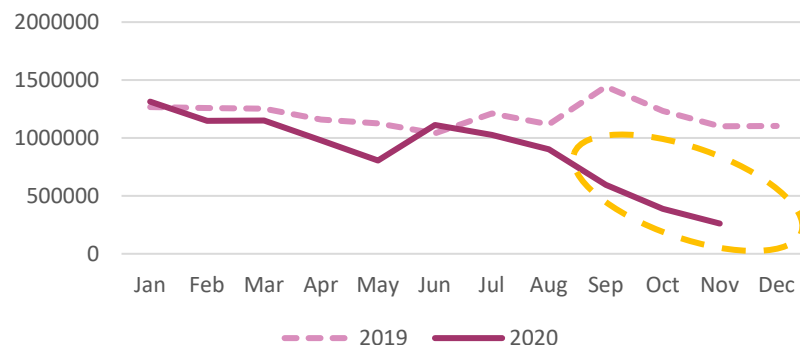


In this example, adjusting for reporting rates shows that the trend in increasing facility births was inaccurate, which can have implications if being used to inform MNCAH programme management and planning.

# Indicator performance may *appear* to “change” over time as reporting completeness improves

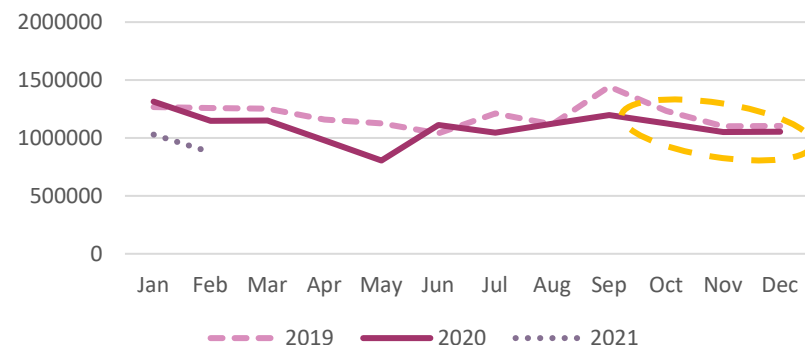
## Data as of November 2020

Total number of ANC visits

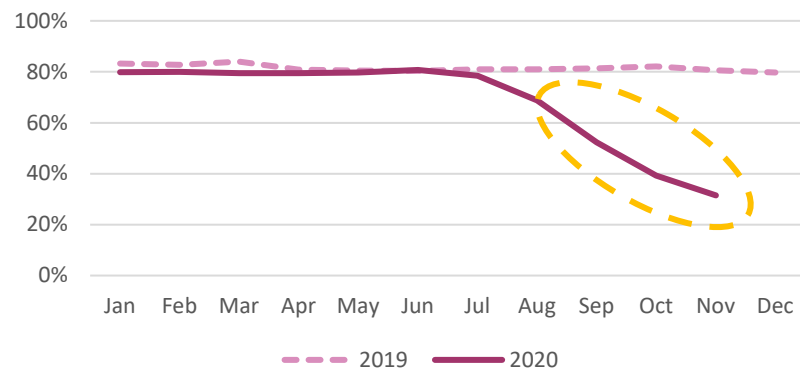


## Data as of February 2021

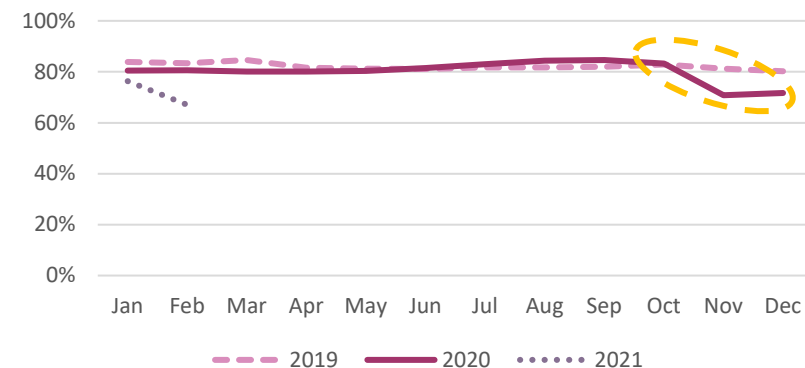
Total number of ANC visits



## HMIS monthly summary reporting rate (%)



## HMIS monthly summary reporting rate (%)



As more reports were submitted by facilities in February 2021, the reported decline in antenatal visits in September to November 2020 was shown to be due to the incomplete reporting in December 2021 and not a true change in service utilization. **It is good practice to note the timestamp of when the data were extracted with your analysis/visualization.**

## 2. Internal consistency of reported data

“**Internal consistency of the data** relates to the **coherence of the data being evaluated**.

Internal consistency metrics examine: 1) coherence between the same data items at different points in time; 2) coherence between related data items; and 3) comparison of data in source documents and in aggregated reports.”<sup>1</sup>

### Data quality metrics for internal consistency of reported data

Presence of outliers

Consistency over time

Consistency between indicators

Consistency of reported data and original records

## Internal consistency of reported data

Data quality metric: Presence of outliers

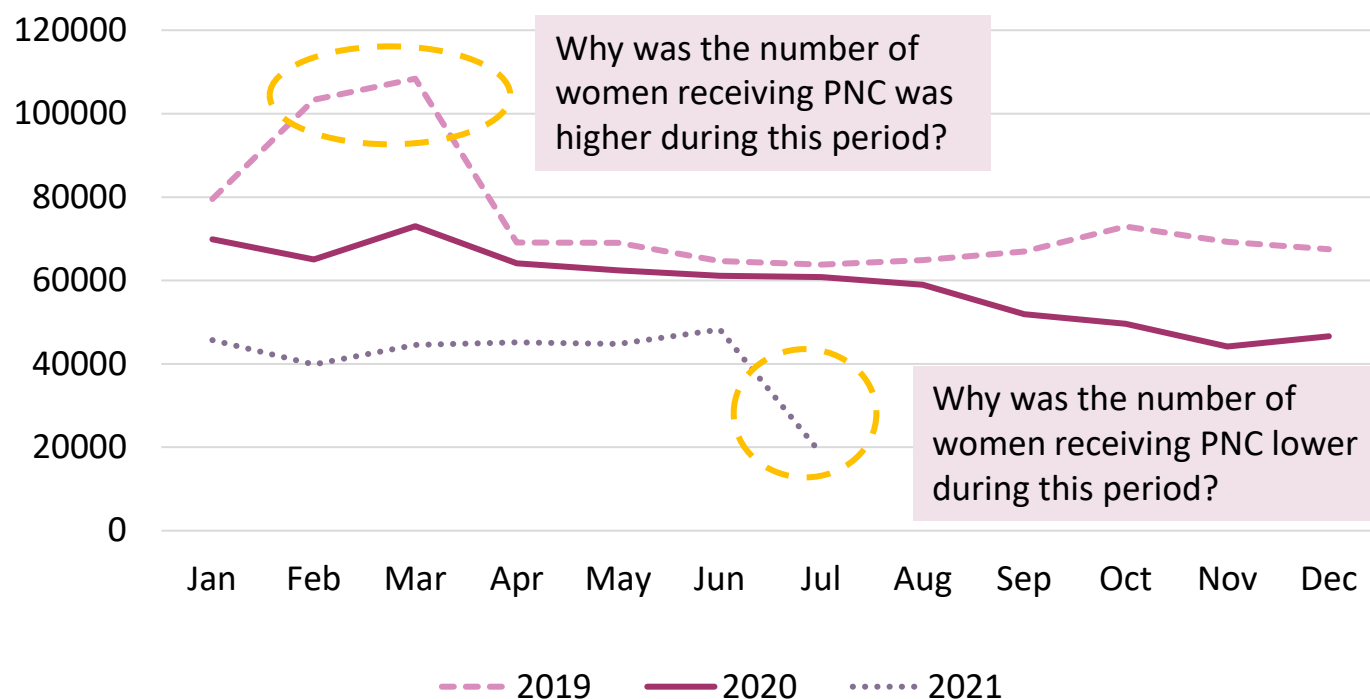
Outliers are values in the data that are **extreme or implausible relative to other values from the same reporting entity**.

- Outliers are identified by comparing individual values with other values in a series (e.g. monthly values over the course of a year) from the same reporting entity.
- Typically a standard is used to identify a value as an outlier in comparison to other values in the series (e.g. multiples of the standard deviation away from the mean (or median) of the values in the series).
- Not all extreme values are data quality problems since gaps or sudden changes in service delivery are relatively common (e.g. stock-outs, intensification campaigns, etc.). Outliers should be investigated to determine the underlying cause.
- If extreme values are determined to derive from data quality problems (e.g. key-punch errors) they should be corrected in the database according to the standard processes in place in countries.

# Internal consistency of reported data

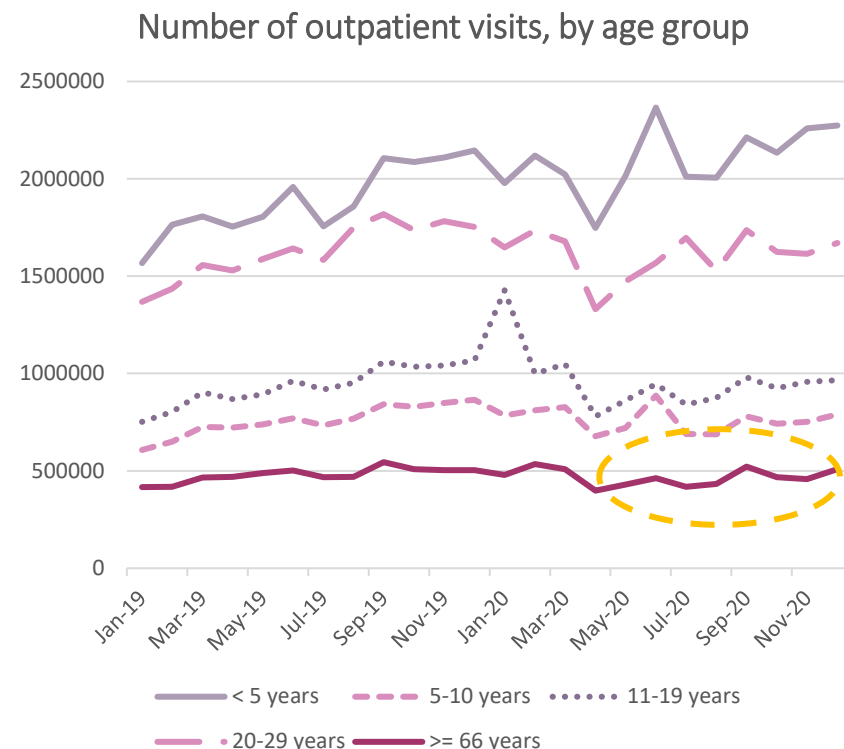
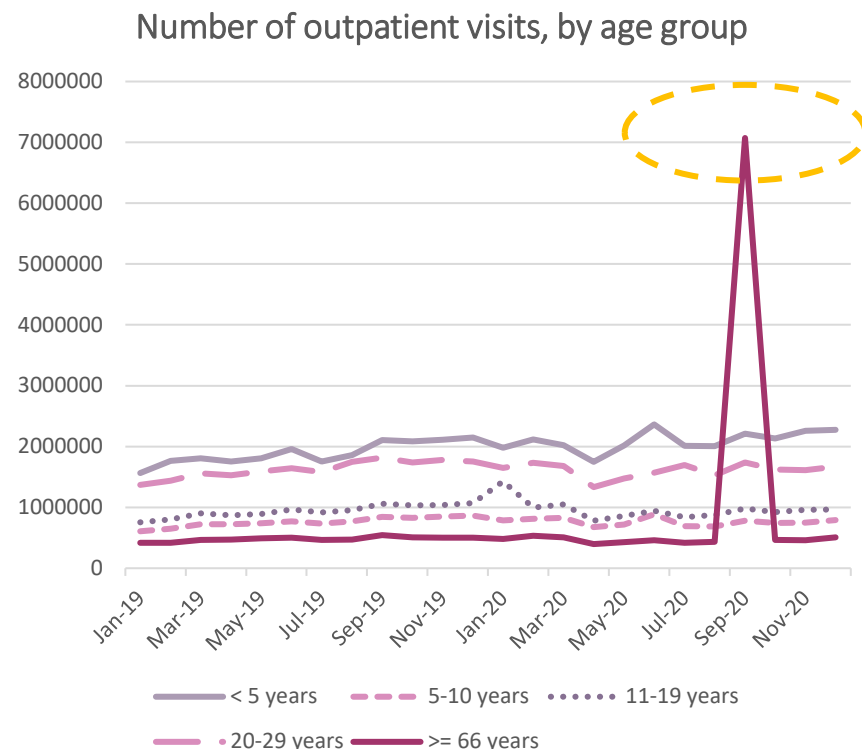
## Outliers

### Example: Number of women who received postnatal care within 1 day of delivery by year



Reviewing the data across multiple time periods helps to identify values that are notably higher or lower than in previous reporting periods. Not all abnormal values are necessarily incorrect (i.e. data quality issue), but should be noted and for further investigation to determine the reason.

# Outlier due to transcription error – easily corrected



Outpatient visits among individuals  
>= 66 years of age in September  
2020 was an obvious outlier

Verify original  
data

Corrected data in the RHIS

Since data quality problems can become masked in the aggregate data it helpful to “drill down” to lower levels (i.e. into the disaggregated data) to better pinpoint source of the error. If the data entry error can be easily verified, you can correct the data.



# Internal consistency of reported data

Data quality metric: Consistency over time

**Data should be relatively consistent over time (i.e. from month to month, or from year to year)**

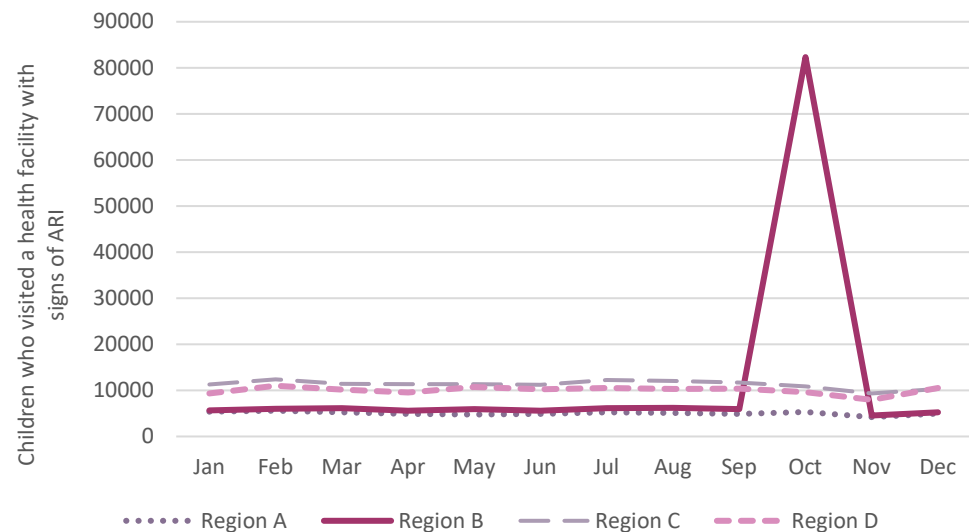
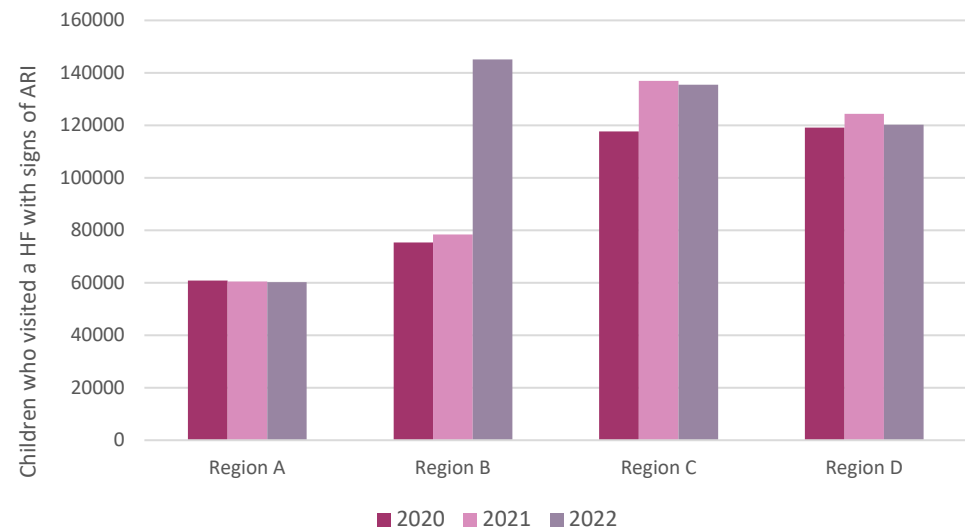
- While we expect the values of some indicators to rise or fall over time, the changes should usually tend to be gradual, as interventions take effect and coverage rises, or disease incidence declines.
- Sudden changes in indicator values over time can indicate a data quality problem.

**Two ways to identify data quality problems in trends over time:**

- Visual inspection of trend using graphs.
- Comparing the current value with historical values by calculating the ratio.
  - Indicators or programmes with expected change (increases or decreases in indicators) - comparison of current year to the value predicted from the trend in the 3 preceding years.
  - Indicators or programs expected to remain constant - comparison of current year to the average of 3 preceding years.
- Discrepancies are usually indicative of data quality problems and need to be investigated to determine the actual cause.

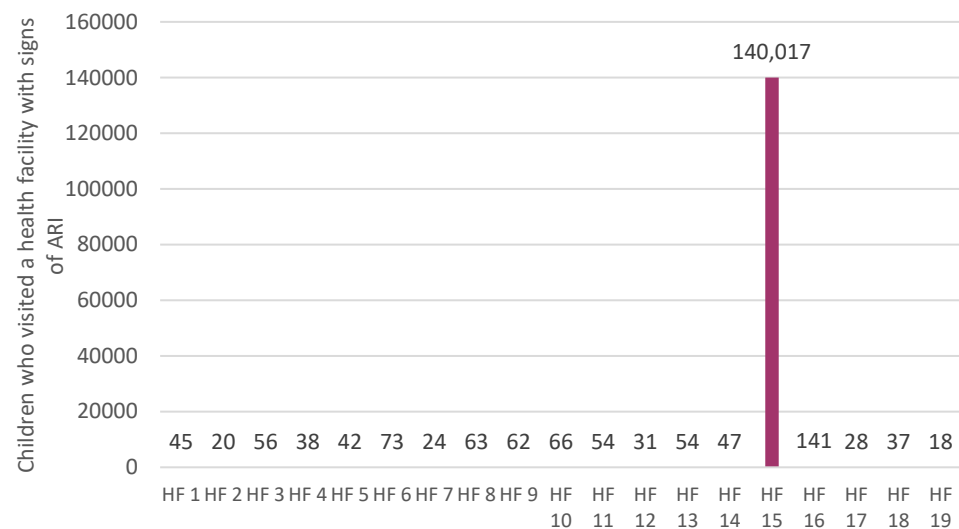
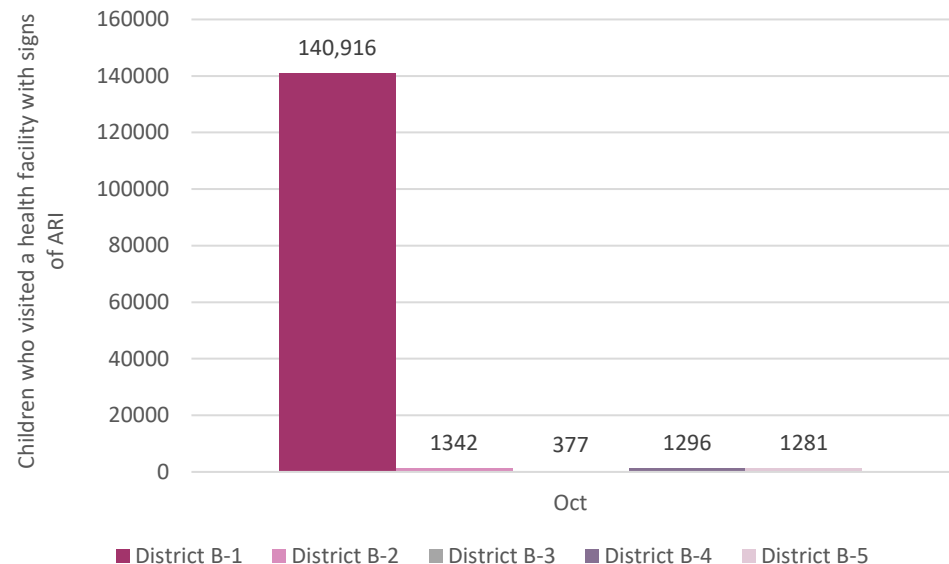
## Example: Visual examination of reviewing consistency over time

- Examine the graphic depicting the Number of children who visited a health facility for an acute respiratory infection, by region and year (2020-2022).
- Notice the sharp increase in Region B in 2022.
- We further examine the monthly trend in the indicator for 2022.
- The outlier value appears to occur in October in Region B.
- To further understand the issue, we need to “drill down” into the data reported from Region B.



## Example: Visual examination of reviewing consistency over time (cont.)

- In Region B, District 1 has an implausible value relative to the other districts in the region.
- Drilling down into District 1 shows where the anomalous value originates (Facility 15).
- To correct the value, a manager may need to follow up with: a) the district to which it was reported in case the monthly report on file has the correct value, or, B) the facility to determine the correct value.



## Internal consistency of reported data

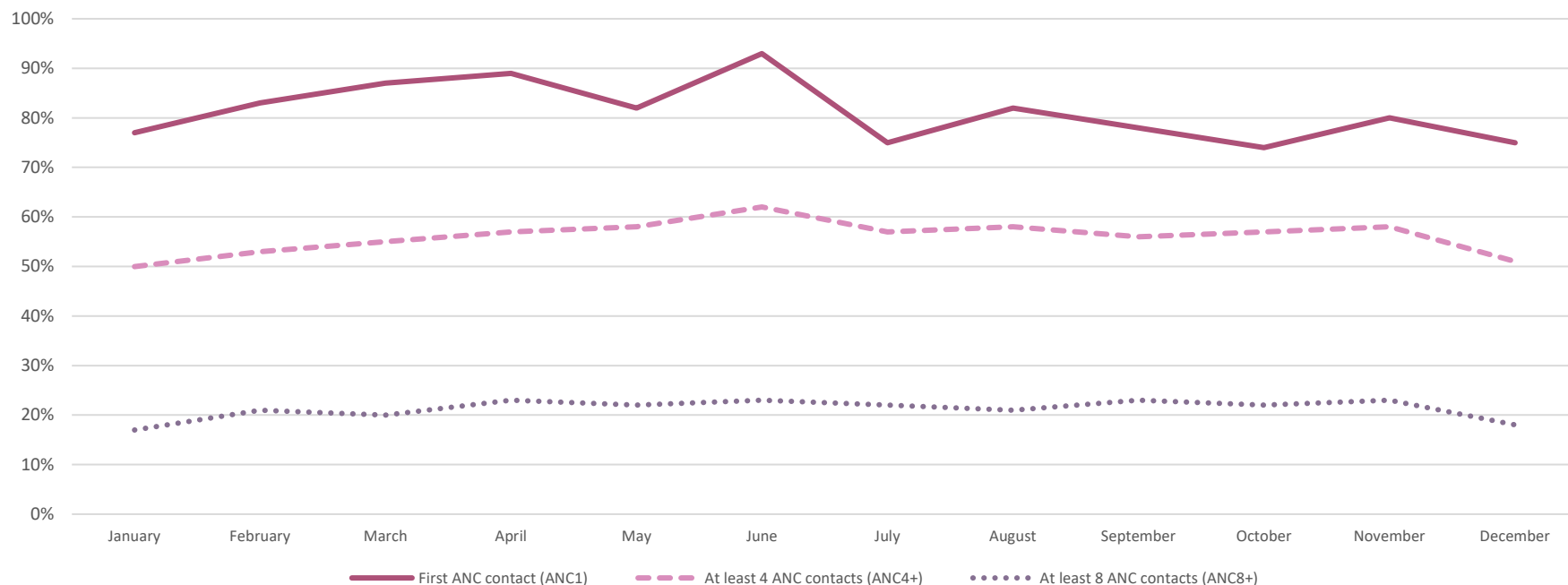
Data quality metric: Consistency between related indicators

---

- Indicators which have a predictable relationship can be examined together to determine if the relationship holds in the reported data.
- Some predictable relationships are that that one is always bigger than the other (or smaller) or that they should have roughly the same value.
- When the predicted relationship does not hold in the reported data a data quality problem is likely at fault.
  - For example, you expect that first ANC visit should always be bigger than fourth ANC visit (ANC1 > ANC4+).

## Example: Visual examination of consistency between related indicators

Proportion of pregnant women with antenatal care visits(s) in health facilities each month, by 1<sup>st</sup>, 4<sup>th</sup>, and 8<sup>th</sup> contact, 2022  
(Mock data)



In this example, you note that the reported proportion of women receiving their first ANC contact in facilities is higher than the proportion receiving their fourth or eighth contact, which follows the expected pattern for these indicators.

In June, the increase in ANC4+ mirrors the increase in ANC1.

### 3. Consistency with external data sources: External comparison/cross-checks

---

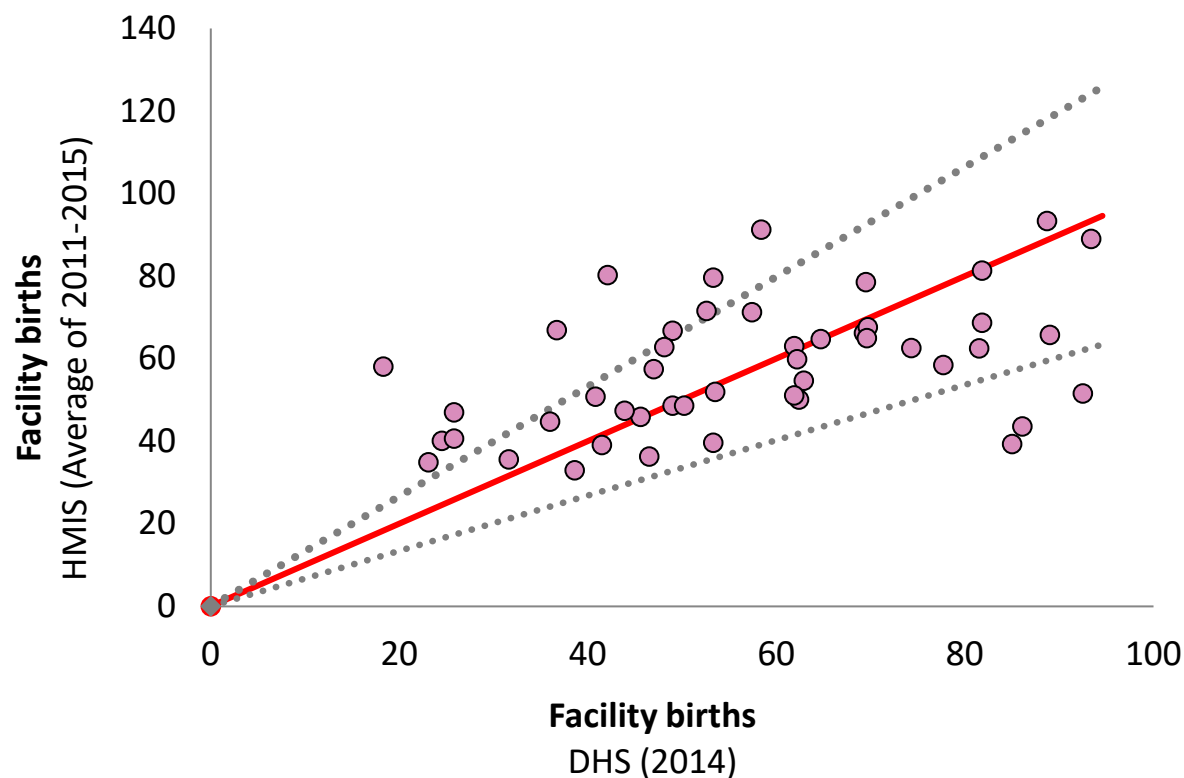
**External comparison** refers to the assessment of the “**level of agreement between two sources of data** measuring the same health indicator”.<sup>1</sup>

#### Data quality metrics for external comparison of data sources

Consistency between routine data from the HMIS and data from population-based surveys (or other alternative data sources)

---

## Example: Analysis of data quality - Consistency with external data source



### Facility births:

Consistency between demographic health survey (DHS) 2014 and health management information system (HMIS) data (average of 2011 – 2015).

The percentage of deliveries in institutions reported the HMIS during 2011-2015 is inconsistent with DHS in 13 counties.

Adapted content from original slide from:

## Example: Consistency with external data source

Comparison of HMIS and survey coverage rates for ANC1. Any difference between the two sources that is  $\geq 33\%$  is highlighted in red.

District	Facility Coverage Rate	Survey Coverage Rate	Ratio of Facility to Survey Rates	% Difference the two rates
A	1.05	0.95	1.11	11%
B	0.93	0.98	0.95	5%
C	1.39	0.90	1.54	54%
D	1.38	0.92	1.50	50%
E	0.76	0.95	0.80	20%
National	1.10	0.94	1.17	17%



**Consistency of population data** “involves determining the **adequacy of the population data used** in evaluating the performance of health indicators”.<sup>1</sup>

#### Data quality metrics for consistency of population data

Consistency of population trends and comparison of related population estimates (i.e. between the population data used for calculating health service coverage and other sources of population estimates)

---

## Example 1 – External consistency of population denominators

Comparison of national and subnational administrative unit ratios of official government live birth estimates. Administrative units with difference  $\geq \pm 10\%$  are highlighted in red.

District	National Bureau of Statistics (NBS) estimate of surviving infants	Expanded programme on immunization (EPI) estimate of surviving infants	Ratio of NBS to EPI estimates
District 1	12,216	16,248	0.75
District 2	10,824	12,612	0.86
District 3	7,393	8,988	0.82
District 4	5,884	6,204	0.95
District 5	4,567	4,812	0.95
.....	.....	.....	.....
.....	.....	....	.....
National	1,553,306	1,678,858	0.93

## Example 2: Examining consistency in population trends using a chart

\*Denominator data quality



EDIT

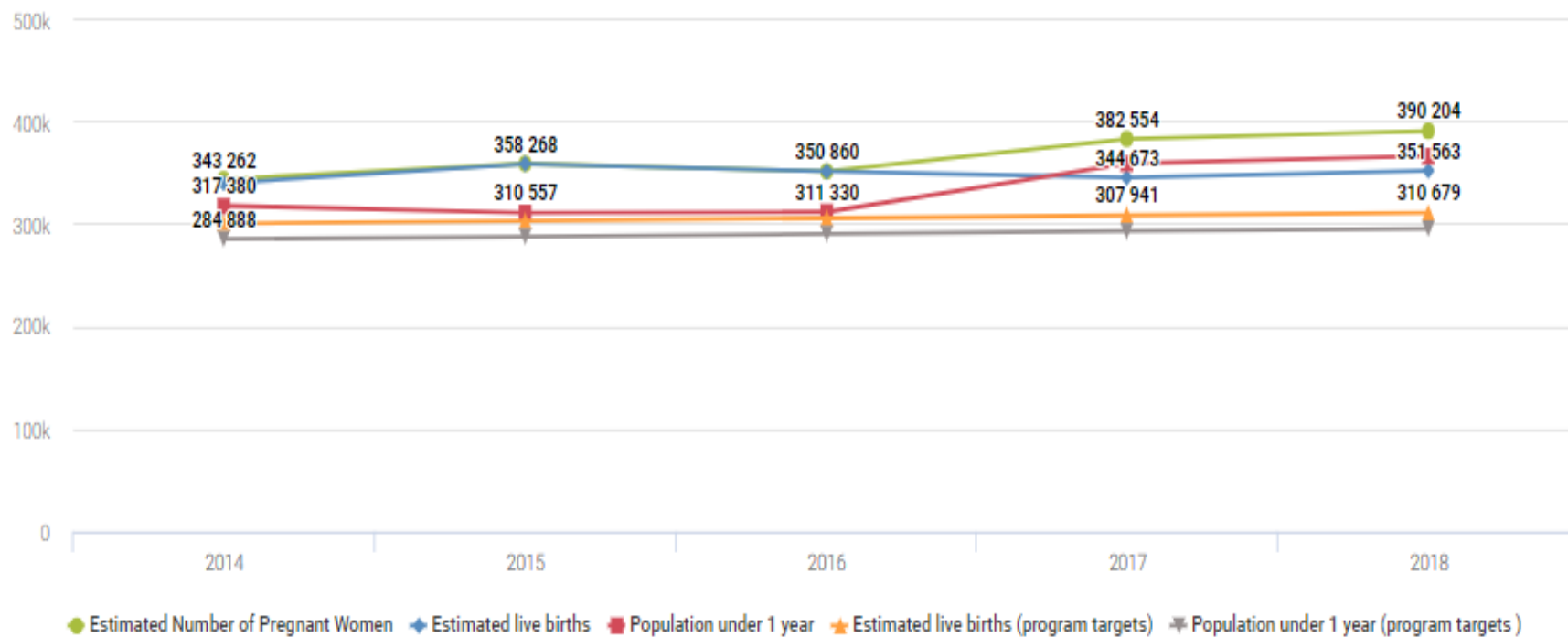
SHARE

FILTER

Denominators 01: 5 year trend in estimated targets



National



# WHO data quality tool for DHIS2

- Specialized tools are available to assist with conducting data quality analyses.
- The [WHO Data Quality Tool for DHIS2](#) and can be installed on the local instance of a dhis2 platform.
- Alternatively, these analyses can also be done in commonly available software, such as Microsoft Excel. Ask the relevant officer to help set up and conduct the analyses with HMIS data.



### FOR DISCUSSION

- What do trends in completeness and timeliness of reporting tell you? Why are they important?
- What does the trend in an indicator over time tell you about the indicator?
  - What is the expected trend?
  - How can comparing a recent trend to a historical trend inform you about data quality?
- Name some examples of pairs of MNCAH indicators that have a predictable relationship.
  - How can this relationship be used to consider data quality?
  - If the expected relationship between two indicators is not found what are some plausible reasons to explain the observed relationship?
- What are some of the reasons for discrepancies between routine and survey indicator values, in addition to data quality?
- What can impact the data quality of population data?



MNCAH managers review RHIS data regularly to assess programme performance. Prior to reviewing analysed data, the data should be checked for quality, though often through review of analysed data, data quality issues are identified.

## In reviewing MNCAH RHIS data, managers may ask:

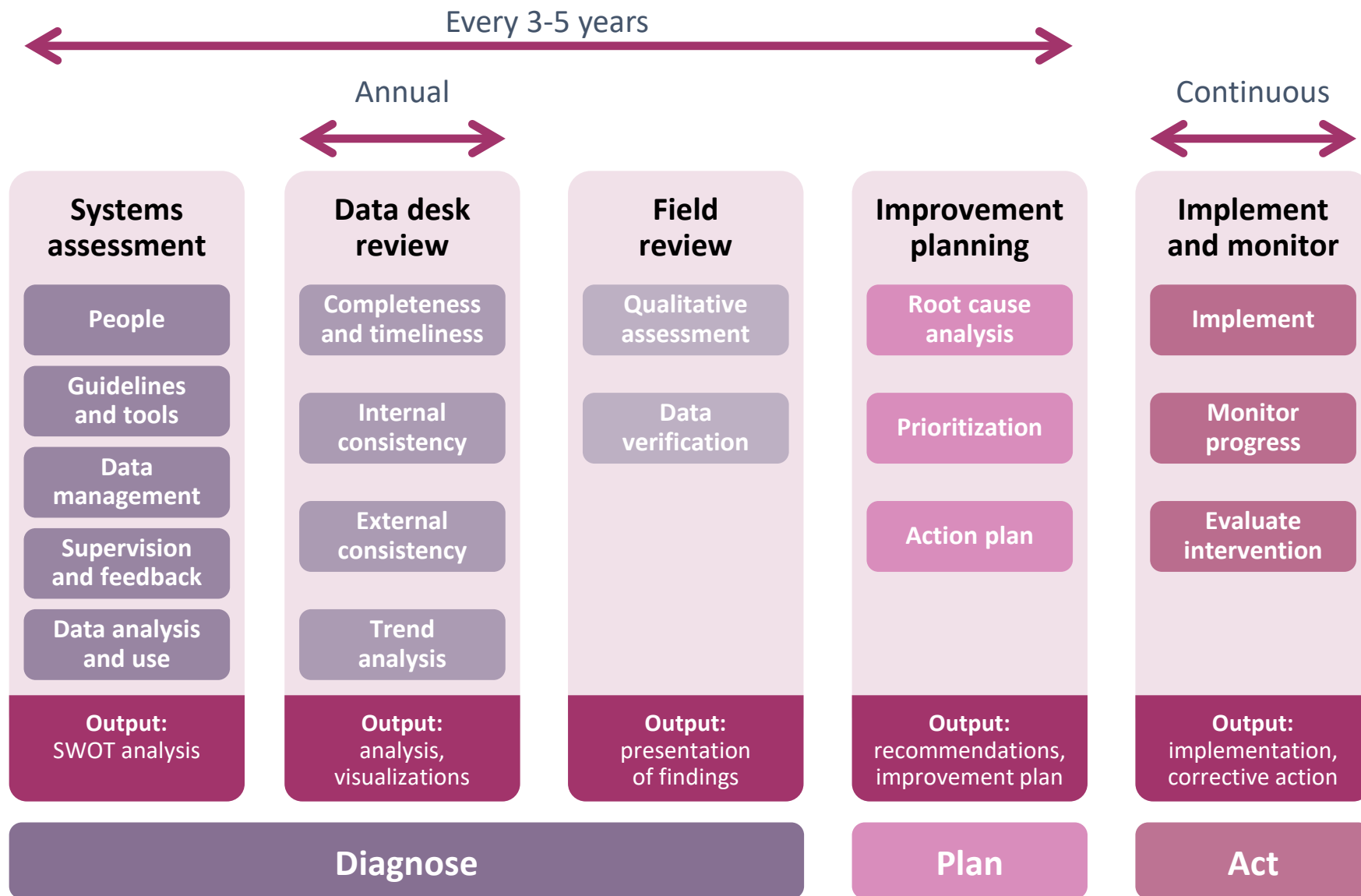
**Are changes in reported utilization of MNCAH services due to actual changes in service utilization?  
Are the changes in reported cases of illnesses or deaths due to real increases/decreases in these outcomes?**

- Was there a marked increase or decrease in reported utilization of any MNCAH service?
- Was there a health worker strike, commodity stock out or another reason why facilities could not provide a specific MNCAH service?
- Was there an intervention campaign or intensification effort that could have increased the delivery of any MNCAH services?
- Is there a seasonal event that could impact access to health facilities (e.g., roads blocked due to flooding) or expected increases or decreases in an illness (e.g., seasonality of malaria)?

**Could the changes in reported utilization of MNCAH services be due to data quality issues?**

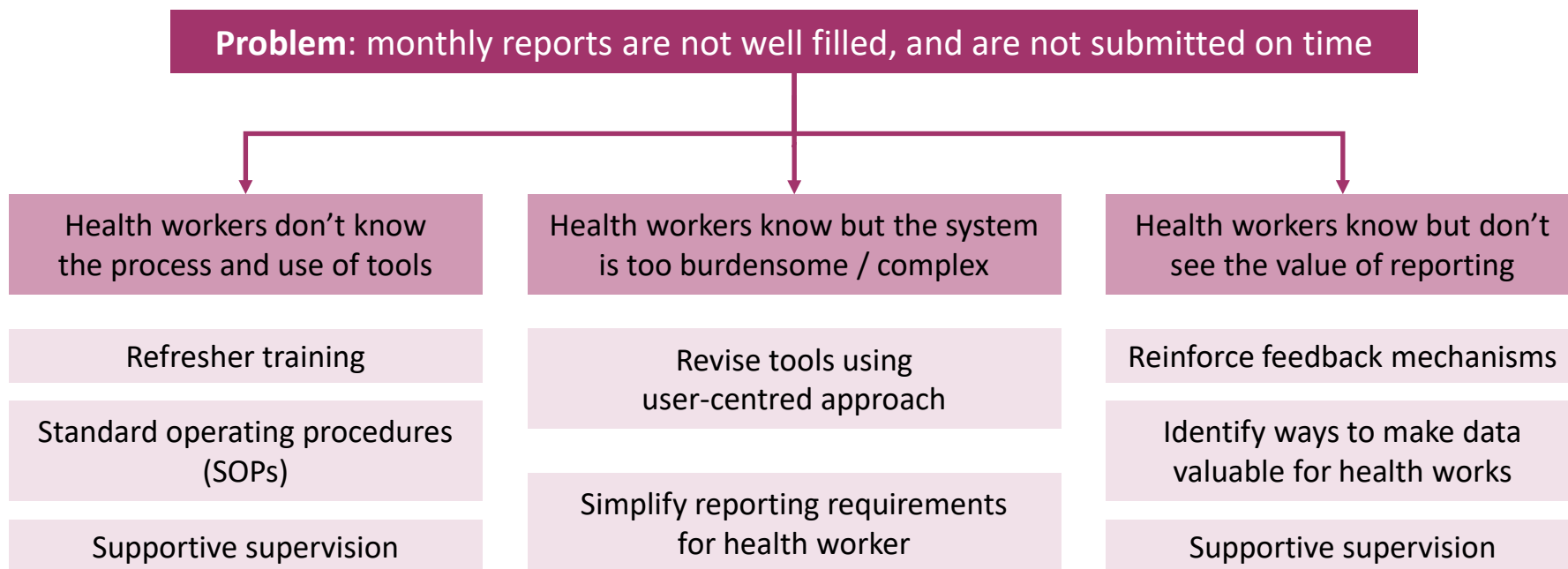
- What is the reporting completeness? Have all facilities, especially high-volume sites, submitted reports?
- Are the same number of facilities reporting each period? Are the same facilities reporting each period?
- Was there a facility (or facilities) that reported extreme values (i.e., outliers)? If we remove those facilities, does the trend look the same?

# Assessing data quality - When and how?



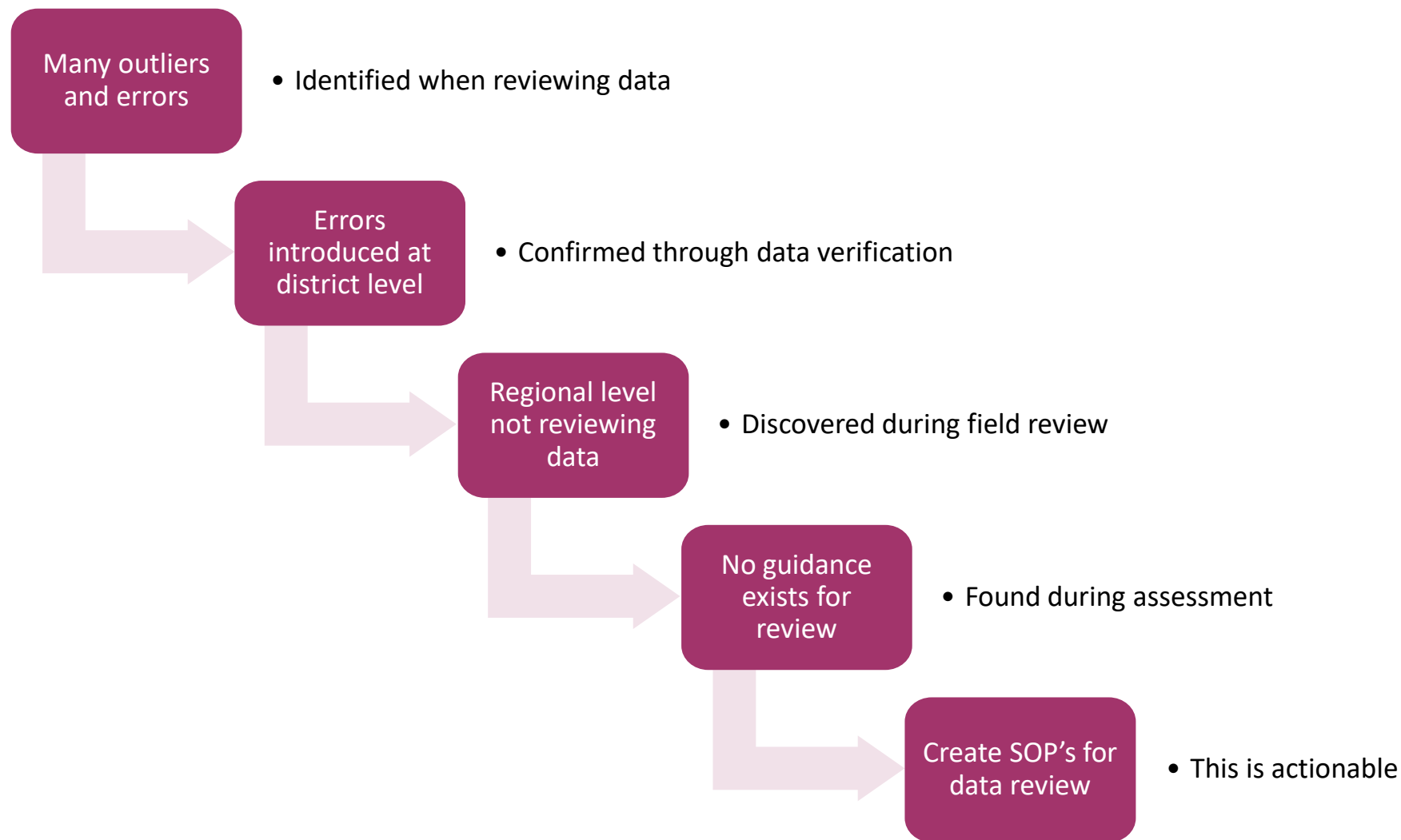


# Doing root cause analysis is a critical step in improving data quality

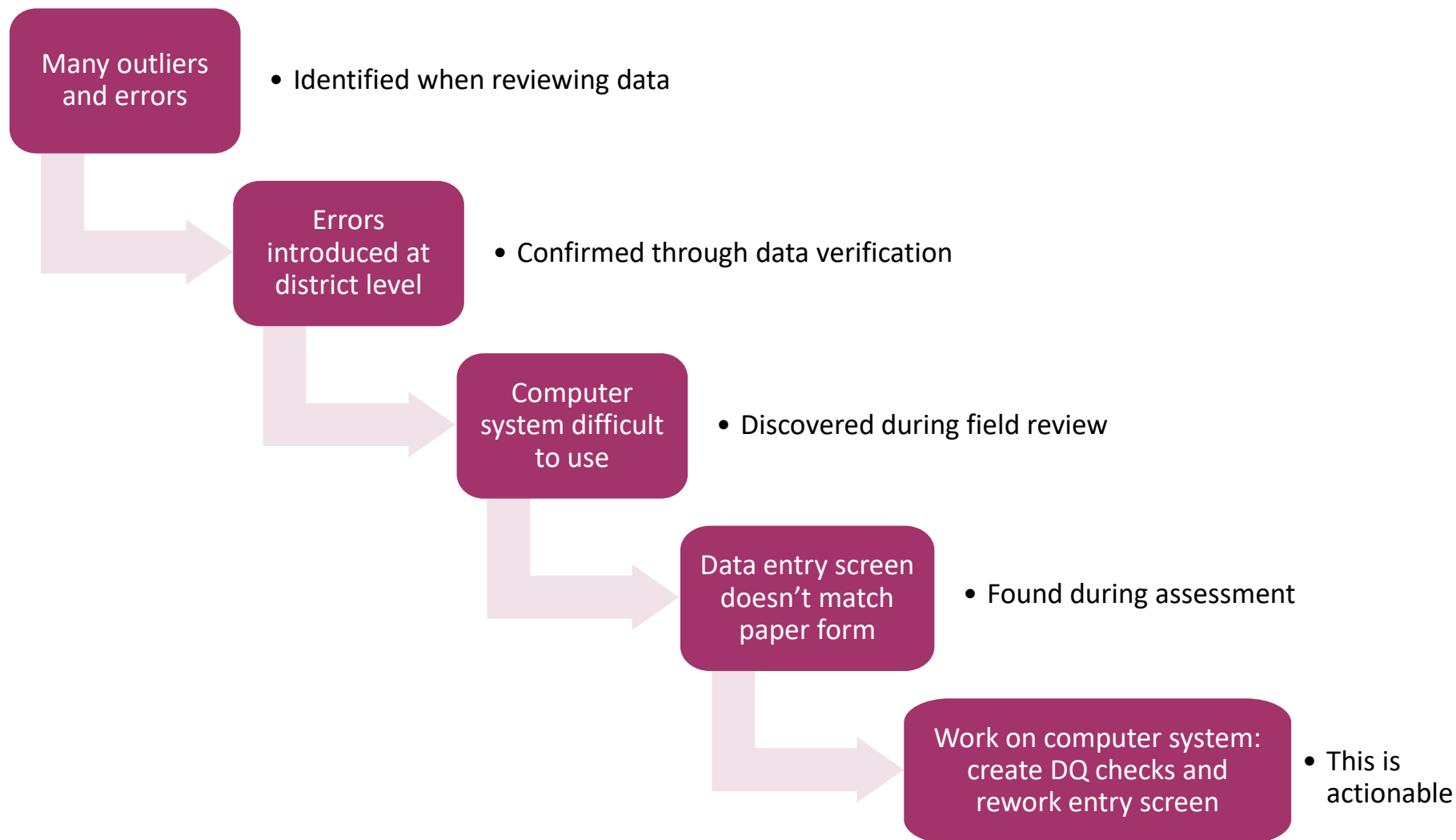


- Now that problems are detected, we need to understand their root causes: why are we seeing these problems?
- Recommendations that address the root causes will be more actionable.
- Prioritize feasible and potentially impactful recommendations.
- Develop an action plan: timebound, clear responsibilities, budgets, how to measure success?

# Management process for addressing data quality issue: Example 1



## Management process for addressing data quality issue: Example 2



- Complete exercises under **Data quality** of MNCAH data in *Companion exercises to strengthen analysis and use of health facility data for MNCAH.*
- Review responses in plenary.