



**MATGRACE
CONSULTING LLC**

DATA MANAGEMENT AND ANALYSIS

Training

Discover Our Expertise

Matgrace Consulting LLC exists to solve the critical issues facing our clients, both large and small. Our unique approach is not only what differentiates us, but also what makes us successful.

We provide a broad range of services and solutions to help organizations facilitate change, achieve their vision and optimize performance and productivity.



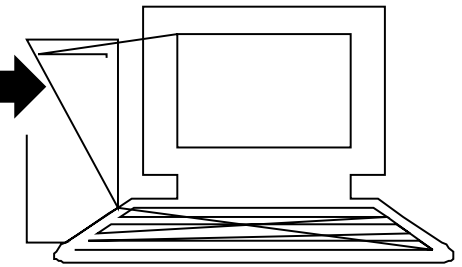
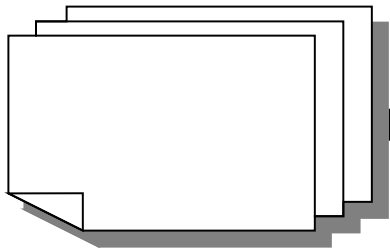
Introduction

INTRODUCTION

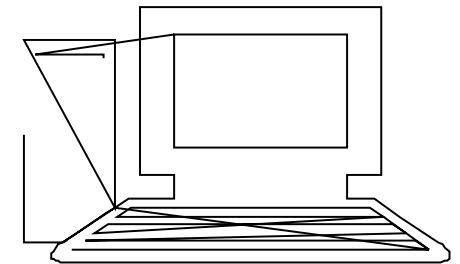
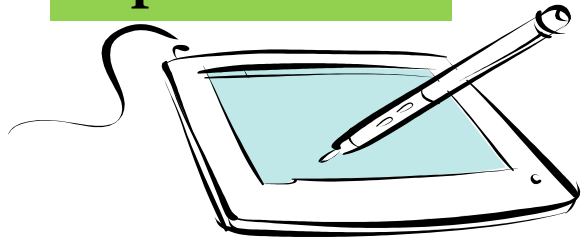
- Data management is an important component in M&E and deserves extra attention and diligence
- M&E teams should invest a significant part of their time and effort in data management
- M&E teams should understand the basic concepts of data management
- Data management policies and procedures should be clearly defined

DATA CAPTURE

Paper based data



Paperless data



Personal Digital
Assistant (PDA)

Database



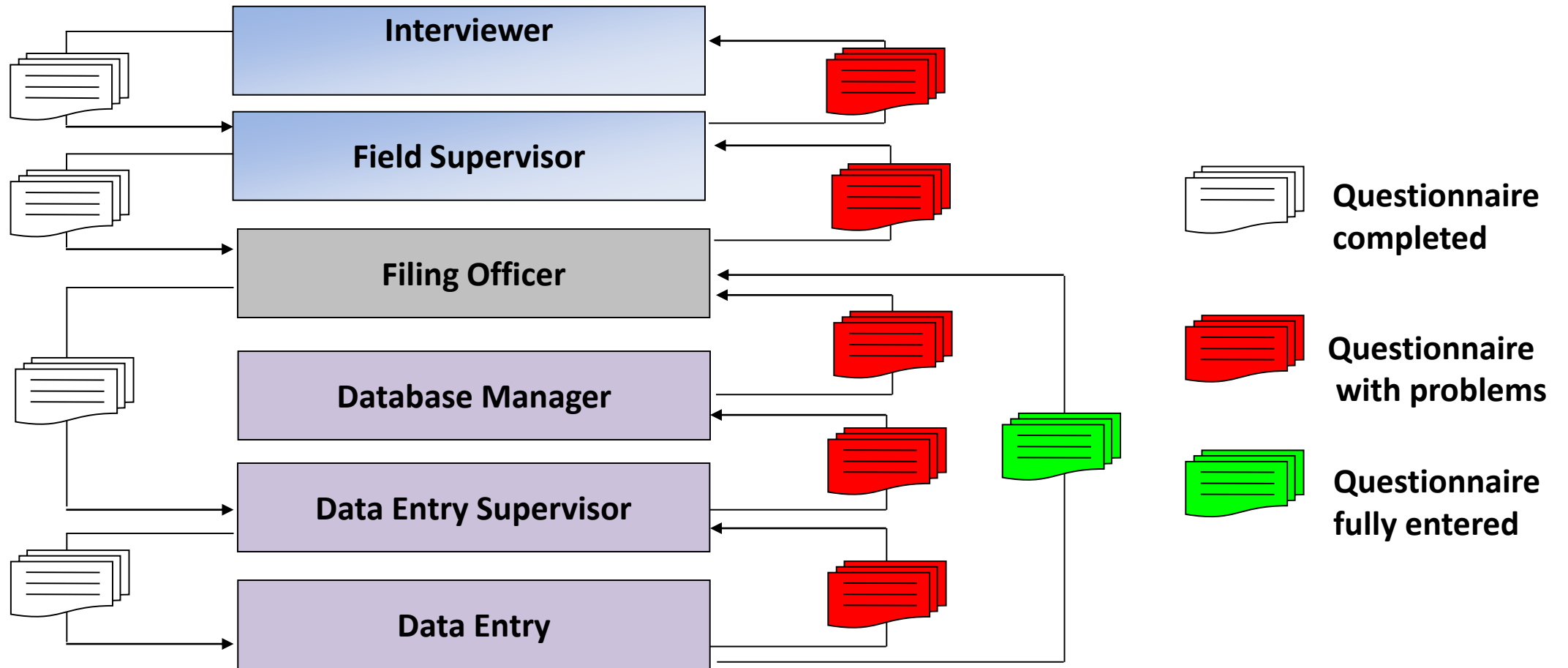
Data Capture, *cont.*

- Plan data capture carefully
- Decide on which software you will be using
- Define your database structure (tables or data files)
- Develop data entry screen (should be user-friendly and include check for plausible values)
- Make provision for double- entry

Set Quality Target

Aspect	Critical level
Consistency/validation	99%
Error (range check)	100%
Double entry	100%

Form/Questionnaire Flow

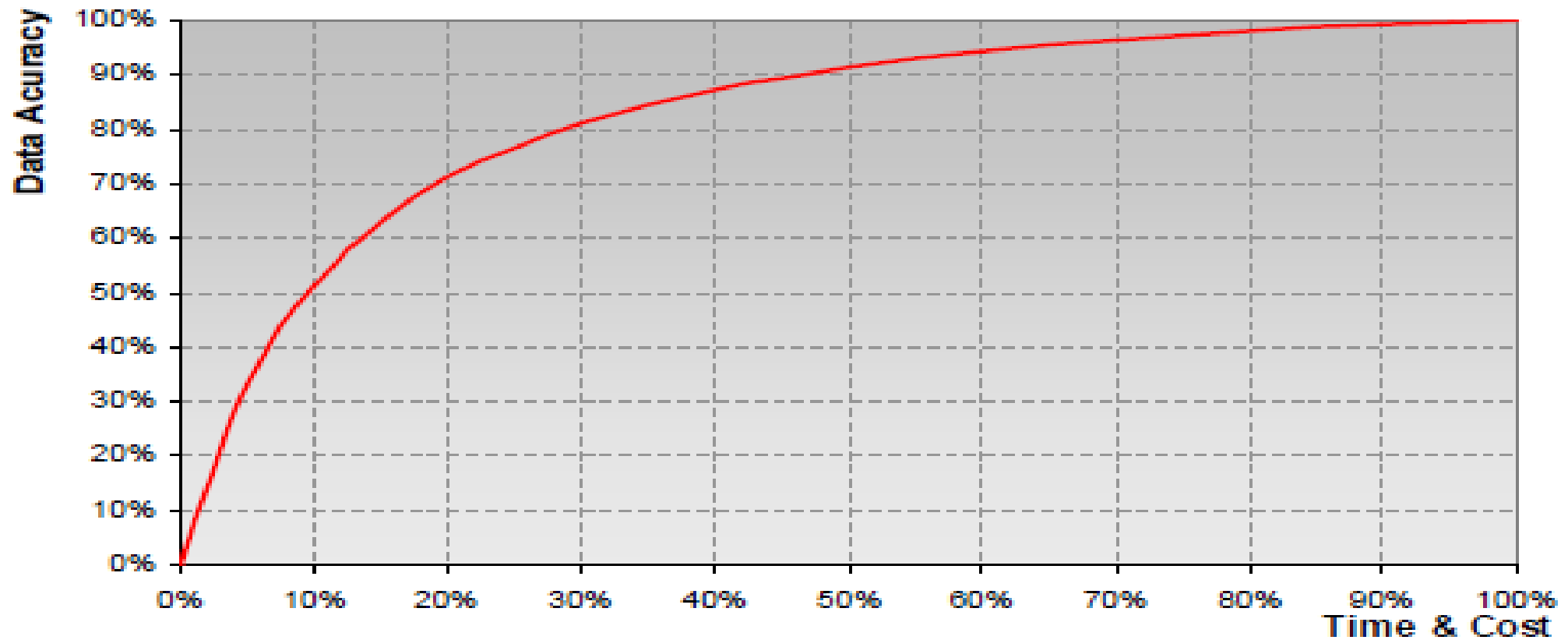


Data Cleaning

- Check completeness of the data
- Check consistency- compare variables
- Check plausibility (value with acceptable range)
- Check for duplicates
- Check for outliers (run basic freq, mean)

Data Cleaning, *cont.*

Data Cleaning Trade-off curve



Data Security

- Access to data should be restricted (Password)
- Final analytical data should be anonymous
- Make sure to do a regular data backup-daily- weekly-monthly...
- If possible store a copy of your back up off-site

Other Aspects to Consider

- **Data Ownership:** Who has the legal rights to the data and who retains the data
- **Data Retention:** Length of time one needs to keep the project data
- **Data Sharing:** How project data and results are disseminated, and when data should not be shared

Session Objectives

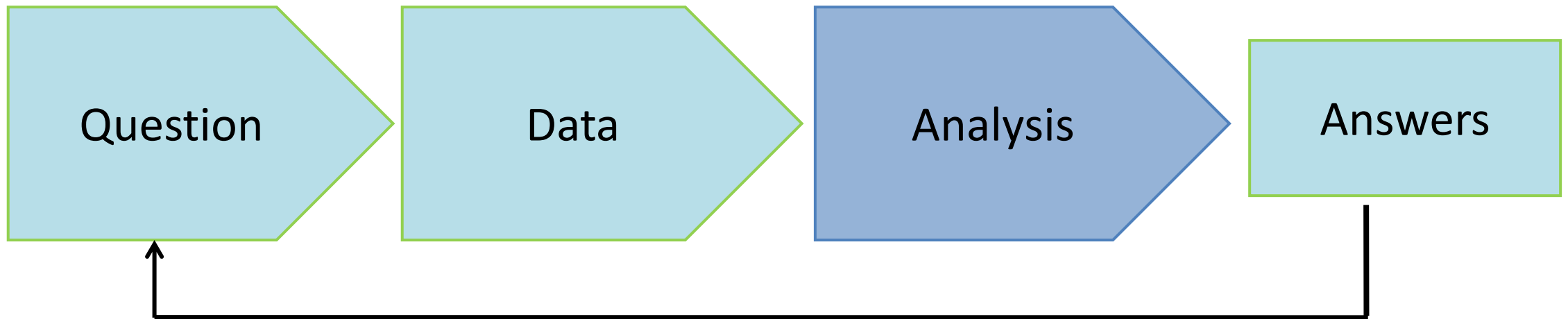
1. Strengthen knowledge of terminology used in data analysis and interpretation
2. Strengthen skills in data analysis and interpretation
3. Improve capacity to summarize data
4. Strengthen effective communication methods

What is Data Analysis?

- The process of **understanding** and **explaining** what findings actually mean. Turning raw data into useful information
- Provide answers to questions being asked at a program site or research questions being studied
- The greatest amount and best quality data mean nothing if not properly analyzed, or, if not analyzed at all

What is Data Analysis?, *cont.*

Analysis is looking at the data in light of the questions you need to answer



How would you analyze data to determine, “Is my program meeting it’s objectives?”

Is Our Program on Track?

- **Analysis:** Compare program targets and actual program performance to learn how far you are from target
- **Interpretation:** Why you have or have not achieved the target and what this means for your program
- May require more information

Examples of Analysis

Compare actual performance against targets

Indicator	Progress (6/12/13)	Target (1/30/14)
Number of persons trained on case management	15	100

Comparing current performance to prior year

Indicator	2011	2012
No. of LLIN distributed	50,000	167,000

Compare performance between sites or groups

Indicator	District A	District B
Number of fever cases tested for malaria by clinics	3,500	8,000

Statistical Measures

- **Measure of central tendency**
 - Mean
 - Median
 - Mode
- **Measure of variation**
 - Range
 - Variance and standard deviation
 - Interquartile range
 - Proportion, Percentage
- **Ratio, Rate**

Mean

Sum of the values divided by the number of cases.
Also called *average*

$$\bar{y} = \frac{\sum y_i}{n}$$

Very sensitive to variation

Average number of confirmed malaria cases per month

Month	Cases 2008
Jan	30
Feb	45
Mar	38
April	41
May	37
Jun	40
Jul	70
Aug	270
Sep	280
Oct	200
Nov	100
Dec	29

Total number of cases

$$\sum y_i = 1,180$$

Number of observations

$$n = 12$$

Mean number of cases

$$\bar{y} = \frac{1,180}{12} = 98.2$$

Median

- Represents the middle of the ordered sample data
- For odd sample size, the median is the middle value
- For even, the median is the midpoint/mean of the two middle values

Not sensitive to variation

Median number of confirmed malaria cases

Month	Cases 2008	Cases 2009
Dec	29	24
Jan	30	29
May	37	32
Mar	38	35
Jun	40	39
April	41	39
Feb	45	42
Jul	70	65
Nov	100	80
Oct	200	150
Aug	270	200
Sep	280	-

Median for 2008

$$\text{median} = \frac{41 + 45}{2} = 43$$

Median for 2009

$$\text{median} = 39$$

Mode

- Value that occurs most frequently
- It is the least useful (and least used) of the three measures of central tendency

Mode number of confirmed malaria cases

Month	Cases 2008	Cases 2009
Dec	29	24
Jan	30	29
May	37	32
Mar	38	35
Jun	40	39
April	41	39
Feb	45	42
Jul	70	65
Nov	100	80
Oct	200	150
Aug	270	200
Sep	280	-

Mode for 2008

mode = none

Mode for 2009

mode = 39

Practice Calculations

- What is the mode, mean and median parasitemia for the following set of observations?

1.5, 1.8, 2.5, 4.1, 8.3, 1.2, 1.9, 0.6

- Answers:

- Mean = 2.74
- Median = 1.85
- Mode = none

- Would you use Mean or Median?
- Answer: Median
- Use Median when you have a large variation between high and low numbers
- Use Mean when there is not a huge variation between the values

Ratio

- Comparison of two numbers
- Expressed as:
 - a to b, a per b, a:b
 - 2 household members per (one) mosquito net, a ratio of 3:1
- All individuals included in the numerator are not necessarily included in the denominator

Proportion

- A ratio in which all individuals in the numerator are also in the denominator
- Example: If a clinic has 12 female clients and 8 males clients, then the proportion of male clients is $8/20$ or $2/5$

M M M M M
M M M

F F F F F
F F F F F
F F

Percentage

- A way to express a proportion
- Proportion multiplied by 100
 - Example: Males comprise $\frac{2}{5}$ of the clients or, 40% of the clients are male (0.40×100)

Important to know: What is the whole? An orange? An apple?
All clients? All clients on with a fever?

Why do we want to know the percentage?

- Helps us standardize so that we are able to compare data across facilities, regions, countries
- Better conceptualize what needs to be done
 - Percentage helps us to track progress on our targets

Rate

(Under five mortality rate)

Probability of Dying Under Age Five per 1,000 Live Births

- A quantity measured with respect to another measured quantity
- Number of cases that occur over a given time period divided by population at risk in the same time period

Nation	Under five mortality rate per 1,000 live births in 2008
France	4
Ghana	76
Sierra Leone	194
Afghanistan	257

Source: UNICEF: Statistics and Monitoring by Country

Annual Parasite Incidence (API)

Number of microscopically confirmed malaria cases detected during one year per unit population

$$\text{API} = \frac{\text{Confirmed malaria cases during 1 year}}{\text{Population under surveillance}} \times 1000$$

Most Common Software

- Microsoft Access
- Microsoft Excel
- Epi-Info
- SPSS
- Stata
- SAS

Learning Objectives

1. Learn to calculate descriptive statistics and run cross tabs in Excel and EpiInfo
2. Identify situations in which more complicated analysis is necessary

END

THANK YOU

Matgrace Consulting LLC

Annapolis Maryland United States.

2406453951

Matthewo@matgrace.com

