TRAINING ON DATA MANAGEMENT

& ANALYSIS OF STD/HIV DATA

RESOURCE BOOK

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Data Management and Analysis of HIV/AIDS Data

RESOURCE BOOK

National STD/AIDS Control Programme (NSACP), Sri Lanka

&

VHS-CDC Project

The Voluntary Health Services (VHS), India

Supported by Centers for Disease Control and Prevention (CDC/DGHT-India)

T.T.T.I. Post, Rajiv Gandhi Salai,

Taramani, Chennai – 600 113, Tamil Nadu, INDIA.

Ph.: +91-44-22541965 | Email: vhs.cdcproject@gmail.com

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES VHS WE SERVE THE PEOPLE WHO SERVE YOU

Book Title	:	Resource Book on Data Management and Analysis of HIV/AIDS Data.
No. of pages	:	233 pages including wrapper pages
Year of Publication	:	August 2019
Technical Guidance & Mentoring team	:	Dr Joseph D Williams, Director Projects, VHS Dr Ariyaratne Manathunge, Consultant-Venereologist, SIMU, NSACP
Team of Authors	:	Dr T Ilanchezhian, Senior Technical Advisor, VHS-CDC Project Dr Yujwal Raj, Technical Advisor (SI), VHS-CDC Project
Compiled & designed by	:	Ms T Sudha, Senior Programme Associate, VHS-CDC Project
© Copyright	:	VHS-CDC Project, Voluntary Health Services, Chennai–113.

This publication was supported by the Grant or Cooperative Agreement Number 6 NU2GGH001087-05-02, funded by the Centers for Disease Control and Prevention, (CDC) US and PEPFAR. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention, US PEPFAR or the Department of Health and Human Services.

Index

S. No.	Contents	Page No.
	Acronyms	5
	Foreword	7
	Acknowledgement	9
I	Data Management and Analysis of HIV/AIDS Data – An Introduction	11
II	Presentations and exercise formats:	12
1	Identifying & mapping data sources for HIV/AIDS analysis	13
1.1	EXERCISE 1: Identifying Programmatic Questions & Mapping Data Sources	
	Part-A: List all the Known HIV/AIDS Data Sources & Identify the Key Information Areas	19
	Part-B: Mapping Data Sources/ Datasets with Programmatic Questions	20
2	Principles of Database Management	21
2.1	EXERCISE 2: Exercise Dataset 2	36
	EXERCISE 2: Reviewing the Database	37
2.2	Data Management Systems - Assessment Tool	38
2.3	Data Management Checklist	40
3	Understanding Datasets of NSACP	42
	Understanding Programme Datasets under NSACP	61
3.1	EXERCISE 3: Assessment of NSACP Datasets	67
4	Variables & Indicators	68
4.1	EXERCISE 4: Variables and Indicators	
	PART A: Review the NSACP datasets - Write the Variable values	80
	PART B: Classify the Indicators	81
5	Data Quality Assessment	83
6	Data Adjustments & Validation	107
6.1	Exercise Dataset – Excel Sheet	115
7	Basic Programmatic Analysis – Measures & Methods	120
8	Basic & Advanced Functions in MS Excel	124
9	Commonly used Formulae in MS Excel	131
10	HIV/AIDS Specific Analysis from Programme Data	137
11	Data Triangulation	140
12	Communication of Data Analysis Results	157
13	Data Use for Decision Making - A Systems Approach	162
14	How to present the Data/ Information	175
15	Introduction to BIOSTATISTICS in Clinical Research	196
16	Statistical Package for the Social Sciences (SPSS)	220
17	Key Learnings & Next Steps	231

Acronyms

ACASI	Audio Computer Assisted Self Interview
AIDS	Acquired Immunodeficiency Syndrome
ANC	Ante-Natal Care
AR	Attributable Risk
ART	Anti-Retroviral Treatment
CAPI	Computer Assisted Personal Interviewing
CDC	Centers for Disease Control and Prevention
CON'T	Continued
CSS	Cross-Sectional Study
CSS	Case-Control Study
C&S	Care & Support
CST	Care, Support & Treatment
DD	Data Dictionary
DGHT	Division of Global HIV & TB
DM	Data Management
DQA	Data Quality Assurance
EPI Unit	Epidemiology Unit
FcFT	Facilitator cum Feedback Team
FGD	Focus Group Discussion
FSW	Female Sex Worker
GIS	Geographical Information Systems
HIV	Human Immunodeficiency Virus
HSS	HIV Sentinel Surveillance
IBBS	Integrated Biological and Behavioral Surveillance
IDI	In-Depth Interviews
IEC	Information Education Communication
ITDM	International Training on Data Management
KAP	Knowledge Attitude and Practice
KP	Key Population
Litt.	Literature
M&E	Monitoring and Evaluation
MSM	Men who have Sex with Men
NSACP	National STD/AIDS Control Programme
OR	Operational Research
OR	Odds Ratio

PAR	**Population Attributable Risk**
PEPFAR	**President's Emergency Plan for AIDS Relief**
PHI	**Public Health Inspector**
PHLT	**Public Health Laboratory Technician**
PHNS	**Public Health Nursing Sister**
PLHIV	**People Living with Human Immunodeficiency Virus**
PM	**Project Management**
PMTCT	**Prevention of Mother To Child Transmission**
PPT	**Power-Point Presentation**
PrEP	**Pre-Exposure Prophylaxis**
PRT	**Peer Review Team**
RCT	**Randomized Controlled Trial**
REC	**Research Ethics Committees**
RR	**Risk Ratio / Relative Risk**
STD	**Sexually Transmitted Diseases**
STI	**Sexually Transmitted Infections**
SI	**Strategic Information**
SIMU	**Strategic Information Management Unit**
TA	**Technical Assistance**
TB	**Tuberculosis**
TNA	**Training Needs Assessment**
VHS	**Voluntary Health Services**
Vs.	**Versus**
WHO	**World Health Organization**

Foreword

Dr Rasanjalee Hettiarachchi
Director
National STD/AIDS Control Programme (NSACP)
Sri Lanka

I am happy to write a foreword to this Resource Book on the *"Data Management and Analysis of HIV/AIDS Data"* developed/ compiled by VHS-CDC Project by engaging experts/ consultants in consultation and coordination with SIMU-NSACP team based on the training needs identified. This Resource Book has been brought out by VHS-CDC Project with the support of CDC for capacity building of the SI team including: SIMU team, Consultant-Venereologists & Medical Officers and Public Health Inspectors and Nursing Officers.

This Resource Book on "Data Management and Analysis of HIV/AIDS Data" will be of very much useful for imparting training and using as ready reckoner/ reference material for the participants.

This Resource Book is primarily introducing the basic principles and approaches of Data Management; various methods of Data Quality Assessment; validation; adjustments; basic skills in statistical analysis of program and epidemiological data; software packages used for statistical analysis; and guidelines for presentation, dissemination and use of data for programmatic purposes. This Resource Book will also contribute for enhancing knowledge and skills supported with hands-on training.

Training and capacity building are the key elements of VHS-CDC Project in providing Technical Assistance to NSACP on Strategic Information with the support of CDC/DGHT-India. VHS-CDC Project with the support of CDC in partnership with NSACP is planning for conducting the training on Data Management at different levels for building the capacity of SI team in the country.

This Resource Book will be used in the training on Data Management and Analysis of HIV/AIDS Data. The training programs are designed to build the data skills of participants in order to enhance the data quality, improve the data analysis and strengthen the use of HIV/AIDS data for epidemiological & programmatic decision making.

On behalf of NSACP, I wish to express my sincere thanks to Dr Joseph D Williams, Director Projects-VHS for his immense support in ensuring partnerships and continue to support in providing TA. We also appreciate the strategic support being extended by Dr T Ilanchezhian, Senior Technical Advisor, VHS-CDC Project for coordinating with NSACP and SIMU in providing technical assistance on strategic information and managing and coordinating the series of capacity building initiatives.

Thanks to VHS-CDC Project team, resource trainer/ consultant for the support extended in development of this Resource Book and contribution in conducting training program.

My gratitude should go to Dr Melissa Nyendak, Country Director, CDC/DGHT-India for the strategic leadership and guidance in providing Technical Assistance to NSACP, Ministry of Health, Nutrition & Indigenous Medicine, Govt. of Sri Lanka and CDC team for their support and guidance in these technical assistance initiatives.

Appreciate Dr Ariyaratne Manathunge, Consultant-Venereologist & Coordinator-SIMU, NSACP for his strategic leadership in coordinating the technical cooperation initiatives on TA to NSACP on SI with VHS-CDC Project, CDC team and contributions and support extended in bringing out this Resource Book and meaningful & successful conduct of the Training on Data Management.

Dr Rasanjalee Hettiarachchi
Director
National STD/AIDS Control Programme (NSACP), Sri Lanka

Acknowledgement

Dr Joseph D Williams
Director Projects
The Voluntary Health Services (VHS)
Chennai/INDIA

The Voluntary Health Services (VHS-CDC Project) with the support of Centers for Disease Control and Prevention (CDC/DGHT-India) in partnership with National STD/AIDS Control Programme (NSACP), Ministry of Health, Nutrition & Indigenous Medicine, Govt. of Sri Lanka is providing TA to NSACP on Strategic Information through a technical partnership initiative on:

- Enhance SIM Unit capacity to utilize electronic & manual program data for decision making;
- Improve capacity of SIM Unit to carryout management, analysis, documentation, and dissemination of summary program data reports;
- Improve capacity of SIM Unit to conduct and disseminate results of operational research;
- Consultation with stakeholders on monitoring & documentation of accomplishments & sustainability plans.

As part of this technical cooperation initiatives, VHS-CDC Project is providing capacity building initiatives, system strengthening, documentation and dissemination. In accordance with the capacity building initiatives, the project is organizing a series of training programs. VHS-CDC Project with the support of CDC/DGHT-India and in partnership with NSACP is conducting *"Training on Data Management and Analysis of HIV/AIDS Data"* for SIMU team, Consultant-Venereologists & Medical Officers, Public Health Inspectors, Development Officers and Nursing Officers.

To support this training, VHS-CDC Project is bringing out a Resource Book on Data Management and Analysis of HIV/AIDS Data. This resource book has been developed for the purpose of:

- Using in the training program for participants;
- Use as a reference material during the training and as and when required;
- Use this reference material for conducting/ capacity building of their team members in the respective Peripheral STD clinics.

This Resource Book has been developed with presentations, exercises, tools and other related reading materials by international professional trainers.

We thank Dr Rasanjalee Hettiarachchi, Director-NSACP for her leadership, supportive guidance in technical cooperation initiative.

We wish to acknowledge & thank Dr Ariyaratne Manathunge, Consultant - Venereologist, NSACP for his continuous support, strategic guidance and cooperation being extended in execution of this technical cooperation initiatives. Appreciate his strenuous support in bringing out this resource book and for conducting the training. Acknowledge the support extended by SIMU team, senior consultants in NSACP, SI team in peripheral STD clinics and key stakeholders.

We sincerely thank & acknowledge the technical guidance & support being extended by Dr Melissa Nyendak, Country Director, CDC/DGHT-India, Mr Lokesh Upadhyaya, Associate Director for Management & Operations, CDC/DGHT-India and CDC team. Wish to thank Ms Srilatha Sivalenka, Public Health Specialist, CDC/DGHT-India for her support in this cooperation initiatives.

We would like to thank Dr Yujwal Raj, Technical Advisor (SI), VHS-CDC Project for his support and contribution in developing resource materials.

We would like to thank Dr T Ilanchezhian, Senior Technical Advisor for his initiative, systematic support, strategic planning and contributions in bringing out this Resource Book.

We thank Ms T Sudha, Senior Programme Associate, VHS-CDC Project for her support in compiling, designing and bringing out this Resource Book.

We thank Mr B Kamalakar, Finance Controller and Mr S Sathyaraju, Associate Manager Finance, VHS-CDC Project and admin team for their support in publishing this Resource Book.

We greatly appreciate the fullest cooperation extended by NSACP & SIMU in this technical cooperation initiatives and for conducting capacity building initiatives.

Dr Joseph D Williams
Director Projects
The Voluntary Health Services (VHS), Chennai/INDIA

I. Data Management and Analysis of HIV/AIDS Data – An Introduction:

GOAL: To build the data skills of NSACP staff in order to enhance the data quality, improve the data analysis and strengthen the use of HIV/AIDS data for epidemiological & programmatic decision making under NSACP.

OBJECTIVES:

- To build the understanding of the NSACP staff on the programmatic & epidemiological databases under NSACP;
- To introduce the basic principles and approaches of data management;
- To orient participants on methods of data quality assessment, validation & adjustments;
- To build the basic skills in statistical data analysis of program and epidemiologic data;
- To briefly introduce various software packages used for statistical analysis; and
- To improve the presentation, dissemination and use of data for programmatic purposes.

METHODOLOGIES:

- Active Learning through discussions & review of examples & case studies
- Learning by Doing
- Individual Hands-on/ Practical Exercises
- Group Exercises
- Parallel work on selected data

OUTCOMES:

- Identified important questions/ topics of programmatic relevance suitable for secondary data analysis.
- Exposed participants to basic principles and methods of data management.
- Enhanced knowledge and skills on analyzing the data and use of data under NSACP through hands-on practice on examples and actual program data.
- Improved skills on effective use of data to make evidence-based decision making under the program.
- Evolved a data analysis plan as a follow-up to the training and identified the next steps.

II. Presentations and exercise formats

1. Identifying & mapping data sources for HIV/AIDS analysis

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

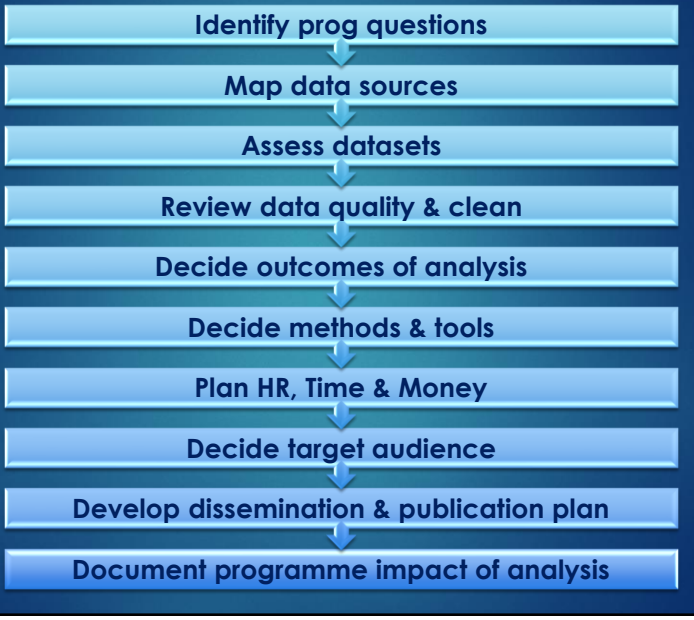
Planning Program Data Analysis (1)

1. Identify programmatic questions/ issues for decision making in your programme
2. Map the data sources/ datasets (Review these questions & identify, which data can help answer these questions)
3. Assess the datasets for key aspects before analysis
4. Undertake data quality assessments, adjust & validate the data wherever necessary (Data Cleaning)
5. Decide the key outcomes of the analysis
6. Decide the type of analysis, methods & tools
7. Plan resources for analysis – HR, Time, Money

Planning Program Data Analysis (2)

8. Decide the target audience
9. Develop the presentation, dissemination & publication plan (Way to present & disseminate the results of analysis, for programmatic improvements/ changes in implementation)
10. Document the impact of your analysis (details of the means & outcomes of bringing in programmatic changes due to the results of your analysis)

Planning Program Data Analysis



Programme Areas

Programmatic Questions (1)

- Epidemic questions
  - Why are HIV cases increasing at this clinic/ in this area?
  - What is the profile of the new HIV cases detected at the clinic?
  - What are the key drivers of HIV epidemic in this region?
  - Which areas/ regions have higher/ rising HIV epidemic?
- Progress & Priority questions
  - How is the scale up of STD services in this province?
  - Why STD/ HIV service uptake is low at a clinic?
  - What has been the progress against NSP/ Global targets in Sri Lanka?
  - Has the targets for prevention been achieved? Gaps? Reasons?
  - Which provinces should be prioritised for EMTCT?
  - How to improve ART adherence rates?

Programmatic Questions (2)

- Performance questions
  - Which STD clinics are the best performing in the country?
  - Why service delivery performance is poor in a province?
  - What are the top ten service delivery facilities that need focus under NSACP?
  - How can the performance parameters be improved at ART centres?
- Questions on evidence & information gaps
  - How to generate evidence on role of international migration in HIV epidemic?
  - What are the key information gaps that SI should focus in next two years?
  - Which provinces suffer from lack of adequate good quality data?
  - What are the barriers to strengthen data use for decision making under the programme?

Programmatic Questions?

Let us list out & discuss the programmatic questions/ Issues for decision making, that you are concerned about, in your area of work under NSACP!!!

(Use board/ display charts for the group work)

Fit for Secondary Analysis?

- Now, let us review these questions and identify, which data can help answer these questions?
- And is this data already collected elsewhere (Secondary Data)?
- Or does it need to be collected fresh from the field (Primary data)?

Rich Evidence Base of NSACP

- Expanded HIV Sentinel Surveillance System
- Rich Programme Data from STD/HIV Clinics & outreach
- Data from ANC clinics & PPTCT Centres
- Data from Blood Banks, TB Clinics, Hospitals & Labs
- Key Population Mapping & Size Estimations
- Integrated Biological & Behavioural Surveillance
- Intervention Data from Partner Agencies
- HIV Modeling, Estimations & Projections
- Increasing no. of Research Studies in HIV/AIDS
- DHS, Census Projections, Migration Statistics

List out Data Sources!!!

- Let us list out all the known available data sources on HIV/AIDS
- Consider all programme data sources, surveys, other research data, other health data that may directly or indirectly give related info
- You can also list out any unpublished datasets
- For each dataset, list out the key info areas
- Use Exercise 1 Format Part A

Map Prog Qs with Datasets

- Use Exercise 1 Format Part B
- List the programmatic questions on the left side
- If the data needs primary field data collection, mention 'primary'
- If it can be answered from secondary data, mention the s.no. of the dataset from Part A

EXERCISE 1: IDENTIFYING PROGRAMMATIC QUESTIONS & MAPPING DATA SOURCES

Part-A: List all the Known HIV/AIDS Data Sources & Identify the Key Information Areas

S. No.	Known HIV/AIDS Data Sources/ Datasets	Key information areas/ Broad themes

Exercise 1: Part-B: Mapping Data Sources/ Datasets with Programmatic Questions

S. No.	Programmatic Question/ Issue for Decision Making	Probable Data Sources/ Datasets for Analysis *(Mention S.Nos. from Part A Table)*

2. Principles of Database Management

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Overview

- Data – Info - Knowledge
- Basics of Data Management
- Data lifecycle
- Database structure
- Key principles

Discussion???

- What does good data management means to you?
- What are good things in data management system at your STD/HIV clinic?
- What are the challenges you face in data management at your STD/HIV clinic?
- What are the things that may be improved, with respect to data management at your STD/HIV clinic?

Data - Information - Knowledge

DATA: *Facts concerning people, objects, events or other entities. Databases store data.*

INFORMATION: *Data presented in a form suitable for interpretation.*

Data is converted into information by programs and queries. Data may be stored in files or in databases. Neither one stores information.

KNOWLEDGE: *Insights into appropriate actions based on interpreted data.*

Data Management

- Good data management practices ensure that data are of high quality (reliable, consistent, and complete) as well as readily available to stakeholders
- Data management entails putting personnel, policies, procedures, and organizational structures in place to ensure that data are accurate, consistent, secure, and available
- Encompasses all components of data flow from the data collection tools used during service delivery to the databases used for data storage/accessibility along with all intermediate steps

Purposes of Data Management Systems

- Monitoring and evaluation of control programs
- Plan actions, programs, and resources
- To prioritize the allocation of health resources
- To provide the basis for epidemiological research
- Accountability

Purposes of Data Management Systems

- Beneficiary Management Level
  - Provision of better healthcare
- Health-Facility Management Level
  - Health facility functions
- System Management Level
  - Planning; M&E
  - Governance
  - Public health interventions

Data Management Systems

- Data Flow: the process by which data are transferred or "moved" from the primary data sources to the database(s)
- Database: the structured storage of data that will be accessed for data use
- Data use: the active use (visualization, analyses, interpretation) of data

Data Lifecycle

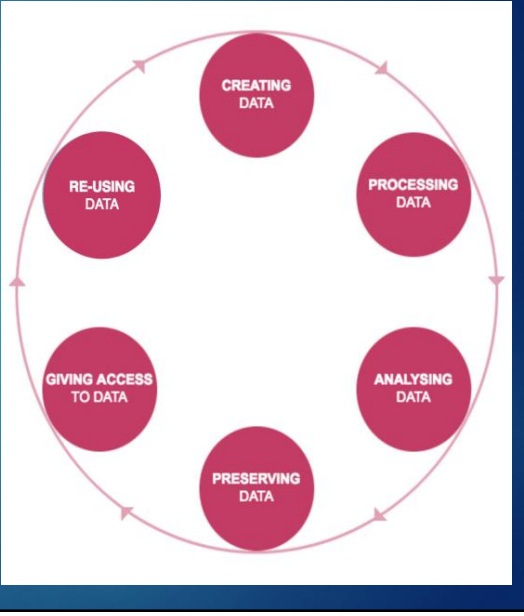
Creating data:

- Design research or program data collection/ acquisition
- Plan management- formats, storage etc
- Plan consent / sharing
- Identify existing data sources
- Collect data – experiment, observe, measure

Data Lifecycle

Processing data:

- Enter data, digitize, transcribe, translate
- Check, validate, clean data
- Anonymise data as required
- Describe data
- Manage and store data



Data Lifecycle

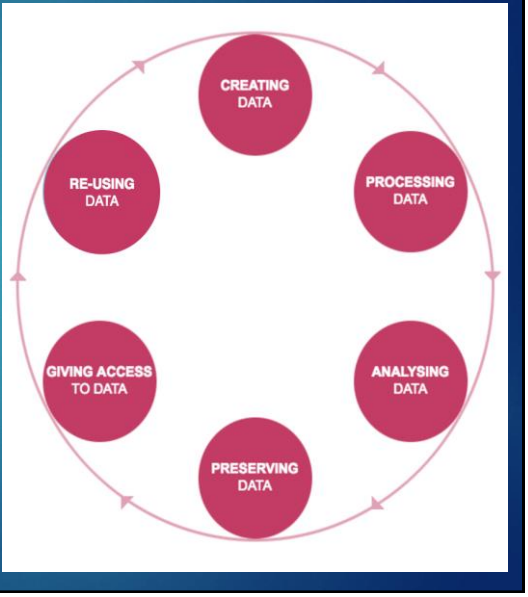
Analyzing data:

- Conduct analysis
- Interpret data
- Produce various outputs
- Develop reports and publications
- Prepare for preservation

Data Lifecycle

Preserving data:

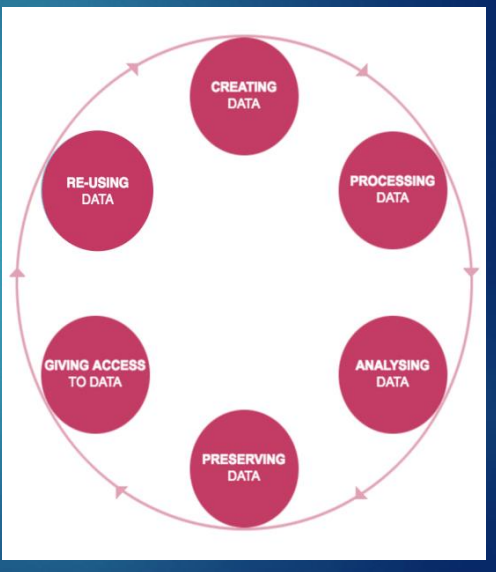
- Migrate data to appropriate format
- Migrate data to suitable medium
- Back up and storage
- Create metadata and documentation
- Archive data



Data Lifecycle

Giving access to data:

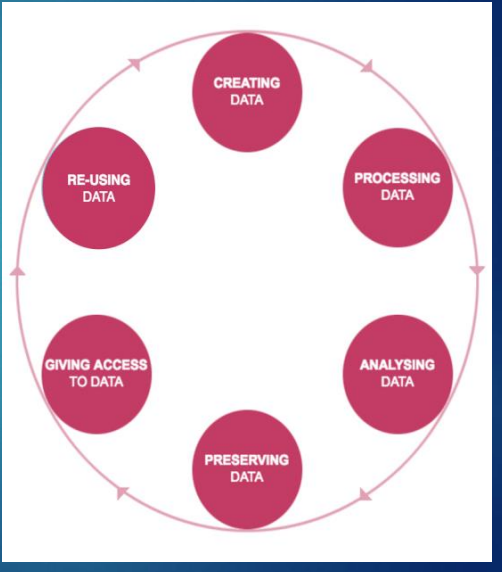
- Distribute data
- Share data
- Control access
- Establish copy right
- Promote data



Data Lifecycle

Re-using data:

- Follow up research
- New activities or research
- Undertake reviews
- Scrutinize findings
- Teach and learn



Principles of Good Data Management Plan

1. Data / indicator definition and audit trail
2. Efficient and timely data flow
3. Effective data verification and cleaning procedures
4. Usable data storage
5. Data access policy/procedures

Metadata

"Data about data"

- Description of fields
- Display and format instructions
- Structure of files and tables
- Security and access rules
- Triggers and operational rules
- Data collection details

Coding & Data Dictionary

- The person who designs the coding for data entry is not necessarily the one who does data analysis.
- Hence we need a dictionary of all codes for all variables and questions, explaining what code means what, and relates to what question, what variable…
- A data dictionary is a descriptive list of names, definitions, and attributes of data elements collected in an information system or database
- Why Data dictionary
  - Standardization of terms used in the data base for common understanding of different users
  - Enhances inter operability across systems

Efficient and Timely Data Flow

- Document and monitor the flow of data from point of data collection to data storage including data entry and storage, mechanisms of data transfer etc.
  - how are data first collected (e.g. paper forms, excel sheets by staff or via the Ministry of Health, directly entered into a database)?
  - How does data typically flow from the site to the district?
  - How does data flow from the district to the region?
  - Who typically are responsible for transferring data between steps?
  - Document the steps till data storage (excel, Access database, shared drive, etc.) and staff responsible for transferring data

Data Preparation

- Data Review
- Data Quality Assessment
- Coding & Data Dictionary
- Data Cleaning: Adjustments & Imputations

Data Analysis

- Computation & Recoding
  - Retain original data
  - Give appropriate variable names
- Generation of tables
- Tests and Interpretation
- Summarising results, writing & publications

Data Storage & Retrieval

- The data storage system must be designed to meet the requirements and data use needs and allow for simple and reliable access to the data
- Hard & soft forms
- Migration to suitable form & medium
- Back-up & storage
- Tables, Identifiers & Query Options
- Central database or databases linked in such a way that data can be easily and automatically combined as required

Data Protection & Sharing

- Infrastructure security measures and user access permissions
  - Physical safety - Lock & key
  - Restricted access - Password protect all digital files/ databases; access control based and privileges based on roles
- Confidentiality (Anonymising) - agreements for collectors, entry operators and supervisors
- Sharing protocol for authorized access
- Data recovery plans where there is loss of data

Maximizing Data Usefulness

- Data Use Plan can guide the design of the data management system
- It is essential to take the time to collect our data properly the first time to avoid re-collecting data
- Review workflows to ensure that we avoid re-processing data as much as possible increases our efficiency
- Electronic data systems and automated reporting drastically reduces data manipulation

Long-Term Planning

- What will happen to my data after my project ends?
- How can I appraise the value of my data?
- What are my options for archiving and preserving my data?
- What are my options for publishing and sharing data?

Database & DBMS

DATABASE: *A shared collection of interrelated data designed to meet the varied information needs of an organization.*

DATABASE MANAGEMENT SYSTEM: *A collection of programs to create and maintain a database.*

Define – Construct – Manipulate

Advantages of Database Processing

- More information from same data
- Shared data
- Balancing conflicts among users
- Controlled redundancy
- Consistency
- Integrity
- Security
- Increased productivity

Creating Database Structure

- Cases or Records or Rows
- Fields or Variables or Columns
- Primary key/ Unique ID/ Case ID
- Column headers
- Data Labels
- Value Labels
- Type of Data
- Data limits/ checks
- Metadata

Some Key Principles

- One row for one case
- No duplicate variables/ column heads
- No duplicate primary key
- No sub-totals in rows; sub-totals/ totals in columns are OK
- No merged cells; No merged headings
- No blank cells (Fill blank cells with some code or Impute)
- No two data types in one column (Text/Num/Code)
- Short & crisp variable names; Not too long
- Colour code in Excel
- Numeric data is better than text

Compiled Datasets

If dataset is a compilation of multiple datasets on the same cases:

- Ensure column alignment
- After joining the datasets, create a unique ID/ primary key
- Unique ID can be created by joining two three variables. E.g. Province, Facility Name & Month
- Avoid different spellings of same case
- Ensure uniformity in the data type & format under each column

Exercise 2: Review Database

- Review the database shared with you
- Identify the following from the 'Exercise 2 Dataset' given to you
  - What is a case in this database?
  - No. of cases
  - No. of fields and their names
  - Primary key
  - Metadata is adequate and clear
  - Any of the key principles violated in the dataset?

EXERCISE DATASET

				2003		2004		2005		2006		2007		2008	
State	District_name	Site_name	Site_Type	NT	NP	NT	NP	NT	NP	NT	NP	NT	NP	NT	NP
Andhra Pradesh	Visakhapatnam	Priyadarshini Service Organization, Vishakhapatnam	FSW	250	32	250	35	250	39	250	27	250	22	248	12
Andhra Pradesh	East Godavari	East Godawari	FSW	250	113	250	102	250	67	250	23	250	39	244	45
Andhra Pradesh	Prakasam	Lakshmi Development Society, Ongle, Prakasam	FSW	250	61	250	27	250	18	250	11	250	18	235	16
Andhra Pradesh	Hyderabad	Hyderabad	FSW	250	40	250	25	250	30	250	24	249	18	250	37
Andhra Pradesh	Kurnool	Parameswari, Kurnool	FSW	250	22	250	25	250	8	249	6	249	9	249	12
Andhra Pradesh	Warangal	Warangal	FSW	250	32	250	47	250	32	249	22				
Andhra Pradesh	Guntur	Needs Society, Chilakaluripet, Guntur	FSW			250	36	250	33	250	15	250	32	213	19
Andhra Pradesh	West Godavari	Action for Development, Bhimavaram (New 07)	FSW									249	41	247	35
Andhra Pradesh	Khammam	JAGRUTI (New 07)	FSW									250	39	247	66
Andhra Pradesh	Adilabad	AIRTDS,Mancherial (New 07)	FSW									217	10	249	38
Andhra Pradesh	Nalgonda	ANKITA (New 07)	FSW									250	25	227	21
Andhra Pradesh	Srikakulam	Swageti Project,Youth Club of Bejjipuram (New 07)	FSW									250	15	243	10
Andhra Pradesh	Warangal	MARI, Hnamkonda (New 07)	FSW									120	8	248	12
Andhra Pradesh	Khammam	JAGRUTI (New 07)	FSW			248	12								
Andhra Pradesh	Nalgonda	ANKITA (New 07)	FSW					250	11						

FSW: Female Sex Workers; NT: Number Tested; NP: Number Positive

REVIEWING THE DATABASE

Review the 'Exercise 2 Dataset' given to you and fill the following table.

S. No.	Observe/ Identify	Observations/ Details
1	Is it a single dataset or compiled dataset? Why do you say so?	
2	What is a case in this database?	
3	No. of cases	
4	No. of fields and their names	
5	Primary key	
6	Metadata is adequate and clear	
7	**Any of the key principles violated in the dataset?**	
A	One row for one case	
B	No duplicate variables/ column heads	
C	No duplicate primary key	
D	No sub-totals in rows; sub-totals/ totals in columns are OK	
E	No merged cells; No merged headings	
F	No blank cells (Fill blank cells with some code or Impute)	
G	No two data types in one column (Text/Num/Code)	
H	Short & crisp variable names; Not too long	

	Data Management Systems - Assessment Tool	National	State	District	Action to be taken
	Standard operating procedures have been written that define roles and responsibilities for data compilation, reporting, data analysis, dissemination and use.				
	There is a comprehensive, singular, master list of health facilities, with unique facility identifier and service domain, that includes the private sector and special facilities (military, etc.).				
	There is a formal mechanism to update and keep current the master facility list (e.g., a census of all facilities is conducted every 5 years).				
	Data collection systems for client data (e.g. clinical episodes) are standardized across all implementing partners and donors.				
	Personnel (clinicians and other staff) have been trained in the collection of the client data, and for the input of the data into the computer database (where applicable).				
	Printed guidelines are available at all health facilities (and in applicable community-based programs) to assist with client-level data collection.				
	Health data (paper or electronic) are stored appropriately and according to national policies.				
	There is a schedule/plan for update, reproduction, and distribution of data collection tools.				
	The data flow pattern (i.e., data flow from client encounter forms -> summary tools [e.g., a register or tally sheet] -> periodic aggregate reporting form) is clearly defined and understood by staff.				
	There are printed guidelines available at all health facilities (and in applicable community-based programs) to assist with data compilation and reporting.				
	Relevant staff at health facilities (and in applicable community-based programs) have received training on data compilation and reporting.				
	Data disaggregations by key stratifiers (age, sex, geography) are maintained during compilation and transfer in order to permit equity analysis.				
	Data transfer to the next level occurs in a timely way, making use of innovation and IT where appropriate and available.				
	There is a data quality assurance plan that is shared with health programs, other government ministries, donors and other stakeholders to guide activities aimed at improving data quality.				
	Routine health data quality assurance standards are defined and enforced, including completeness, timeliness, accuracy, integrity, and consistency over time.				
	Roles and responsibilities for data quality are assigned at each level, including verification of data, summarizing data quality issues, and developing and implementing improvements strategies.				

	Data Management Systems - Assessment Tool	National	State	District	Action to be taken
	Training and capacity development for data quality assurance are provided at facility, district, and national levels using standard methods.				
	Systematic and comprehensive assessments of facility data quality are conducted regularly in advance of health sector planning, including analysis of completeness, timeliness, accuracy, and consistency over time (e.g., data quality review) and which result in published reports describing data quality issues and plans to address them.				
	Data management staff conducts regular checks of accuracy and completeness of data prior to submitting reports to the next level (using automated electronic checks where appropriate).				
	Data quality assurance is linked to the health sector planning cycle in the country so that information on data quality is available prior to the use of data for planning.				
	There is collaboration between the MOH, government agencies (e.g., national statistics office) and other national stakeholders (e.g., donors, universities, etc.) on data quality assurance so that assessments are conducted with an element of independence (i.e., no conflict of interest).				

Data Management Checklist

	Databases	1	2	3	4
	Name of database				
	1. Type of data				
	What kinds of data - survey, interview, observation, machine or instrument collected, physical samples, models, etc. - are you collecting?				
	What formats - paper, digital, image, audio, other - will the data be in?				
	Will it be reproducible? What would happen if it got lost or became unusable later?				
	2. Data formats and standards				
	Do you have data dictionaries, code books or other documentation to explain terms, variable names, codes and abbreviations used?				
	Have you provided documentation describing how the data were collected or created?				
	Have you used standard collection methods, standard data formats, and standard file format choices (if these exist for your research field)?				
	3. Data access policies				
	Have you removed personal or sensitive information from your data to ensure privacy protection?				
	Have you established who owns the copyright of your data?				
	Do you have documentation on how institutional and personal credit should be acknowledged for your data?				
	Are your data, records, and files labeled and logically organized?				
	Have you used consistent and easy to understand file names?				

Data Management Checklist

	Databases	1	2	3	4
	Name of database				
	4. Data use and distribution				
	How will your data be made available?				
	Do you plan to limit re-use or re-distribution of your data? If so, why and for how long?				
	5. Data preservation and archiving				
	Have you made arrangements for the long-term storage and preservation of your data (both physical and digital collection items)?				
	Do you have data security plans in place to ensure that copies of your data are stored and backed up on a regular basis?				
	Are you using data formats and software that enable sharing and ensure long-term validity of data, such as non-proprietary software and software based on open standards?				
	When converting from one format to another, have you checked that no data are lost or changed in the process?				

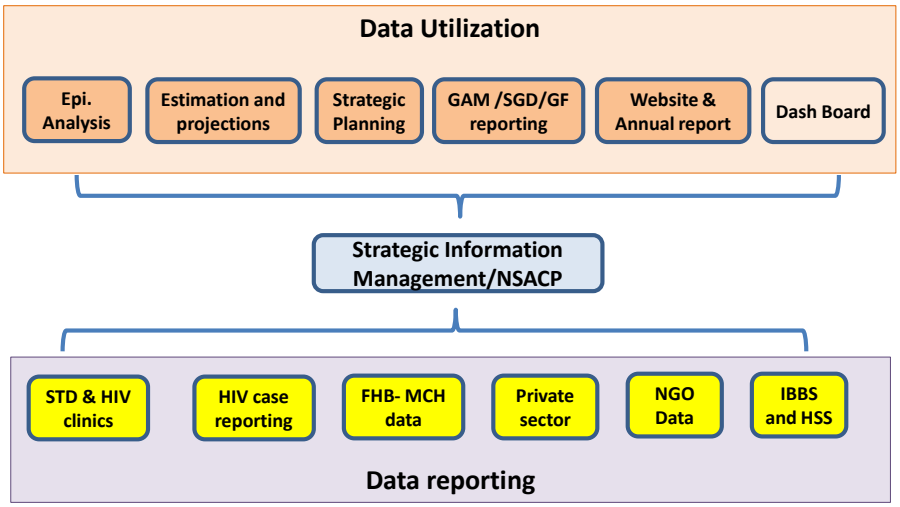
3. Understanding datasets of NSACP

Majority of HIV testing and all HIV treatment is happening in Public STD clinics of NSACP

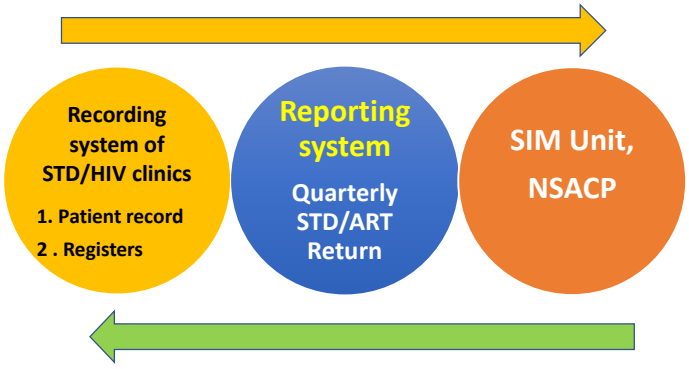
33 STD Clinics (22 has ART facility)

Main Data Systems of NSACP

- HIV case reporting system (Epid unit)
- Monitoring of STD services (SIM unit)
- Monitoring of ART services (SIM unit)
- HIV surveillance
  - HIV Sentinel Surveillance (HSS)
  - Integrated Biological and Behavioural Surveillance (IBBS)



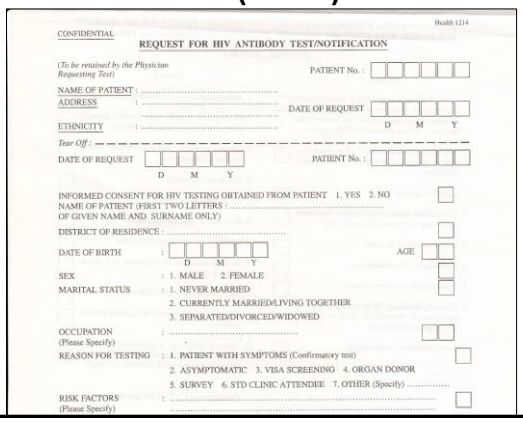
Data Management From STD & HIV clinics



HIV Case Reporting System

Case Reporting

1. HIV
   - H.1214 form (OLD)



Health 1214

CONFIDENTIAL

REQUEST FOR HIV ANTIBODY TEST/NOTIFICATION

(To be retained by the Physician Requesting Test)

PATIENT No. :

NAME OF PATIENT :

ADDRESS :

DATE OF REQUEST D M Y

ETHNICITY :

Tear Off :

DATE OF REQUEST D M Y

PATIENT No. :

INFORMED CONSENT FOR HIV TESTING OBTAINED FROM PATIENT 1. YES 2. NO

NAME OF PATIENT (FIRST TWO LETTERS : OF GIVEN NAME AND SURNAME ONLY)

DISTRICT OF RESIDENCE :

DATE OF BIRTH : D M Y

AGE

SEX : 1. MALE 2. FEMALE

MARITAL STATUS : 1. NEVER MARRIED
2. CURRENTLY MARRIED/LIVING TOGETHER
3. SEPARATED/DIVORCED/WIDOWED

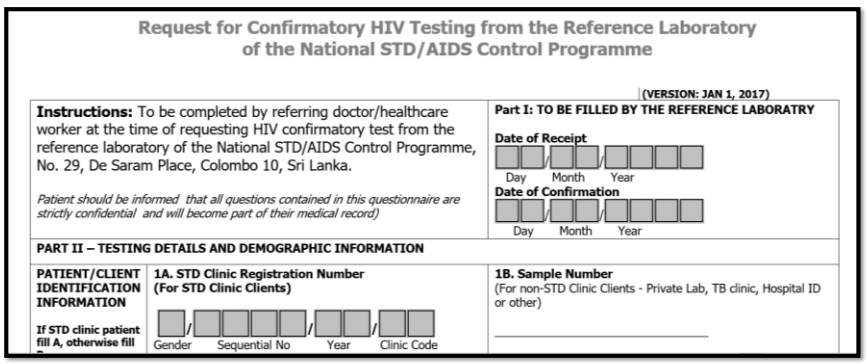
OCCUPATION (Please Specify) :

REASON FOR TESTING : 1. PATIENT WITH SYMPTOMS (Confirmatory test)
2. ASYMPTOMATIC 3. VISA SCREENING 4. ORGAN DONOR
5. SURVEY 6. STD CLINIC ATTENDEE 7. OTHER (Specify) .....

RISK FACTORS (Please Specify) :

Case Reporting

1. HIV
   - H.1214 form (modified in 2017)



Request for Confirmatory HIV Testing from the Reference Laboratory of the National STD/AIDS Control Programme

(VERSION: JAN 1, 2017)

Instructions: To be completed by referring doctor/healthcare worker at the time of requesting HIV confirmatory test from the reference laboratory of the National STD/AIDS Control Programme, No. 29, De Saram Place, Colombo 10, Sri Lanka.

Patient should be informed that all questions contained in this questionnaire are strictly confidential and will become part of their medical record)

Part I: TO BE FILLED BY THE REFERENCE LABORATRY

Date of Receipt Day / Month / Year

Date of Confirmation Day / Month / Year

PART II – TESTING DETAILS AND DEMOGRAPHIC INFORMATION

PATIENT/CLIENT IDENTIFICATION INFORMATION

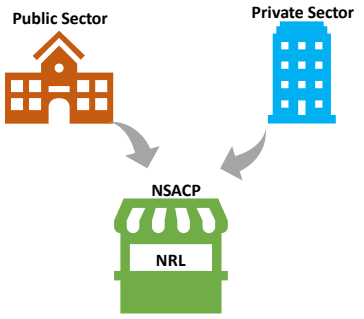
If STD clinic patient fill A, otherwise fill

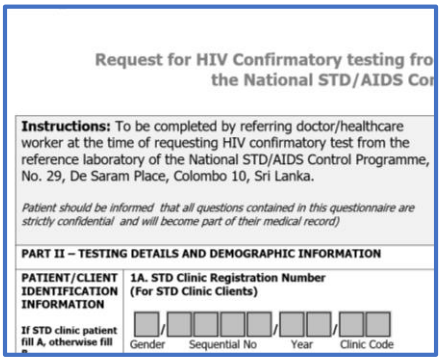
1A. STD Clinic Registration Number (For STD Clinic Clients)

Gender / Sequential No / Year / Clinic Code

1B. Sample Number
(For non-STD Clinic Clients - Private Lab, TB clinic, Hospital ID or other)

HIV case Reporting system





Request for HIV Confirmatory testing fro
the National STD/AIDS Cor

Instructions: To be completed by referring doctor/healthcare worker at the time of requesting HIV confirmatory test from the reference laboratory of the National STD/AIDS Control Programme, No. 29, De Saram Place, Colombo 10, Sri Lanka.

Patient should be informed that all questions contained in this questionnaire are strictly confidential and will become part of their medical record)

PART II – TESTING DETAILS AND DEMOGRAPHIC INFORMATION

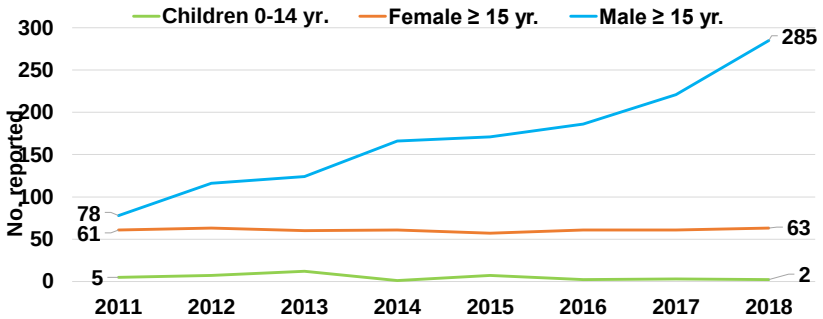
PATIENT/CLIENT IDENTIFICATION INFORMATION

If STD clinic patient fill A, otherwise fill

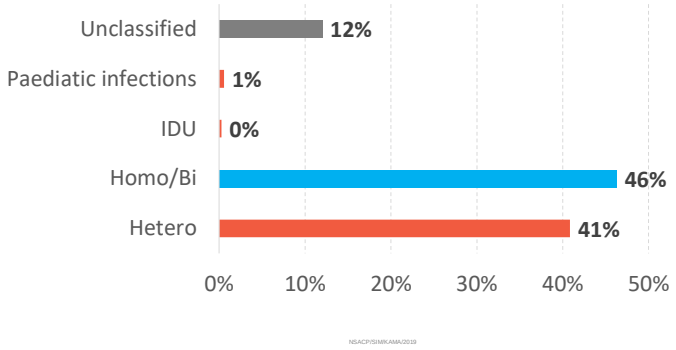
1A. STD Clinic Registration Number (For STD Clinic Clients)

Gender / Sequential No / Year / Clinic Code

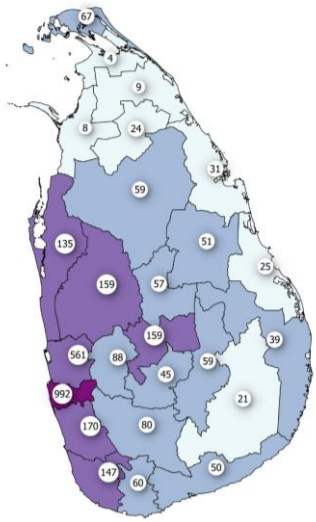
Trends of reported HIV by Age and Sex 2011-2018



Probable Mode of Transmission of New HIV Cases in 2018 (N=350)

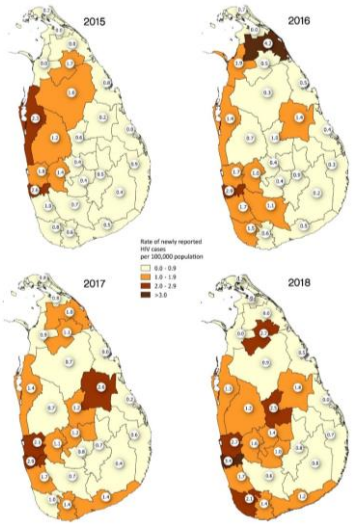


All HIV infections reported 1987-2018



Rate of newly reported HIV infections/ 100,000 population reported in 2015, 2016, 2017 and 2018

- Colombo
- Gampaha
- Puttalam
- Kegalle



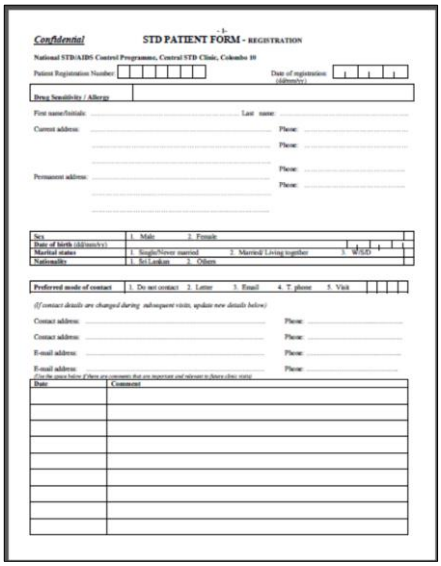
Monitoring and Evaluation of STD services

Source: SIMU/NSACP/2014

Sources of STD Data

1. **STD clinics**
   - **Patient records**
   - **STD clinic registers**
   - **STD quarterly return**
2. **Private sector, Blood bank etc**
   - **HIV and syphilis testing data**

SIM Unit/National STD/AIDS Control Programme-2015



Confidential

STD PATIENT FORM - REGISTRATION

National STD/AIDS Control Programme, Central STD Clinic, Colombo 10

Patient Registration Number | | | | | | | Date of registration (dd/mm/yy) | | | | | | |

Drug Sensitivity / Allergy

First name/Initials: ..... Last name: .....

Current address: ..... Phone: .....

..... Phone: .....

..... Phone: .....

Permanent address: ..... Phone: .....

.....

.....

Sex	1. Male 2. Female
Date of birth (dd/mm/yy)						
Marital status	1. Single/Never married 2. Married/Living together 3. W/S/D
Nationality	1. Sri Lankan 2. Others

Preferred mode of contact	1. Do not contact 2. Letter 3. Email 4. T. phone 5. Visit

(If contact details are changed during subsequent visits, update new details below)

Contact address: ..... Phone: .....

Contact address: ..... Phone: .....

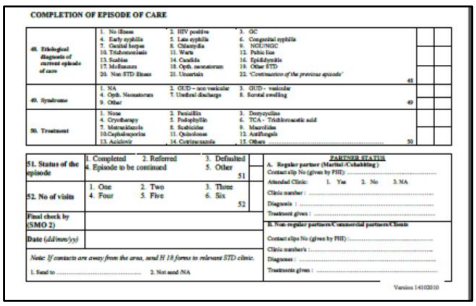
E-mail address: ..... Phone: .....

E-mail address: ..... Phone: .....

(Use the space below if there are comments that are important and relevant to future clinic visits)

Date	Comment

STD Patient Form



COMPLETION OF EPISODE OF CARE

49. Etiological diagnosis of current episode of care	1. No illness 4. Early syphilis 7. Genital herpes 10. Trichomoniasis 13. Scabies 17. Molluscum 20. Non STD illness	2. HIV positive 5. Late syphilis 8. Chlamydia 11. Warts 14. Candida 18. Opth. neonatorum 21. Uncertain	3. GC 6. Congenital syphilis 9. NGU/NGC 12. Pubic lice 16. Epididymitis 19. Other STD 22. *'Continuation of the previous episode'*	48
49. Syndrome	1. NA 4. Opth. Neonatorum 9. Other	2. GUD – non vesicular 7. Urethral discharge	3. GUD – vesicular 8. Scrotal swelling	49
50. Treatment	1. None 4. Cryotherapy 7. Metronidazole 10. Cephalosporins 13. Acyclovir	2. Penicillin 5. Podophyllin 8. Ketoconazole 11. Quinolones 14. Cotrimoxazole	3. Doxycycline 6. TCA - Trichloroacetic acid 9. Macrolides 12. Antifungals 15. Others	50

51. Status of the episode	1. Completed 4. Episode to be continued	2. Referred	3. Defaulted 5. Other	51
52. No of visits	1. One 4. Four	2. Two 5. Five	3. Three 6. Six	52
Final check by (SMO 2)				
Date (dd/mm/yy)				

Note: If contacts are away from the area, send H 18 forms to relevant STD clinic.

1. Send to ..... 2. Not send /NA

PARTNER STATUS

A. Regular partner (Marital /Cohabiting)

Contact slip No (given by PH) : .....

Attended Clinic: 1. Yes 2. No 3. NA

Clinic number : .....

Diagnosis : .....

Treatment given : .....

B. Non regular partners/Commercial partners/Clients

Contact slips No (given by PH): .....

Clinic number/s : .....

Diagnoses : .....

Treatments given : .....

Version 14102010

STD Registers

1. Main Register
2. Subsequent Visit Register
3. Outpatient Blood Testing Register
4. Interview and Contact Tracing Register
5. IEC/BCC/ Awareness Programme Register
6. HIV testing and counseling Register

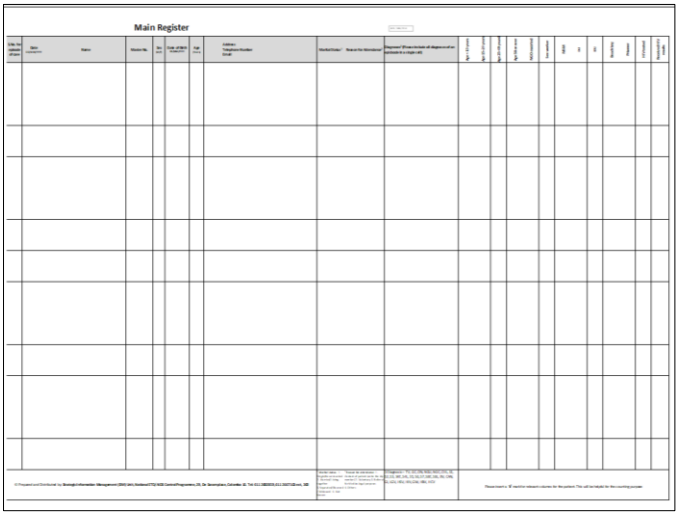
Source: SIMU/NSACP/2014

Cont., STD Registers

7. Condom Distribution Register
8. Commercial Sex Worker Register
9. Outreach Blood Survey Register
10. Defaulter Register
11. Antenatal Syphilis Register
12. Pre-employment/Visa Screening

Source: SIMU/NSACP/2014

Main Register


Main Register

Main Register

- S.No. for episode of care
- Date DD/MM/YYYY
- Name
- Master No.
- Sex (M/F)
- Date of Birth DD/MM/YYYY
- Age (Years)
    - Address
      Telephone Number
      Email
- Marital Status[1]
- Reason for Attendance[2]
- Diagnoses[3] (Please include all diagnoses of an episode in a single cell)

- Age < 15 years
- Age 15-24 years
- Age 25-49 years
- Age 50 or over
- NGO escorted
- Sex worker
- MSM
- DU
- IDU
- Beach boy
- Prisoner
- HIV tested
- Received HIV results

QUARTERLY STD RETURN

Quarterly STD Return

QUARTERLY RETURN FROM STD CLINICS IN SRI LANKA *(Revision: 29.05.2017)*

Name of the STD clinic : _____

Period of the return : __ /__ / 20_ _ to __ /__ / 20_ _ __(___ Quarter of 20_ _)

Return completed by (Name and designation) _____ **Checked by (Name and designation)** _____

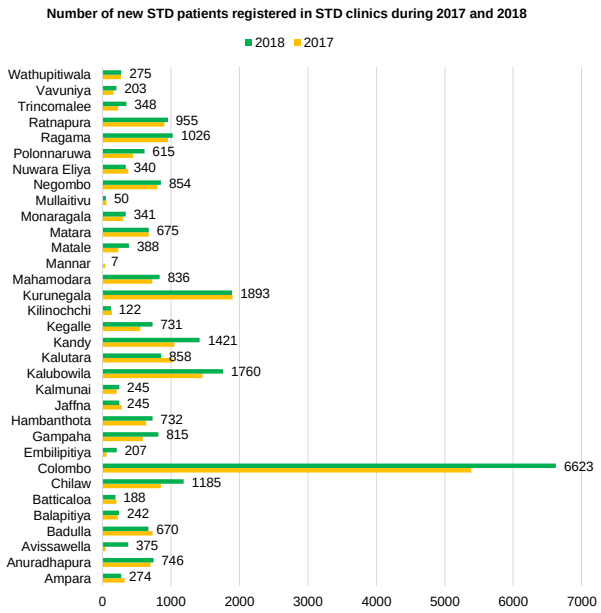
Date of completion : __ / __ / 20_ _

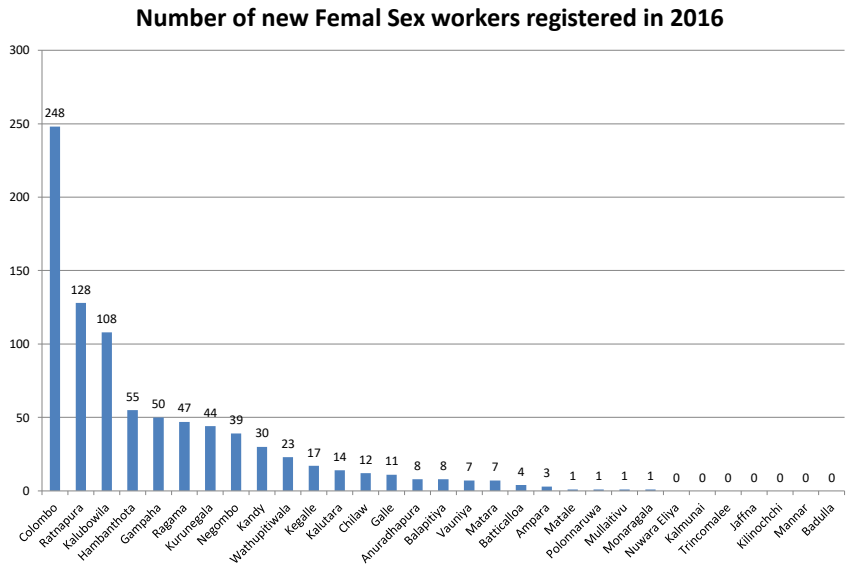
Table 1. Total number of new diagnoses* by age group and sex *(Source: Main Register- 2017)*

SN	Name of the Disease	Male					Female				
		<15 year	15-24 year	25-49 year	50+ year	Total	<15 year	15-24 year	25-49 year	50+ year	TOTAL
1.1	Infectious syphilis										
1.2	Late syphilis										
1.3	Early Congenital Syphilis										
1.4	Late Congenital Syphilis										
1.5	Gonorrhoea and presumptive GC										
1.6	Opthalmia neonatorum										
1.7	NGU/NGC										
1.8	Chlamydia										
1.9	Genital herpes										
1.10	Genital warts										
1.11	Pelvic inflammatory disease (PID)										
1.12	Trichomoniasis										
1.13	Candidiasis										
1.14	Bacterial vaginosis										
1.15	Other STIs										
1.16	**TOTAL STI**										
1.17	Non STI/Uncertain										
1.18	No illness										
1.19	**GRAND TOTAL**										

Excel STD Database

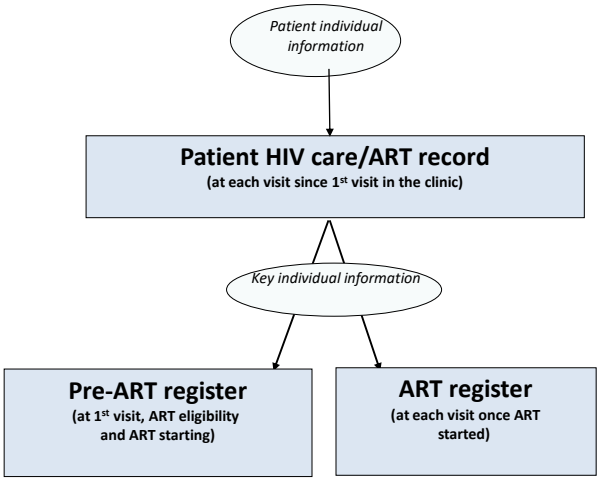
New Patients Registered of all STD Clinic in 2017				New Patients Registered of all STD Clinic in 2018			
	New patients registered				New patients registered		
	Male	Femal	2017		Male	Femal	2018
Ampara	138	188	326	Ampara	135	139	274
Anuradhapura	376	330	706	Anuradhapura	397	349	746
Avissawella	26	18	44	Avissawella	207	168	375
Badulla	320	410	730	Badulla	269	401	670
Balapitiya	112	109	221	Balapitiya	131	111	242
Batticaloa	64	138	202	Batticaloa	57	131	188
Chilaw	359	496	855	Chilaw	628	557	1185
Colombo	3470	1918	5388	Colombo	4394	2229	6623
Embilipitiya	23	37	60	Embilipitiya	104	103	207
Gampaha	294	295	589	Gampaha	399	416	815
Hambanthota	364	272	636	Hambanthota	411	321	732
Jaffna	160	121	281	Jaffna	160	85	245
Kalmunai	115	93	208	Kalmunai	122	123	245
Kalubowila	814	645	1459	Kalubowila	1044	716	1760
Kalutara	560	466	1026	Kalutara	424	434	858
Kandy	515	533	1048	Kandy	636	785	1421
Kegalle	243	312	555	Kegalle	382	349	731

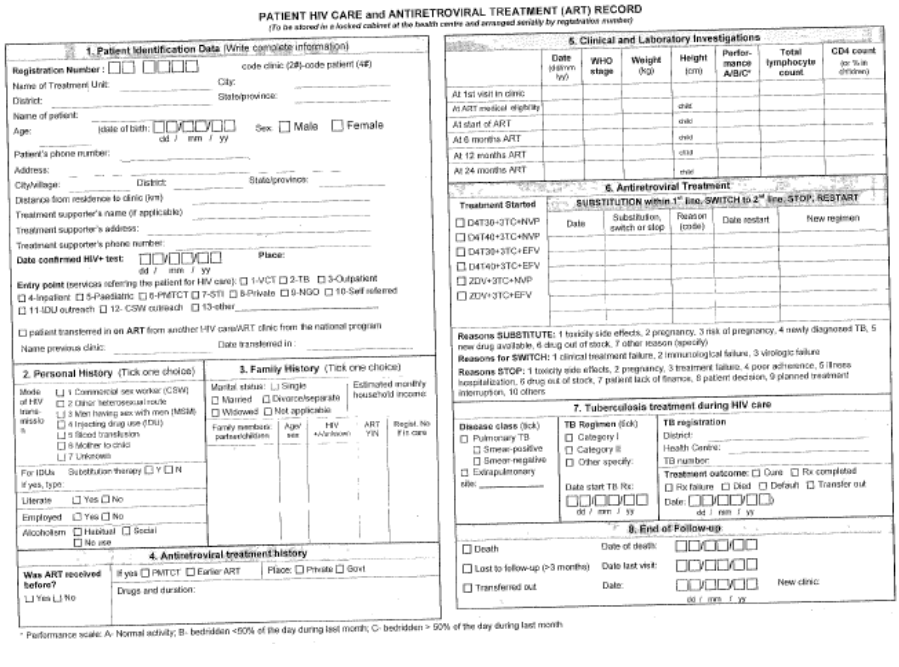




Monitoring and Evaluation of ART centers

WHO Paper-based ART Monitoring system





PATIENT HIV CARE and ANTIRETROVIRAL TREATMENT (ART) RECORD

(To be stored in a locked cabinet at the health centre and arranged serially by registration number)

1. Patient Identification Data (Write complete information)

Registration Number : ☐☐ ☐☐☐☐ code clinic (2#)-code patient (4#)

Name of Treatment Unit: City:

District: State/province:

Name of patient:

Age: Idate of birth: ☐☐/☐☐/☐☐ (dd / mm / yy) Sex ☐ Male ☐ Female

Patient's phone number:

Address:

City/Village: District: State/province:

Distance from residence to clinic (km)

Treatment supporter's name (if applicable)

Treatment supporter's address:

Treatment supporter's phone number:

Date confirmed HIV+ test: ☐☐/☐☐/☐☐ (dd / mm / yy) **Place:**

Entry point (services referring the patient for HIV care): ☐ 1-VCT ☐ 2-TB ☐ 3-Outpatient ☐ 4-Inpatient ☐ 5-Paediatric ☐ 6-PMTCT ☐ 7-STI ☐ 8-Private ☐ 9-NGO ☐ 10-Self referred ☐ 11-IDU outreach ☐ 12- CSW outreach ☐ 13-other

☐ patient transferred in **on ART** from another HIV care/ART clinic from the national program

Name previous clinic: Date transferred in :

2. Personal History (Tick one choice)

Mode of HIV transmission: ☐ 1 Commercial sex worker (CSW) ☐ 2 Clinic heterosexual route ☐ 3 Men having sex with men (MSM) ☐ 4 Injecting drug use (IDU) ☐ 5 Blood transfusion ☐ 6 Mother to child ☐ 7 Unknown

For IDUs Substitution therapy ☐ Y ☐ N

If yes, type:

Literate ☐ Yes ☐ No

Employed ☐ Yes ☐ No

Alcoholism ☐ Habitual ☐ Social ☐ No use

3. Family History (Tick one choice)

Marital status: ☐ Single ☐ Married ☐ Divorce/separate ☐ Widowed ☐ Not applicable

Estimated monthly household income:

Family members: partner/children	Age/ sex	HIV +/-/Unknown	ART Y/N	Regist. No if in care

4. Antiretroviral treatment history

Was ART received before? ☐ Yes ☐ No	If yes ☐ PMTCT ☐ Earlier ART	Place: ☐ Private ☐ Govt
	Drugs and duration:	

5. Clinical and Laboratory Investigations

	Date (dd/mm /yy)	WHO stage	Weight (kg)	Height (cm)	Performance A/B/C*	Total lymphocyte count	CD4 count (or % in children)
At 1st visit in clinic							
At ART medical eligibility				child:			
At start of ART				child:			
At 6 months ART				child:			
At 12 months ART				child:			
At 24 months ART				child:			

6. Antiretroviral Treatment

Treatment Started	SUBSTITUTION within 1st line, SWITCH to 2nd line, STOP, RESTART			
	Date	Substitution, switch or stop	Reason (code)	Date restart	New regimen
☐ D4T30+3TC+NVP					
☐ D4T40+3TC+NVP					
☐ D4T30+3TC+EFV					
☐ D4T40+3TC+EFV					
☐ ZDV+3TC+NVP					
☐ ZDV+3TC+EFV					

Reasons SUBSTITUTE: 1 toxicity side effects, 2 pregnancy, 3 risk of pregnancy, 4 newly diagnosed TB, 5 new drug available, 6 drug out of stock, 7 other reason (specify)

Reasons for SWITCH: 1 clinical treatment failure, 2 immunological failure, 3 virologic failure

Reasons STOP: 1 toxicity side effects, 2 pregnancy, 3 treatment failure, 4 poor adherence, 5 Illness hospitalization, 6 drug out of stock, 7 patient lack of finance, 8 patient decision, 9 planned treatment interruption, 10 others

7. Tuberculosis treatment during HIV care

Disease class (tick) ☐ Pulmonary TB ☐ Smear-positive ☐ Smear-negative ☐ Extrapulmonary site:

TB Regimen (tick) ☐ Category I ☐ Category II ☐ Other specify:

Date start TB Rx: ☐☐/☐☐/☐☐ (dd / mm / yy)

TB registration District: Health Centre: TB number:

Treatment outcome: ☐ Cure ☐ Rx completed ☐ Rx failure ☐ Died ☐ Default ☐ Transfer out Date: ☐☐/☐☐/☐☐ (dd / mm / yy)

8. End of Follow-up

☐ Death Date of death: ☐☐/☐☐/☐☐

☐ Lost to follow-up (>3 months) Date last visit: ☐☐/☐☐/☐☐

☐ Transferred out Date: ☐☐/☐☐/☐☐ (dd / mm / yy) New clinic:

* Performance scale: A- Normal activity; B- bedridden <50% of the day during last month; C- bedridden > 50% of the day during last month

Pre ART Register

PRE ART REGISTER Month: _____ Year: _____ Name of Clinic / Hospital: _____

	1 Date 1st entered into HIV care at this clinic	2 Registration number	3 Patient's name, Address and Contact number	4 Age	5 Sex: M / F	6 HIV Confirmation test: Date	6 HIV Confirmation test: Place	7 Entry point code 1 to 13*	8 Risk factor code 1to7**	9 Literate Y/N	10 Employed Y/N	11 CPT*** Date Start	12 Date of TB Screening & Result#, Category Regimen Date Rx start	13 Date medically eligible for ART	14 Why medically eligible?	15 Date ART started	16 End of follow-up before ART: Date of death	16 End of follow-up before ART: Date lost to FU (last visit)	16 End of follow-up before ART: Date transferred
1															WHO stage CD4 #/% ILC#				
2															WHO stage CD4 #/% ILC#				
3															WHO stage CD4 #/% ILC#				
4															WHO stage CD4 #/% ILC#				
5															WHO stage CD4 #/% ILC#				
6															WHO stage CD4 #/% ILC#				
7															WHO stage CD4 #/% ILC#				
8															WHO stage CD4 #/% ILC#				
9															WHO stage CD4 #/% ILC#				

ART Register

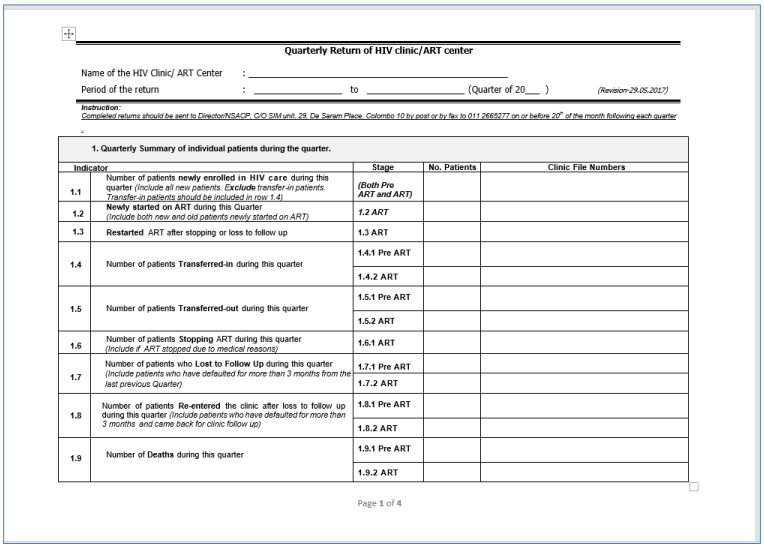
MONTH: July YEAR: 2004

Monthly visits: • 1st row, write patient outcome: on treatment **(OT)** if patient picked up ART drugs, stopped **(ST)** if ART was stopped by the doctor, missing **(MIS)** if the patient missed the appointed visit, lost to follow-up **(LFU)** if the patient is missing >3 months, restart **(RS)** if ART

we. 2	mo. 1	mo.2	mo.3	mo.4	mo.5	mo.6	mo.7	mo.8
OT	OT	OT	OT	OT	OT	OT		
A	A	A	A	A	A	A		
OT	OT	ST	ST	ST	RS	OT		
A	A	A				A		
OT	OT	OT	OT	OT	OT	OT		
A	B	B	A	A	B	B		
OT	OT	OT	MIS	OT	OT			
A	A	A		C	C			
OT	OT	OT	OT	OT	OT	OT		
A	A	A	A	A	A	A		
OT	OT	OT	OT	MIS	MIS			
A	A	C	C					

2 missing persons in January 2005 at 6 months FU

Quarterly Return from ART Centers



Quarterly Return of HIV clinic/ART center

Name of the HIV Clinic/ ART Center : _____

Period of the return : _____ to _____ (Quarter of 20___) *(Revision 29.05.2017)*

Instruction:
Completed returns should be sent to Director/NSACP, C/O SIM unit, 29, De Saram Place, Colombo 10 by post or by fax to 011 2665277 on or before 20th of the month following each quarter

1. Quarterly Summary of individual patients during the quarter.

Indicator		Stage	No. Patients	Clinic File Numbers
1.1	Number of patients **newly enrolled in HIV care** during this quarter *(Include all new patients. Exclude transfer-in patients. Transfer-in patients should be included in row 1.4)*	***(Both Pre ART and ART)***		
1.2	**Newly started on ART** during this Quarter *(Include both new and old patients newly started on ART)*	***1.2 ART***		
1.3	**Restarted** ART after stopping or loss to follow up	**1.3 ART**		
1.4	Number of patients **Transferred-in** during this quarter	**1.4.1 Pre ART**		
		1.4.2 ART		
1.5	Number of patients **Transferred-out** during this quarter	**1.5.1 Pre ART**		
		1.5.2 ART		
1.6	Number of patients **Stopping** ART during this quarter *(Include if ART stopped due to medical reasons)*	**1.6.1 ART**		
1.7	Number of patients who **Lost to Follow Up** during this quarter *(Include patients who have defaulted for more than 3 months from the last previous Quarter)*	**1.7.1 Pre ART**		
		1.7.2 ART		
1.8	Number of patients **Re-entered** the clinic after loss to follow up during this quarter *(Include patients who have defaulted for more than 3 months and came back for clinic follow up)*	**1.8.1 Pre ART**		
		1.8.2 ART		
1.9	Number of **Deaths** during this quarter	**1.9.1 Pre ART**		
		1.9.2 ART		

Page 1 of 4

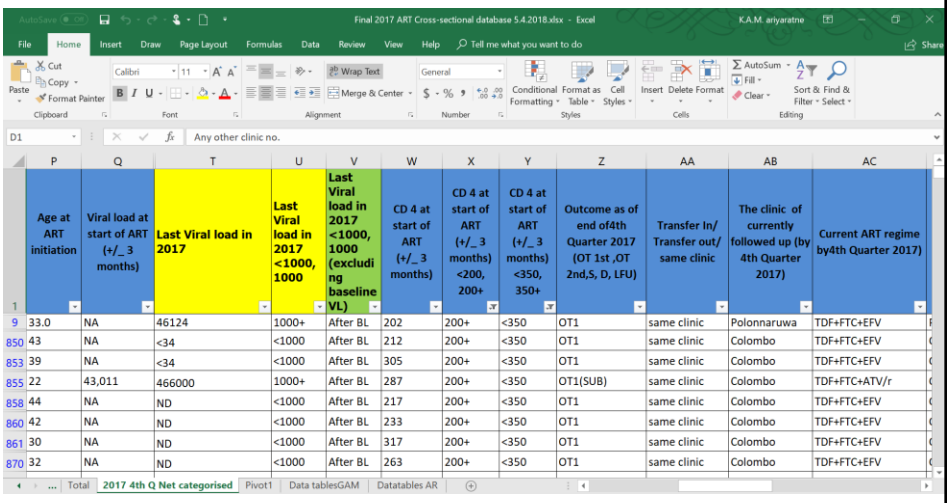
Databases of HIV clinics

- **1. Cross-sectional ART database**
- **2. Cohort database**

Cross-sectional ART database

- Example of an Excel file
- CD4 count at baseline
- Viral load at 12 months of ART

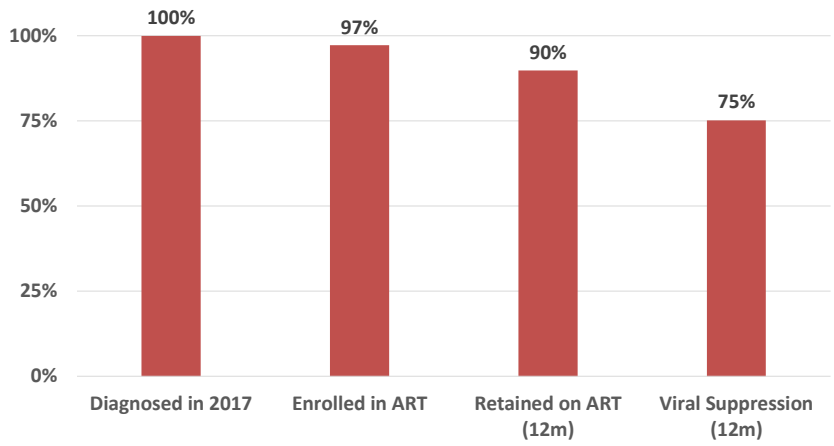
Cross-sectional ART database



Final 2017 ART Cross-sectional database 5.4.2018.xlsx - Excel

	P	Q	T	U	V	W	X	Y	Z	AA	AB	AC
1	Age at ART initiation	Viral load at start of ART (+/_ 3 months)	Last Viral load in 2017	Last Viral load in 2017 <1000, 1000 (excluding baseline VL)	Last Viral load in 2017 <1000, 1000 (excluding baseline VL)	CD 4 at start of ART (+/_ 3 months)	CD 4 at start of ART (+/_ 3 months) <200, 200+	CD 4 at start of ART (+/_ 3 months) <350, 350+	Outcome as of end of4th Quarter 2017 (OT 1st ,OT 2nd,S, D, LFU)	Transfer In/ Transfer out/ same clinic	The clinic of currently followed up (by 4th Quarter 2017)	Current ART regime by4th Quarter 2017)
9	33.0	NA	46124	1000+	After BL	202	200+	<350	OT1	same clinic	Polonnaruwa	TDF+FTC+EFV
850	43	NA	<34	<1000	After BL	212	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV
853	39	NA	<34	<1000	After BL	305	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV
855	22	43,011	466000	1000+	After BL	287	200+	<350	OT1(SUB)	same clinic	Colombo	TDF+FTC+ATV/r
858	44	NA	ND	<1000	After BL	217	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV
860	42	NA	ND	<1000	After BL	233	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV
861	30	NA	ND	<1000	After BL	317	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV
870	32	NA	ND	<1000	After BL	263	200+	<350	OT1	same clinic	Colombo	TDF+FTC+EFV

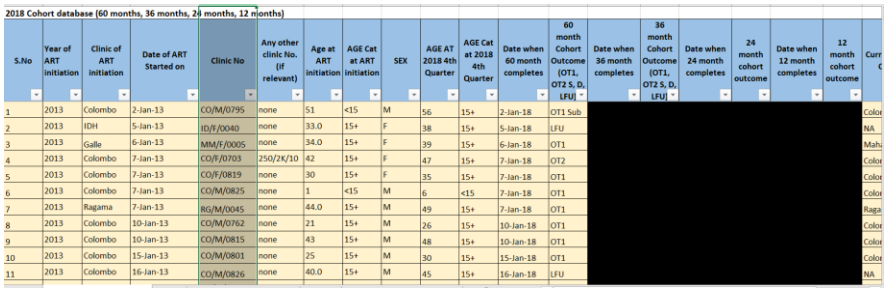
Cascade Analysis



Cohort database

- **Example of Excel file**
- **Status after**
  - **12 months,**
  - **24 months and**
  - **60 months**

Cohort database



2018 Cohort database (60 months, 36 months, 24 months, 12 months)

S.No	Year of ART initiation	Clinic of ART initiation	Date of ART Started on	Clinic No	Any other clinic No. (if relevant)	Age at ART initiation	AGE Cat at ART initiation	SEX	AGE AT 2018 4th Quarter	AGE Cat at 2018 4th Quarter	Date when 60 month completes	60 month Cohort Outcome (OT1, OT2 S, D, LFU)	Date when 36 month completes	36 month Cohort Outcome (OT1, OT2 S, D, LFU	Date when 24 month completes	24 month cohort outcome	Date when 12 month completes	12 month cohort outcome	Curr
1	2013	Colombo	2-Jan-13	CO/M/0795	none	51	<15	M	56	15+	2-Jan-18	OT1 Sub							Colo
2	2013	IDH	5-Jan-13	ID/F/0040	none	33.0	15+	F	38	15+	5-Jan-18	LFU							NA
3	2013	Galle	6-Jan-13	MM/F/0005	none	34.0	15+	F	39	15+	6-Jan-18	OT1							Mah
4	2013	Colombo	7-Jan-13	CO/F/0703	250/2K/10	42	15+	F	47	15+	7-Jan-18	OT2							Colo
5	2013	Colombo	7-Jan-13	CO/F/0819	none	30	15+	F	35	15+	7-Jan-18	OT1							Colo
6	2013	Colombo	7-Jan-13	CO/M/0825	none	1	<15	M	6	<15	7-Jan-18	OT1							Colo
7	2013	Ragama	7-Jan-13	RG/M/0045	none	44.0	15+	M	49	15+	7-Jan-18	OT1							Raga
8	2013	Colombo	10-Jan-13	CO/M/0762	none	21	15+	M	26	15+	10-Jan-18	OT1							Colo
9	2013	Colombo	10-Jan-13	CO/M/0815	none	43	15+	M	48	15+	10-Jan-18	OT1							Colo
10	2013	Colombo	15-Jan-13	CO/M/0801	none	25	15+	M	30	15+	15-Jan-18	OT1							Colo
11	2013	Colombo	16-Jan-13	CO/M/0826	none	40.0	15+	M	45	15+	16-Jan-18	LFU							NA

Status after
- **12 months,**
- **24 months and**
- **60 months**

Summary-Databases

- **HIV case report database**
- **STD clinic database**
- **ART Cross sectional database**
- **ART cohort database**

3. Understanding Programme Datasets under NSACP

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Programme Data – Common Issues

- Collected for monitoring purposes; Using data for other purposes should be attempted with caution
- Mostly aggregate data; not individual level data
- Limited scope for slice & dice
- Varying quality from different centres at different times
- Changing formats and reporting mechanisms
- Changing programme strategies affect the data reported
- Scale up of centres
- Interest & ability; absence/ change of personnel affects data
- Commodity supplies & stock outs affect data
- Largely paper-based; needs computerisation

HIV Testing Data

- Mostly no. of tests; not individuals tested
- Limited information on risk profiling
- Beneficiary segmentation
- Routes of transmission data
- Couple testing and sero-discordance
- Linkages to ART

PPTCT Data

- Mostly no. of tests; not individual pregnant women
- Documentation of trimester and gravida
- Background profile data
- Documentation of multiple testing is challenging
- Public sector – Private sector overlaps
- Linkages with ART
- Positive pregnancy follow-up data, baby testing & outcomes
- Longitudinal data often not maintained

STD Clinic Data

- Footfalls vs STD Episodes vs Individuals
- Background profiles of STD Clinic Attendees
- Syndromic data vs Etiologic data
- Linkage to HIV testing
- Re-occurrence of STDs; Incidence vs Prevalence
- Correct denominators for case rates, incidence, prevalence etc.
- Partner testing data

ART Data

- Aggregate vs Individual data
- Cross-sectional vs longitudinal data
- Demographic, Epidemiological, clinical data
- PLHIV profiles
- Sero-discordance
- Viral suppression data
- Key population treatment data
- Building testing & treatment cascades at sub-national level

Assessment of Data Sources

- Utility & Usability:

S.No	Criteria	Score: 3	Score: 2	Score: 1
1	Explains Epidemic	Explains very well	Explains moderately	Doesn't explain
2	Reflects Programme Performance	Reflects very well	Reflects moderately	Doesn't Reflect
3	Availability of Data at the Desired Level	Easily Available	Not Easily Available	Not Available
4	Feasibility of Extraction & Use	Easy	Difficult	Very Difficult

Issues at various steps (1)

- Collecting data/ Gathering Info from the beneficiaries or patients
  - Skipping questions
  - Way of asking questions & eliciting information
  - Interviewer bias
  - Judgmental approach
  - Workload, timings & staff cooperation
- Documenting in Registers
  - Using the correct registers & formats
  - Inefficient documentation – Info split into multiple registers
  - Duplication
  - Transcribing errors
  - Filling the register without enquiring
  - Blank cells filled later or left unfilled
  - Illegible writing/ improper noting down

Issues at various steps (2)

- Compilation/ Aggregation
  - Lack of clarity of definitions/ Absence of data dictionary
  - Errors in manual counting
  - Wrong method of including/ excluding criteria
  - Computation Errors/ Errors in data that needs to be computed from register
- Data entry
  - Typo errors in data entry
  - Mismatch b/w register & data entered
  - Incomplete data entry/ gaps/ blanks
  - Entry in wrong fields/ wrong format

Issues at various steps (3)

- Reporting
  - Non-reporting/ Irregular reporting
  - Lack of timeliness in reporting
  - Reporting in wrong/ outdated formats
- Compilation & Analysis
  - Compilation errors
  - Dataset merging issues
  - Missed/ excluded in compilation
  - Aggregation errors
  - Errors/ Wrong approaches in analysis

Exercise 3: Assess NSACP Datasets

- Review the NSACP Dataset that you have brought for the following. Each group to take up one prog dataset.
  - Time period
  - Geographic scope
  - Data lifecycle for the dataset
  - Assessment of Utility & Usability
  - Probable issues with the dataset at various steps
    - Collecting data
    - Documenting in Registers
    - Compilation/ Aggregation
    - Data entry
    - Reporting

EXERCISE 3: ASSESSMENT OF NSACP DATASETS

Review the NSACP Dataset brought by you in your group and fill the following table.

S.No.	Attributes of Dataset	Observations/ Details	
1	Name of the dataset		
2	Reference Period of dataset		
3	Geographic Scope	National/ Provincial/ District/ Clinic Details:	
4	**Data Lifecycle**	**Who performs this function?**	**When/ At what frequency?**
A	Collecting data		
B	Creating dataset		
C	Processing data		
D	Analysing data		
E	Preserving data		
F	Have access to data		
G	Publish results		
H	Using & reusing data		
5	**Assessment of Utility & Usability**	**Score the following on a scale of 3, where 3 is good, 2 is moderate, 1 is low**	
A	Explains epidemic		
B	Reflects programme performance		
C	Availability of data at the desired level		
D	Feasibility of extraction & use		
6	**Probable issues with the dataset at various steps**	**Discuss & write the key issues affecting the dataset at each step below.**	
A	Collecting data		
B	Documenting in Registers		
C	Counting & Aggregation		
D	Data entry		
E	Reporting		
F	Compilation & Analysis		

4. Variables & Indicators

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VHS VOLUNTARY HEALTH SERVICES WE SERVE THE PEOPLE WHO SERVE YOU

Overview

- What & Why
- Levels of Measurement
- Types of Variables
- Indicators
- Types of Indicators
- Computing Indicators

Definition

Definition of an Element in Chemistry

- … substance consisting of atoms
- … chemically the simplest substances
- … cannot be broken down using chemical methods.
- … can only be changed into other elements using nuclear methods.

Definition of Variable – The Element of Research

- … consisting information
- … in simplest form
- … cannot be broken down into more basic info
- … can be changed into other variables by some technique

Variables… Diff. Viewpoints

- … that changes from person to person, from observation to observation (think in terms of height, weight, Hb status)
- … either a result of some force or is itself the force that causes a change in another variable (think in terms of Obesity & Hypertension)
- … characteristics of interest in a study that has different values for different subjects or objects (think in terms of 'proportion of ANC having Hb<7gm/dl)

Why to focus on key variables?

- … Provide focus when writing the Introduction section.
- … Major terms to use when searching for research articles for the Literature Review.
- … The key variables are the terms to be operationally defined.
- … Provide focus to the Methods section.
- … The Instrument will measure the key variables.

Variables are the "things" we're measuring, or collecting data on, or forming groups on, in order to conduct our research study & answer the questions

Why Levels of Measurement?

- Indicate the relationship among the values assigned to different attributes of a variable
- E.g. Hypertension – High, Medium, Low; >180/100 mmHg; 180/100 – 120/80 mmHg; <120/80 mmHg
- helps to decide how to interpret the data from that variable..
- helps to decide what statistical analysis is appropriate on the values that were assigned.

Let's check the visual acuity of a patient

- Vision normal / Vision not normal
- Normal vision/ Visual impairment / Severe visual impairment/ Blind
- VA-6/6 to 6/18 , VA-6/18 to 6/60 , VA-6/60 to 3/60, VA-3/60 to NPL

Same characteristic measured in different ways

Qualitative vs Quantitative

BP- 120/80mmHg

BP- 134/86 mmHg

BP- 160/100mmHg

BP- 200/120 mmHg

Hypertensive/ Normotensive

Stage-1
Stage2
Stage-3
Hypertension

Levels of Measurement

- **Nominal**
  - Qualities can not be graded
  - No order preserving transformation
  - E.g. Site of malignancy, constipation present, vomiting absent, Ram, Black
- **Ordinal**
  - Qualities can be graded
  - Order preserving transformation
  - E.g. Stage I, stage II, stage III – CaCx
  - Mild, Moderate and Severe Hypertension
  - For convenience
  - When quantitative scale not available or difficult

Before we go further…
Let's visit Sikkim

Last Metaled Road and permissible limit for tourists

Absolute Zero

Or

Arbitrate Zero

Levels of Measurement... Metric Scale

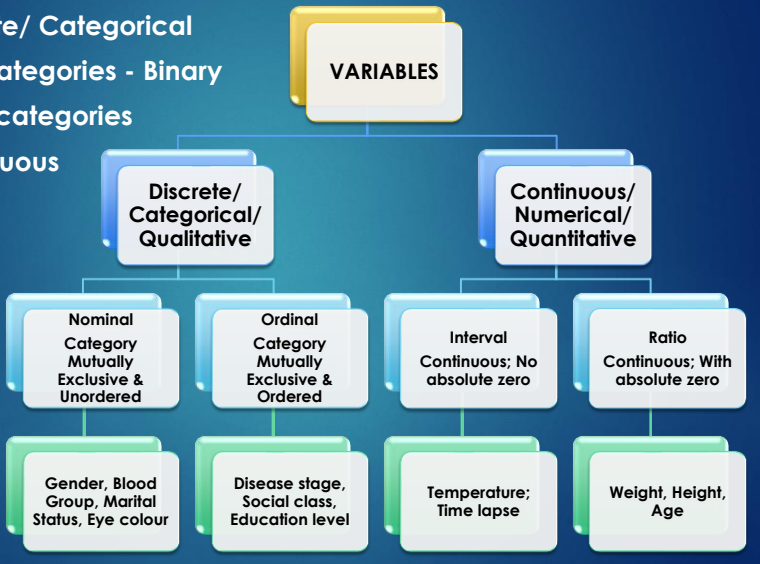
- Interval
  - Measured and have constant, equal distances between values
  - Zero point arbitrary; No Absolute Zero
  - E.g. IQ test - difference between a 100 and a 110 equal to the scoring distance between 120 and 130, but no true zero on this test and an IQ of 140 is not twice as high as an IQ of 70.
  - Distance between two cities
  - Time interval between two doses
  - Cannot say one is double or triple or so many times the other; (Becomes a ratio)

Levels of Measurement... Metric Scale

- Ratio
  - Agreed absolute zero; Same reference point
  - Zero is meaningful
  - E.g. Parity 4 and Parity 0; Height from the sea level;
  - Laboratory tests;
  - Weight of a person (A person weighs 100 kilos - twice as heavy as a person who weighs 50 kilos (measure of zero kilos meaningful)
  - Pulse rate, Respiratory rate, etc.

Types of Variables

- Discrete/ Categorical
  - 2 categories - Binary
  - >2 categories
- Continuous



Let's classify the variables

Table 1.1 Demographic characteristics and selected risk factors for endometrial cancer. Values are numbers (percentages) unless stated otherwise

Characteristic	Cases	Controls
Number of women (*n*)	832	846
Mean (SD) age (years)	55.3 (8.60)	55.7 (8.58)
Education		
No formal education or just elementary school	204 (24.5)	234 (27.7)
Middle school	503 (60.5)	513 (60.6)
College and above	125 (15.0)	99 (11.7)
Marital status		
Unmarried	14 (1.7)	10 (1.2)
Married or cohabiting	724 (87.0)	742 (87.7)
Separated, divorced, or widowed	94 (11.3)	94 (11.1)
Per capita income in previous year (yuan)		
≤4166.7	230 (27.7)	244 (28.9)
4166.8–6250.3	243 (29.2)	242 (28.6)
6250.4–8333.3	57 (6.9)	50 (5.9)
≥8333.3	301 (36.2)	309 (36.6)
No of pregnancies		
None	62 (7.5)	35 (4.1)
1	137 (16.5)	109 (12.9)
2	199 (23.9)	208 (24.6)
3	194 (23.3)	207 (24.5)
4	141 (17.0)	157 (18.6)
≥5	99 (11.9)	130 (15.4)
Cancer among first degree relatives	289 (34.7)	228 (27.0)
Oral contraceptive use	147 (17.7)	207 (24.5)
Regular exercise	253 (30.4)	287 (33.9)
Age at menarche*	14 (13 to 16)	15 (13 to 16)
Age at menopause (among postmenopausal women)*	50.1 (48.6 to 52.5)	49.4 (47.1 to 51.1)
Body mass index*	25.1 (22.7 to 27.9)	23.7 (21.4 to 26.3)

From Xu et al (2004). Soya food intake and risk of endometrial cancer among Chinese women in Shanghai: population-based case–control study. *British Medical Journal*, **328**, 1285–1291: reproduced by permission of the BMJ Publishing Group.
*Median (25th to 75th centile).

What do you say…

- A story of 100 meter race: Three runners are participating from three different districts of Madhya Pradesh . Each runner is assigned a color-coded T-shirt (BLUE/RED/GREEN) to differentiate from each other (Nominal scale). The winner is declared along with the declaration of first runner up and second runner up based on the criteria that who reaches the destination first, second and last. The rank order of runners such as "second runner up as 3", "first runner up as 2" and the "winner as 1" (Ordinal scale). During the tournament, 3 consequent races are arranged at a time interval of 10 day (Interval scale). The time spent by each runner in completing the race (Ratio scale) was also announced.
- Distance between two cities in kilometers
- Divorced, married and widowed
- Mild, moderate and Severe Anaemia
- Number of umbrellas sold in April, July, October
- GE cases among fish eaters and non- fish eaters

Types of Variable Values

- Text variables
  - Coded
  - Un-coded
- Numeric variables
  - Continuous Measures
  - Codes
- Alphanumeric codes

Other Issues

- Identify the correct Source Registers & Sections/ Columns in the register from where the variable is extracted
- Explore the Dis-aggregations available for the variable
  - By gender
  - By age or age category (child/youth/adult/elderly)
  - By occupation
  - By marital status
  - By type of infection (viral/bacterial)
  - By source of referral
  - By place of residence/ location of clinic

Indicators

- Measures computed from one or more variables to reflect a process or an outcome
  - Program element that needs tracking
  - Measures an aspect of a program's performance
- Could be a number, percentage, rate or ratio
- E.g. No. of HIV tests done in a month; % of pregnant women receiving ANC services; STD cure rate; HIV transmission rate; Sex ratio;

Types of Indicators

- Indicators of need/ Epidemic indicators
  - Size of KP, HIV prevalence
- Input Indicators
  - Finance, HR, Commodities
- Process Indicators
  - No. of outreach camps; No. of trainings; Counseling duration; Waiting time at ART clinic
- Output Indicators
  - No. of facilities opened; No. of tests done; No. of KP reached; Cascade indicators
- Outcome Indicators
  - Condom uptake; No. of sexual partners; Health seeking behaviour
- Impact Indicators
  - New HIV infections/ Incidence; AIDS deaths/ Mortality; Survival

Computing Indicators

- Indicator Title
- Indicator Type
- Indicator Definition
- Geographic Scope
- Time Reference
- Numerator & Denominator
- Inflation/ Deflation factors
- Units
- Use/ Utility/ Importance/ Application

Indicator Eg

- Prevalence of Syphilis
- Epidemic Indicator/ Outcome Indicator
- Definition: Percentage of STD patients found reactive for Syphilis in last one year in a fixed geographic area/ at a STD clinic
- Num: No. of STD patients found reactive for Syphilis in last 1 yr
- Den: No. of STD patients tested for Syphilis in last 1 yr
- Inflation/ Deflation factor: Proportion of STD patients tested for Syphilis out of all those who visited STD clinic
- Units: %
- Monitor STD epidemics; Reflect the success of STD control programme

Is it a Variable or Indicator?

- As of December 1, 2016, more than 1 million PLHIV were on antiretroviral treatment in India
- 45.3% of FSW respondents from Western Province in the IBBS (who reported having regular clients) reported having anal sex with regular clients
- 7000 KP were tested for HIV during the FY 2015-16
- Herpetic to Non-herpetic STD Ratio at Colombo clinic is 1.2
- Condoms were distributed to 5,500 youth during the outreach camps

Exercise 4

- Review the NSACP datasets that you have brought
- List out at least 20 variables from the dataset.
- Write the variable values.
- Classify them based on the level of measurement
- Mention the source register & the dis-aggregations available for the variable.
- Evolve 12 indicators – 2 of each type – based on the variables listed above. Classify them and indicate their num, den, units & importance

EXERCISE 4: VARIABLES & INDICATORS

PART A: Review the NSACP datasets that you have brought. List out at least 20 variables from the dataset. Write the variable values. Classify them based on the level of measurement. Mention the source register & dis-aggregations available for the variable.

S. No	Variable	Variable Values	Level of Measurement	Source Register	Disaggregations Available

PART B: Evolve 12 indicators – 2 of each type – based on the variables listed above. Classify them and indicate their num, den, units & importance

Indicator 1	
Type	
Num	
Den	
Units	
Importance	
Indicator 2	
Type	
Num	
Den	
Units	
Importance	
Indicator 3	
Type	
Num	
Den	
Units	
Importance	
Indicator 4	
Type	
Num	
Den	
Units	
Importance	
Indicator 5	
Type	
Num	
Den	
Units	
Importance	
Indicator 6	
Type	
Num	
Den	
Units	
Importance	
Indicator 7	
Type	
Num	
Den	
Units	
Importance	

Indicator 8	
Type	
Num	
Den	
Units	
Importance	
Indicator 9	
Type	
Num	
Den	
Units	
Importance	
Indicator 10	
Type	
Num	
Den	
Units	
Importance	
Indicator 11	
Type	
Num	
Den	
Units	
Importance	
Indicator 12	
Type	
Num	
Den	
Units	
Importance	

5. Data Quality Assessment

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Outline

- What is Data Quality?
- Why DQA? Importance of DQ
- Reasons for poor DQ
- Conclusions from DQA
- Attributes of DQ

GIGO

GARBAGE IN GARBAGE OUT

Data Quality

Services: the REAL world

In the *real world*, project activities are implemented in the field. These activities are designed to produce results that are quantifiable.

Data: the INFORMATION SYSTEM

An information system represents these activities by collecting the results that were produced and mapping them to a recording system.

Data Quality: How well the **information system** represents the **real world**




ATTEMPTING DATA QUALITY ASSESSMENT & VALIDATION MAY NOT ALWAYS ENSURE QUALITY, BUT, AT LEAST MAKES ONE AWARE OF POOR QUALITY, IF ANY, WHILE DRAWING INFERENCE !!

&

GIVES AN OPPORTUNITY TO IMPROVE IT

Why DQA?

- An essential step before actual analysis
  - It is careful consideration of data collected, before embarking on analysis
- To help preparation of the *working* dataset/ database
  - To be used for further analysis
- It may open up further questions
  - E.g. HIV positivity rates from ICTC & HSS at same site, different from each other needs exploration of both datasets
- May require further more analysis
  - Eg. Settling an outlier identified in data
- Empowers the investigator to either adjust, correct, or totally reject the data for consideration in final dataset
  - Ability to make an informed decision

Why Is It Important to Have Quality Data?

- Accountability for funding and results reported increasingly important
- Quality data needed at program level for decision making
- Quality Data needed for any sort of operations research
- Quality data → Robust evidence → Appropriate policy

Reasons for Poor Data Quality?

- Not so well structured or well designed data collection tools
- Ill-framed questions
- Untrained personnel
- Biases
- Documentation errors
- Data entry errors
- Compilation errors
- Fraudulent practices

Common Errors during data collection

- Collecting incorrect information (e.g. information collected not as per the indicator definition)
- Collecting information from incorrect or unauthentic document
- Collecting information for incorrect reporting period
- Documenting the information improperly (e.g. making inadequate/rough notes which become difficult to comprehend later)
- Error in consolidating the data before entering into SIMS

Common Errors during data entry

- Entering data incorrectly (e.g. typing 10 instead of 01)
- Entering data against the wrong indicator
- Incomplete data entry (e.g. cells missed in the data entry format)
- Improper storing of offline data entry formats (e.g. not maintaining files properly in a folder, no or irregular data backup, etc.)
- Not doing version control- so that outdated files are uploaded (i.e. not uploading the latest version after revision)

Common Errors during data validation

- Inbuilt validation check signals are not addressed and corrected
- Additional validation checks like checking outliers/out of range and missing values for indicators, consistency checks across different indicators
- Errors are made in cleaning the data even as validation checks are conducted

Data Quality - At Every Step of the Way

- Designing questionnaire/ Registers/ Tools
- Training of data collectors/ facility staff
- Testing the tools
- Recurrent training
- Developing an operational manual for questionnaire to:
  - Standardize the way of asking
  - Standardize the definitions

Conclusions from DQA

- Data quality issues – Present/ Absent
- Level of data quality – Good/ Average/ Poor
- Identified specific quality issues
- Localise the quality issue (indicator; year/month; site/facility)
- Decisions about the quality issues
  - Accept as they are (Decide on acceptance limits)
  - Verify & Correct from original source (Decide on feasibility)
  - Impute, Adjust or Replace data (Decide the methods)
  - Reject the data/ Exclude from analysis (Decide on rejection limits)

Reject/ Exclude data, when…

- The quality issue cannot be corrected
- It is a true reflection of the reality, but adversely affects the analysis
- The quality issue cannot be adjusted

Impute missing data, when…

- Leaving blank affects the power of analysis
- A large segment of data is missing
- A small segment of data is missing & you don't want to have blank cells
- The variable or the case/ facility/ unit is important for analysis that you don't want to exclude, to avoid loss of other important attributes

Key Attributes of Data Quality

A PIT with four Corners
A P I T CCCC

- Availability/ Reporting Status
- Completeness
- Correctness/ Accuracy/ Validity
- Consistency
  - Consistency over time
  - Consistency – Internal/ Reliability
  - Consistency – External/ Validity/ Representativeness
- Precision
- Others
  - Timeliness
  - Integrity
  - Confidentiality

DQA needs Most Granular Data

- DQA should be performed on the most granular data; not on aggregated data – Facility-wise, month/quarter-wise data; Individual-level data; Not on annual, province level data
- Granular data allows you to localise the error/ drill down to the exact issue
- Granular data allows you to fix the exact issue, without altering the other data which is good
- Recommended to do DQA for programme data at facility or district level; NOT ABOVE

Availability/ Reporting Status

Availability of relevant indicators in reporting format

- Facility level (All RUs in the State):
  - No. of months in a year the RU has reported
  - Expressed as % out of 12
  - Categorised as 0-24%, 25-49%, 50-74% & 75-100%
- District & State levels:
  - Total no. of RU-Months reported
  - Expressed as % out of total no. of RU-Months (Tot no. of RUs in district or state x 12)

Summarise

S.No.	Facility Name	Facility Type	No. of months reported: 2015-16	No. of months reported: 2016-17	Reporting %: 2015-16	Reporting %: 2016-17
1	Facility 1	STD Clinic	10	12	83	100
2	Facility 2	STD Clinic	6	8	50	66
3	Facility 3					

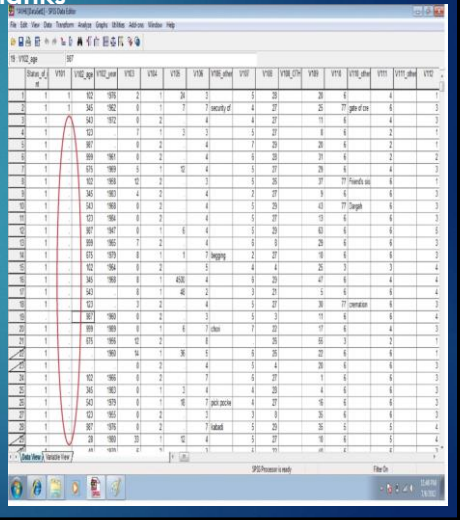
How to assess & conclude?

- Need compiled dataset of one or more facilities
- Calculate Reporting Status
- Identify the non-reported months
- Try to obtain the non-reported data
- Decide acceptance limits:
  - > 75% reporting status in a year: Include in analysis
  - < 75% reporting status in a year: Exclude the facility from analysis
- Impute the missing data if within acceptance limits
  - With average of other months/ neighbouring months

Completeness

- Completeness of data means that all variables for all reporting units are being collected and reported; Reporting formats are completely filled, without any blanks

Number of clients receiving pre-test counselling / information		
ICTC Clients (excluding Pregnant women)	ICTC Clients - Pregnant women	Total ICTC
13365	11835	25200
		0
166	14	180
58	801	859



Completeness

No. of months the report is completely filled for the specified indicators

- Facility level (All RUs in the State):
  - No. of months in a year the RU has reported completely for the specified indicators
  - Expressed as % out of the no. of months in a year the RU has reported
- District & State level:
  - No. of RU-Months that have reported completely for the relevant indicators
  - Expressed as % out of total no. of RU-Months reported

How to assess & conclude?

- Identify the variables needed from the dataset for the analysis.
- Review completeness for the identified variables for each facility
- Identify the facilities & months where the data is missing
- Try to obtain the missing data from the registers
- Decide acceptance limits:
  - > 75% completeness in a year: Include in analysis
  - < 75% completeness in a year: Exclude the facility from analysis
- Impute the missing data if within acceptance limits

Summarise

S.No.	Facility Name	Facility Type	No. of months reported	No. of months reported completely	Completeness %
1	Facility 1		10	10	100
2	Facility 2		14	7	50
3	Facility 3				

Correctness/ Accuracy

- Also sometimes known as validity (Valid data vs Invalid data)
- Accurate data are considered correct: the data measures what it is intended to measure
- Range checking/matching is conducted to check for accuracy
- Example:
  - Recording Male STI syndrome (scrotal swelling) in a female STI syndrome data cell
  - Mentioning age as 234 years (could be 23 or 34)
  - Typing 39 instead of 93 (Typo errors)
  - Ticking the wrong code on the questionnaire

Statistics

Age

N	Valid	129
	Missing	0

Age

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	46	11	8.5	8.5	8.5
	47	8	6.2	6.2	14.7
	48	22	17.1	17.1	31.8
	49	3	2.3	2.3	34.1
	50	21	16.3	16.3	50.4
	51	4	3.1	3.1	53.5
	52	7	5.4	5.4	58.9
	53	3	2.3	2.3	61.2
	54	1	.8	.8	62.0
	55	6	4.7	4.7	66.7
	58	3	2.3	2.3	69.0
	59	1	.8	.8	69.8
	60	5	3.9	3.9	73.6
	62	5	3.9	3.9	77.5
	63	1	.8	.8	78.3
	64	1	.8	.8	79.1
	88	1	.8	.8	79.8
	102	4	3.1	3.1	82.9
	123	4	3.1	3.1	86.0
	345	4	3.1	3.1	89.1
	543	4	3.1	3.1	92.2
	675	3	2.3	2.3	94.6
	987	4	3.1	3.1	97.7
	999	3	2.3	2.3	100.0
	Total	129	100.0	100.0	

Correctness (1)

Consistency b/w registers & reports and b/w two registers for each relevant indicator in selected months

- Facility level:
  - Verification Ratio - b/w registers & reports (Max value – 1)
    - Total no. in the source register in the selected month/ Total no. reported in the monthly report of the selected month
  - Cross-check consistency - b/w two registers (Max value – 100%)
    - No. of cases taken from Register 1 also found in Register 2/ Total no. of cases taken from Register 1
  - Reverse cross-check consistency - b/w two registers
    - No. of cases taken from Register 2 also found in Register 1/ Total no. of cases taken from Register 2

Correctness (2)

Consistency b/w registers & reports and b/w two registers for each relevant indicator in selected months

- District & State levels:
  - No. of RUs that have Verification Ratio of 1 in the selected month for the selected indicator
    - Expressed as % out of total no. of RUs where Trace & Verify is carried out in the district/state
  - No. of RUs that have Cross-check consistency of ≥ 90% in the selected month for the selected indicator
    - Expressed as % out of total no. of RUs where cross-check is carried out in the district/state
  - No. of RUs that have Reverse Cross-check consistency of ≥ 90% in the selected month for the selected indicator
    - Expressed as % out of total no. of RUs where reverse cross-check is carried out in the district/state

How to assess and conclude?

- Conduct range checks for all numeric variables & identify out of range values
- Errors within the range & text data errors can be identified by
  - Verification & Cross-check with registers – Gold standard; Random checks of a few data points, if not all
  - Identifying & verifying implausible data – data that does not seem to be correct/ not likely to be correct
  - Identifying implausible combinations of variables – e.g. age-specific variables & values (age & marital status, age & occupation, age & education), gender-specific variables (female & house-wife, female & received ANC services), etc.
  - Identifying data points that stand out from the rest/ looks like an outlier

How to assess and conclude?

- Highlight the incorrect data points
- Delete confirmed errors & exclude from analysis
- Impute, adjust or replace errors by an appropriate method

Summarise

Correctness		Trace & Verify			Cross Check		
Facility Name	Facility Type	No. in Source Register	No. in Monthly Report	Verification Ratio	No. of cases taken from Register 1	No. of cases found in Register 2	Cross-check Consistency %
Facility 1	STD Clinic	10	10	1.00	10	10	100
Facility 2	STD Clinic	14	15	0.93	10	8	80
Facility 3	STD Clinic	9	5	1.80	10	5	50

Consistency - Internal

- Also known as Reliability
- Refers to the degree of similarity of information obtained when the data are measured and collected in the same method every time
- Can be checked by comparing reported data with different response checks or other primary sources like random respondent check.
- Example:
  - Number of individuals tested: "received post-test counseling" or "collected tests within 7 days"
  - While collecting data on male sexual behavior the number of regular sexual partners is reported to be more than total number of sexual partners in last three months
  - Respondent says s/he is not aware of HIV but states later that s/he has availed HIV testing services.
  - The birth date and age of the respondent are inconsistent

Consistency - Internal

- Example:
  - Number of individuals tested & received pre-test counseling; received post-test counseling & collected test results within 7 days
  - While collecting data on male sexual behavior the number of regular sexual partners is reported to be more than total number of sexual partners
  - Respondent says s/he is not aware of HIV but states later that s/he has availed HIV testing services.
  - The birth date and age of the respondent are inconsistent

Outliers

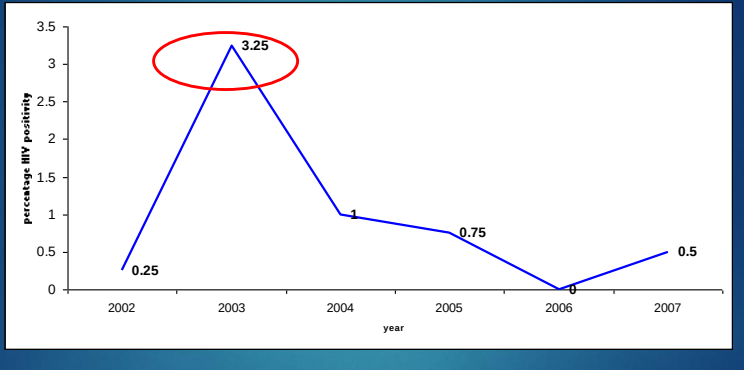
- Outliers are disturbing...it is best to recognize them and make an informed decision to include or exclude the value..
- Outlier: specific values that differ very markedly from the other 'usual' values
- These specific values have an unduly large effect on statistics and may need special handling
- Their presence makes it difficult to draw valid conclusion
- Ideally, all analysis should be performed with and without including the outlier value
  - There might, in fact, be a true reason for the apparently strange data that should not be ignored

Consistency over time

Absence of outliers for each relevant indicator

- Facility level
  - No. of months in a year where there are no outliers in monthly data
  - Expressed as % out of the no. of months in a year the RU has reported
- District & State levels
  - Tot no. of RU-Months where there are no outliers
  - Expressed as % out of total no. of RU-Months reported

Lets see what the trend line looks like...



Trend among women attending ANC clinics at District A, HSS, 2003-07

Compare numbers tested (Target ANC=400, others=250)
HIV positives among ANC attendees, by site and year

District X	HIV Sentinel Surveillance (ANC)		
Year	Number Tested	Number Positive	% positive
2002	400	1	0.25
2003	400	13	3.25
2004	400	4	1
2005	400	3	0.75
2006	400	0	0
2007	400	2	0.5

- Number tested is 400 across years (if <75%, disregard)
- The site is a consistent site for last six years

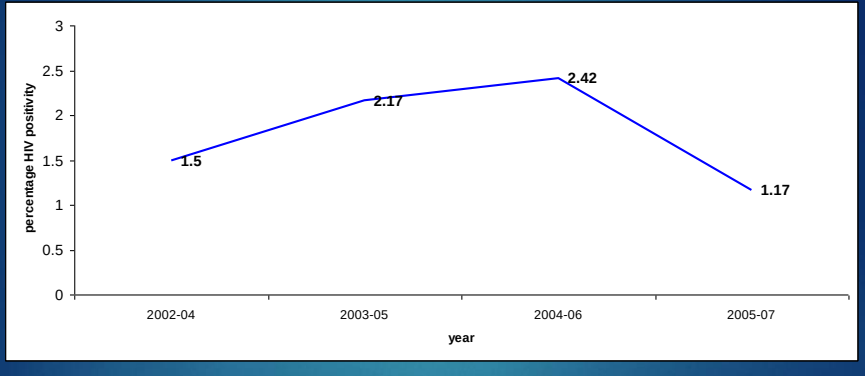
Another example

- While examining trend of HIV positivity among STD clinic attendees, you identify an outlier in one district in one year
- On drilling down, you identify that the outlier is coming from one clinic in the district
- On further drilling down, you identify that the outlier is coming from one quarter data reported by that clinic in that year
- Fixing this one data point will fix the trend

How to assess & conclude

- Identify the outlier
  - Plot a trend line & simple eyeballing of a trend line
  - Build confidence intervals – Identify non-overlapping confidence intervals
  - Figures falling outside 3 standard deviations
- Investigate the outlier
  - Look for data entry errors
  - Examine if the sample size is very low leading to large data fluctuations
  - Look for clustering by day or place or site
  - Check if the profile of the beneficiaries is different from the rest
  - For testing data, look for any consecutive positives due to probable lab contaminations
- Adjust the outlier – Mean imputation; 3-yr moving average

The earlier graph would look something like this by applying 3 years moving average

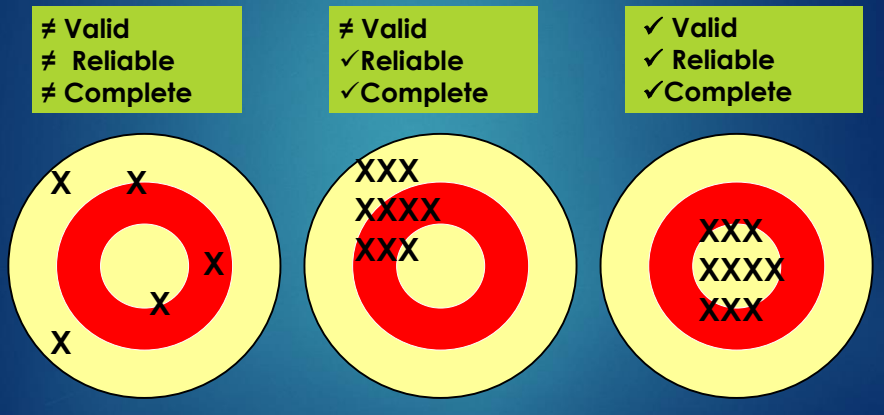


Trend among women attending ANC clinics at District A, HSS, 2003-07 (using 3 years' moving average)

Consistency – External Validity

- External consistency or Validity refers to the representativeness of the data to the population from which it is collected
- This is more applicable to survey data
- For clinic based data, this is not directly applicable as the segment of population who attend the clinic are not exactly same as the general population; But, they can be used as proxy indicators for specific sub-populations
- If coverage levels are high, near 90%, usually, external validity is assumed to be present
- Data with very low sample sizes/ denominators is usually excluded, due to poor external validity

Good Data is Valid, Reliable & Complete



Precision

- This means that the data have sufficient detail. Precise data collection plans/systems ensure that all data dis-aggregations that are required, are collected. Inadequate data affects the quality of the entire analysis
- Example:
  - Totals for positives put on treatment, but no mention of gender or age disaggregations, or type of key population that it serves
  - In a study exploring causes of gender disparity in availing HIV counseling & testing in a particular district, a data collection plan/system lacks precision if it is not designed to record the attitudes of men & women separately

Timeliness

- Data are timely when they are up-to-date (current), and when the information is available on time.
- Example:
  - Service delivery sites submit the previous month's report at different times in a month instead of by a certain date in the month as per reporting timelines.
  - Data collection on " treatment seeking behavior among migrant men for STI/STDs" starts in 2008 when there is no STD clinic in the migration habitation area and ends in 2012 when there are two STD clinics set up by the new TIs in the catchment area.
  - Data is collected for a period which is not within the study period, or from a period wherein no intervention was being carried out.

Integrity

- Integrity is when data generated/collected are protected from deliberate bias or manipulation for political or personal reasons. Manipulation could be done to benefit the project or program.
- Example:
  - Data would lack integrity if an ART centre inflates the number of PLHIV on ART in order to reach the targets defined by NACO
  - Interviewer deliberately records "no sex happened in last one month" to avoid the section on current sexual behavior
  - During training data collectors are encouraged by the implementers to get positive feedback from beneficiaries on an intervention which is being evaluated

Confidentiality

- Confidentiality means that respondents are assured that their information will be maintained according to national and/or international standards for data. This means that personal data are not disclosed inappropriately, and that data in hard copy and electronic form are treated with appropriate levels of security.

Example:

- Care should be taken to ensure that during different steps of research , confidentiality of data is not breached
- Basic ethical principles of the Belmont Report (1979) should be adhered to- Respect for Persons, Beneficence, Justice
- Registers containing personally identifiable information (PII) should be kept in locked cabinets
- Confidentiality agreements for data collectors

Summing Up…

Data Quality Attribute	Indicator to Assess	Key thing to look for
Availability/ Reporting Status	Reporting %	Non-reporting facilities & quarters
Completeness	Completeness %	Missing Data
Correctness/ Accuracy	Verification Ratio & Cross-check Consistency %	Incorrect entries; Invalid data;
Internal Consistency/ Reliability	--	Cross-checks within data
Consistency over time	% of facilities without outliers over a year	Outliers
External Consistency/ Validity/ Representativeness	--	Sample size & Coverage %
Precision	--	Disaggregation available

6. Data Adjustments & Validation

Ministry of Health Sri Lanka | National STD/AIDS Control Programme | Voluntary Health Services

Steps to Improve Data Quality

- Good understanding of registers, columns, variables & definitions
- Good idea of normal or usual range or pattern of data
- Attention & clarity during data recording in the registers
- Correctness & completeness in registers
- Good understanding of mapping of variables/columns in registers to fields/ cells in the reporting formats
- Accuracy in counting, keeping in mind the defining variables & disaggregations
- Focus during data entry to avoid data entry errors
- Self-check/ Self-review of the entered data for DQ before submitting

Adjustment & Validation

- Adjustment refers to taking action to fix the data quality issues and improve the quality of data
- Adjustment also refers to the actions to make the dataset more suitable and appropriate for analysis
- Validation refers to comparing the quality-controlled data with other data/ other source to ensure that we are dealing with plausible data

Summing Up…DQA

Data Quality Attribute	Indicator to Assess	Key thing to look for
Availability/ Reporting Status	Reporting %	Non-reporting facilities & quarters
Completeness	Completeness %	Missing Data
Correctness/ Accuracy	Verification Ratio & Cross-check Consistency %	Incorrect entries; Invalid data;
Internal Consistency/ Reliability	--	Cross-checks within data
Consistency over time	% of facilities without outliers over a year	Outliers
External Consistency/ Validity/ Representativeness	--	Sample size & Coverage %
Precision	--	Disaggregation available

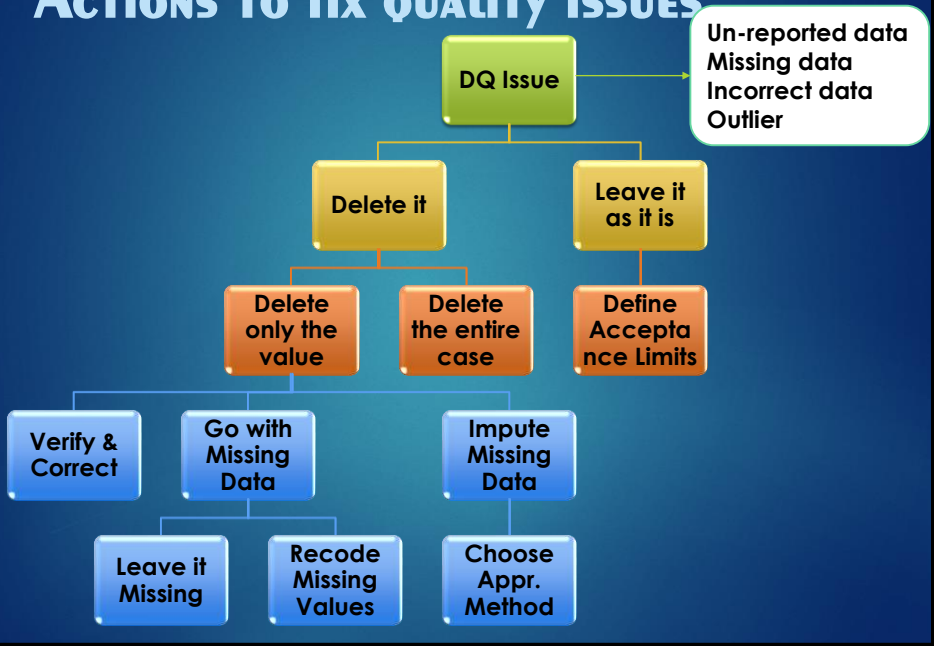
Conclusions from DQA

- Data quality issues – Present/ Absent
- Level of data quality – Good/ Average/ Poor
- Identify specific quality issues
- Localise the quality issue (variable; year/month; site/facility)
- Decisions about the quality issues
  - Accept as they are (Decide on acceptance limits)
  - Verify & Correct from original source (Decide on feasibility)
  - Impute, Adjust or Replace data (Decide the methods)

Actions to fix quality issues

- Leave it as it is, if it is not of great importance or if it does not affect your analysis significantly
- Delete only the concerned data point or Delete the entire case (facility or month or individual); Known as List Deletion
- Verify and correct from the source data/ registers, if possible and feasible
- Impute the missing values or go with missing data/ blank cells if it is acceptable
- Recode the missing values, if required

Actions to fix quality issues

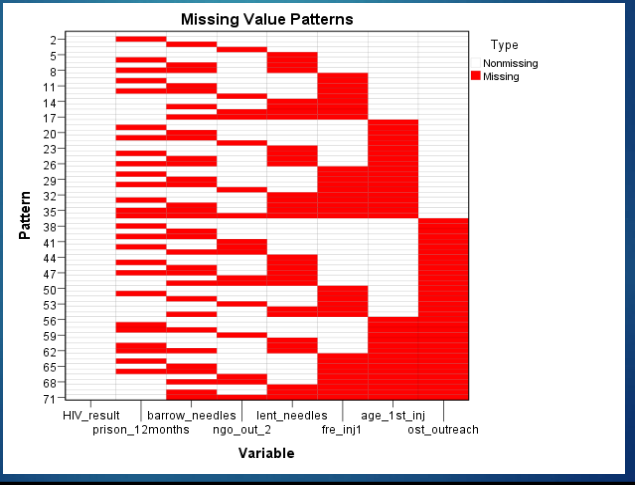


Adjustments

Data Quality Attribute	Key thing to look for	Adjustment
Availability/ Reporting Status	Non-reporting facilities & quarters; Unreported data	Leave, Correct or Impute
Completeness	Missing Data	Leave, Correct or Impute
Correctness/ Accuracy	Incorrect entries; Invalid data;	Leave, Correct or Impute
Internal Consistency/ Reliability	Cross-checks within data	Leave or exclude; Can't do anything
Consistency over time	Outliers	Leave, Correct, Smoothen, adjust or replace
External Consistency/ Validity/ Representativeness	Sample size & Coverage %	Leave or exclude; Can't do anything
Precision	Disaggregation available	Leave; Can't do anything

Analysis of Missing Patterns

- Missing At Random (MAR)
- Missing Completely at Random (MCAR)
- Missing Not At Random (MNAR)



Impute missing data, when…

- Leaving blank affects the power of analysis
- A large segment of data is missing
- A small segment of data is missing & you don't want to have blank cells
- The variable or the case/ facility/ unit is important for analysis that you don't want to exclude, to avoid loss of other important attributes

Methods of Imputation

- Mean Imputation – Replace with average of all months/ all facilities; average of before & after; moving averages;
- Mode Imputation – Replace with the most common value
- Nearest Neighbourhood Imputation – Replace with the value from the nearest similar case/ facility/ district
- Univariate/ Single Imputation Methods
  - Linear imputation
  - Logistic regression imputation
- Multiple Imputation by Chained Equations (MICE)
  - Using Linear Regression/ Predictive Mean Matching method for scale variables
  - Using Logistic Regression for categorical variables
- More advanced methods also available

Reject/ Exclude data, when…

- The quality issue cannot be corrected
- It is a true reflection of the reality, but adversely affects the analysis
- The quality issue cannot be adjusted

Don't Overkill

- Too much manipulation of original data to be avoided
- Too much of imputation of missing data not advised
- Leave the originality of data as much as possible
- Adjust and impute only where it is of significant value addition for the analysis

Other Data Adjustments

- Inflation or deflation factor
- Adjust for specific demographic variables. E.g. Age adjustment
- Standardisation – Adjust to a standard population
- Correction factors
- Calibration

Validation of QC Data

- Compare with external gold standard – Values/ Distribution/ Patterns
- Compare with averages from similar complete data for other years, states, countries, etc
- Validate against known epidemiological and programmatic understanding of the issue and expert consensus

Exercise Dataset

Sno	Province	District	Month	Name of the Facility	Total tested during the month	Total clients received post-test and results	Total ANC positives during the month	Total ANC Positives linked to ART during the month	Total Non ANC positives during the month	Total Non ANC clients linked to ART during the month	Total Positives (ANC+Non ANC)	Positivity Rate (%)	Total Positive clients linked to ART(ANC+Non ANC)	Linkage loss during the month
1	WP	Dist1	JAN 18	ICTC AMALAPURAM	702		0	0	14	12				
2	WP	Dist1	JAN 18	ICTC RAMACHANDRAPURM	483		1	1	12	11				
3	WP	Dist1	JAN 18	ICTC TUNI	652		0	0	10	9				
4	WP	Dist1	JAN 18	ICTC PEDDAPURAM	437		2	2	16	22				
5	WP	Dist1	JAN 18	ICTC DH – RAJAMEHINDRAVARAM	742		0	0	76	69				
6	WP	Dist1	JAN 18	ICTC RMC KAKINADA	722		0	0	46	44				
7	WP	Dist1	JAN 18	ICTC TBIDH GGH KAKINADA	583		0		23	18				
8	WP	Dist3	JAN 18	SA AH - BAPATLA (P&V)	815		0		154	15				
9	WP	Dist3	JAN 18	SA AH - NARSARAOPET (P&V)	1243		2		34	32				
10	WP	Dist3	JAN 18	SA CHC - CHILAKALURIPETA (V)	590		2	2	7	7				
11	WP	Dist3	JAN 18	SA CHC - MACHERLA (V)	350			0	5	7				
12	WP	Dist3	JAN 18	SA CHC - VINUKONDA (I)	315		1	1	9	8				
13	WP	Dist3	JAN 18	SA DH - TENALI (V)	673			0	39	35				
14	WP	Dist3	JAN 18	SA GGH - GUNTUR (V)	1480			0	54	46				
15	WP	Dist3	JAN 18	SA TBH - GUNTUR (I)	422			0	15	13				
16	WP	Dist2	JAN 18	SA AH - GUDIVADA (P&V)	1089		0	0	10	8				
17	WP	Dist2	JAN 18	SA AH - NUZIVEEDU (P&V)	874		0	0	14	13				
18	WP	Dist2	JAN 18	SA CHC - NANDIGAMA-OLD (V)	401		1	3	20	20				
19	WP	Dist2	JAN 18	SA CHC - RAJIVNAGAR-U (V)	602		0	0	17	15				
20	WP	Dist2	JAN 18	SA DH - MACHILIPATNAM (V)	508		0	0	12	15				
21	WP	Dist2	JAN 18	SA GGH - SIDDHARTHA MEDICAL COLLEGE (V)	1670		0	0	78	57				
22	WP	Dist2	JAN 18	SA GGH - VIJAYAWADA (P)	1022		1	1	3	3				
23	NP	Dist4	JAN 18	Dr.R.N.Cooper	1531	1455	0	0	16	9				
24	NP	Dist4	JAN 18	Shatabdi Govandi	1186	1112	0	0	15	12				
25	NP	Dist4	JAN 18	KEM Hos.	3961	3558	0	0	49	39				
26	NP	Dist4	JAN 18	LTMGH	2828	2494	0	0	36	28				
27	NP	Dist4	JAN 18	J.J. Hos.	3429	3237	2	3	48	30				
28	NP	Dist4	JAN 18	Nair Hos.	2046	1978	1	1	31	43				
29	NP	Dist4	JAN 18	G.T. Hospital	987	1112	0	0		8				
30	NP	Dist4	JAN 18	H. Bhagawati Hospital	2015	1845	0	0	24	8				
31	NP	Dist4	JAN 18	Siddhartha Nagar Hospital	749	704	0	0	13	6				
32	NP	Dist4	JAN 18	Rajawadi Hospital	1641	1494	1	1	14	8				
33	NP	Dist5	JAN 18	A F M C, PUNE VCTC (SDLIC2725000220983)	1386	1369	0	0	22	13				
34	NP	Dist5	JAN 18	BHARATHI HOSPITAL & MEDICAL COLLEGE (SDLIC2725000201912)	405	405	0	0	5					
35	NP	Dist5	JAN 18	BHOSARI HOSPITAL (SDLIC2725000202045)	712	683	0	0	9	2				
36	NP	Dist5	JAN 18	BJMC PUNE VCTC (SDLIC2725000202049)	0	0	0	0	0	0				
37	NP	Dist5	JAN 18	CHEST & DISTRICT HOSPITAL AUNDH (SDLIC2725000201797)	702	692	0	0	14	14				
38	NP	Dist5	JAN 18	DR.D Y PATIL MEDICAL COLLEGE & HOSPITAL (SDLIC2725000202047)	327	327	0	0	5	5				
39	NP	Dist5	JAN 18	MAHARASHTRA AROGYA MANDAL (SDLIC2725000201911)	436	436	0		0	0				
40	NP	Dist5	JAN 18	MIMER MEDICAL COLLEGE (SDLIC2725000202046)	714	714	0		4	4				
41	NP	Dist5	JAN 18	RAJIV GANDHI HOSPITAL (KHANSAHEB PESTANJI MATERNITY HOME) (SDLIC2725000216075)	0	0	0		0	0				
42	NP	Dist5	JAN 18	RH MANCHAR (SDLIC2725000202042)	630	630	0	0	6	6				
43	NP	Dist5	JAN 18	RH SHIRUR (SDLIC2725000220634)	244	244	0	0	3	0				
44	NP	Dist5	JAN 18	SDH INDAPUR (SDLIC2725000220633)	287	284	1	1	5	1				
45	NP	Dist5	JAN 18	SILVER JUBILEE GOVT RURAL HOSPITAL,BARAMATI (SDLIC2725000216081)	239	239	0	0	16	16				
46	NP	Dist5	JAN 18	SMT KASHIBAI NAVALE HOSPITAL PPP (SDLIC2725000222763)	1659	1542	0	0	15	10				
47	NP	Dist5	JAN 18	SONAWANE HOSPITAL PMC (SDLIC2725000201798)	398	400	0	0	8	0				
48	NP	Dist5	JAN 18	YASHWANTRAO CHAVAN MEMORIAL HOSPITAL (SDLIC2725000203522)	854	847	2	2	35	29				
49	NP	Dist6	JAN 18	V S GENERAL HOSPITAL I (SDLIC2721001414877)	366	366	0	0	6	6				
50	NP	Dist6	JAN 18	MEENATAI THAKARE THANE(SHIVAJI NAGAR) (SDLIC2721001420585)	198	197	1	1	2	2				
51	NP	Dist6	JAN 18	DR A G JOSHI HOSPITAL (SDLIC2721001421951)	152	152	0	0		2				
52	NP	Dist6	JAN 18	RAJIV GANDHI MEDICAL COLLEGE, KALVA (SDLIC2721001414883)	990	1157		1	15	15				
53	NP	Dist6	JAN 18	SUTIKAGRUH HOSPITAL (SDLIC2721001421955)	413	407		1	2	2				
54	NP	Dist6	JAN 18	RUKMINIBAI HOSPITAL (SDLIC2721001416388)	710	707	0	0	14	14				

Sno	Province	District	Month	Name of the Facility	Total tested during the month	Total clients received post-test and results	Total ANC positives during the month	Total ANC Positives linked to ART during the month	Total Non ANC positives during the month	Total Non ANC clients linked to ART during the month	Total Positives (ANC+Non ANC)	Positivity Rate (%)	Total Positive clients linked to ART(ANC+Non ANC)	Linkage loss during the month
55	NP	Dist6	JAN 18	SHASHTRINAGAR GENERAL HOSPITAL (SDLIC2721001416392)	458	458	0	0	6	6				
56	NP	Dist6	JAN 18	CENTRAL HOSPITAL ULHASNAGAR (SDLIC2721001401795)	765	696	0	0	135	13				
57	NP	Dist6	JAN 18	BALKUM (CHHAYA HOSPITAL, ABERNATH) (SDLIC2721001416961)	689	469	0	0	7	7				
58	NP	Dist6	JAN 18	HEALTH CENTRE BHAYANDAR-W (SDLIC2721001420583)	593	592	0	0	3					
59	NP	Dist6	JAN 18	HEALTH CENTRE MIRA ROAD (SDLIC2721001421947)	586	558	2	2	13	13				
60	NP	Dist6	JAN 18	I G M HOSPITAL, BHIWANDI (SDLIC2721001416389)	404	411	3	3	28	28				
61	NP	Dist6	JAN 18	NAVI MUMBAI (GEN.HOSPITAL),VASHI (SDLIC2721001401796)	1193	1185	0	0	30	30				
62	NP	Dist6	JAN 18	RAJMATA JIJAU MATA VAL RUGNALAYA,AIROLI (SDLIC2721001416959)	1010	997	0	0	7	5				
63	NP	Dist6	JAN 18	PAD. DR D Y PATIL MEDICAL COLLEGE & HOSPITAL, NERUL,NAVI MUMBAI (SDLIC2721001414879)	832	831	0		2	2				
64	NP	Dist6	FEB 18	V S GENERAL HOSPITAL I (SDLIC2721001414877)	330	330	0		10	10				
65	NP	Dist6	FEB 18	MEENATAI THAKARE THANE(SHIVAJI NAGAR) (SDLIC2721001420585)	112	112	0		0	0				
66	NP	Dist6	FEB 18	DR A G JOSHI HOSPITAL (SDLIC2721001421951)	177	176	0	0	2	2				
67	NP	Dist6	FEB 18	RAJIV GANDHI MEDICAL COLLEGE, KALVA (SDLIC2721001414883)	990	960	1	1	15	13				
68	NP	Dist6	FEB 18	SUTIKAGRUH HOSPITAL (SDLIC2721001421955)	378	375	0	0	4	4				
69	NP	Dist6	FEB 18	RUKMINIBAI HOSPITAL (SDLIC2721001416388)	633	629	0	0	10	9				
70	NP	Dist6	FEB 18	SHASHTRINAGAR GENERAL HOSPITAL (SDLIC2721001416392)	444	441		1	7	7				
71	NP	Dist6	FEB 18	CENTRAL HOSPITAL ULHASNAGAR (SDLIC2721001401795)	765	726	0	0	24	33				
72	NP	Dist6	FEB 18	BALKUM (CHHAYA HOSPITAL, ABERNATH) (SDLIC2721001416961)	681	387	0	0	0	0				
73	NP	Dist6	FEB 18	HEALTH CENTRE BHAYANDAR-W (SDLIC2721001420583)	600	498	0	0	4	4				
74	NP	Dist6	FEB 18	HEALTH CENTRE MIRA ROAD (SDLIC2721001421947)	513	506	0	0	10	10				
75	NP	Dist6	FEB 18	I G M HOSPITAL, BHIWANDI (SDLIC2721001416389)	293	292	1	2	30	30				
76	NP	Dist6	FEB 18	NAVI MUMBAI (GEN.HOSPITAL),VASHI (SDLIC2721001401796)	1014	1006	0	0		23				
77	NP	Dist6	FEB 18	RAJMATA JIJAU MATA VAL RUGNALAYA,AIROLI (SDLIC2721001416959)	883	862	0	0	3	3				
78	NP	Dist6	FEB 18	PAD. DR D Y PATIL MEDICAL COLLEGE & HOSPITAL, NERUL,NAVI MUMBAI (SDLIC2721001414879)	699	693	0	0	4	4				
79	NP	Dist5	FEB 18	A F M C, PUNE VCTC (SDLIC2725000220983)	1082	1122	0	0	9	7				
80	NP	Dist5	FEB 18	BHARATHI HOSPITAL & MEDICAL COLLEGE (SDLIC2725000201912)	190	190	1	1	4	4				
81	NP	Dist5	FEB 18	BHOSARI HOSPITAL (SDLIC2725000202045)	282	280	3	3	3	3				
82	NP	Dist5	FEB 18	BJMC PUNE VCTC (SDLIC2725000202049)	951	923	0	0	77	44				
83	NP	Dist5	FEB 18	CHEST & DISTRICT HOSPITAL AUNDH (SDLIC2725000201797)	348	348	8	8	14	14				
84	NP	Dist5	FEB 18	DR.D Y PATIL MEDICAL COLLEGE & HOSPITAL (SDLIC2725000202047)	236	236	1	1	10	10				
85	NP	Dist5	FEB 18	MAHARASHTRA AROGYA MANDAL (SDLIC2725000201911)	225	225	7	7	10	10				
86	NP	Dist5	FEB 18	MIMER MEDICAL COLLEGE (SDLIC2725000202046)	364	364	4	4	0	0				
87	NP	Dist5	FEB 18	RAJIV GANDHI HOSPITAL (KHANSAHEB PESTANJI MATERNITY HOME) (SDLIC2725000216075)	0	0	1	1	0	0				
88	NP	Dist5	FEB 18	RH MANCHAR (SDLIC2725000202042)	301	301	1	1	5					
89	NP	Dist5	FEB 18	SDH INDAPUR (SDLIC2725000220633)	152	152	2	2	5	2				
90	NP	Dist5	FEB 18	SILVER JUBILEE GOVT RURAL HOSPITAL,BARAMATI (SDLIC2725000216081)	119	119	2	2	14	14				
91	NP	Dist5	FEB 18	SMT KASHIBAI NAVALE HOSPITAL PPP (SDLIC2725000222763)	1136	1068	15	15	23	11				
92	NP	Dist5	FEB 18	SONAWANE HOSPITAL PMC (SDLIC2725000201798)	240	234	2	2	66	5				
93	NP	Dist5	FEB 18	YASHWANTRAO CHAVAN MEMORIAL HOSPITAL (SDLIC2725000203522)	356	350	14	3	34	32				
94	NP	Dist4	FEB 18	Dr.R.N.Cooper	1352	1320	0	0	16	5				
95	NP	Dist4	FEB 18	Shatabdi Govandi	814	772	1	1	17	23				
96	NP	Dist4	FEB 18	KEM Hos.	3686	3876	33	1	35	21				
97	NP	Dist4	FEB 18	LTMGH	2521	2146	1	1	53	32				
98	NP	Dist4	FEB 18	J.J. Hos.	3112	2951	2	2	55	45				
99	NP	Dist4	FEB 18	Nair Hos.	1772	1727	0	0	35	31				
100	NP	Dist4	FEB 18	G.T. Hospital	822	806	0	0	14	1				
101	NP	Dist4	FEB 18	H. Bhagawati Hospital	1771	1682	1	1		6				
102	NP	Dist4	FEB 18	Siddhartha Nagar Hospital	596	571		0	19	11				
103	NP	Dist4	FEB 18	Rajawadi Hospital	1395	1309	2	1	9	6				
104	WP	Dist1	FEB 18	ICTC AMALAPURAM	793		0	0	10	10				
105	WP	Dist1	FEB 18	ICTC RAMACHANDRAPURM	544		1	1	20	20				
106	WP	Dist1	FEB 18	ICTC TUNI	6582		0	0	14	12				
107	WP	Dist1	FEB 18	ICTC PEDDAPURAM	522		2	2	19	15				
108	WP	Dist1	FEB 18	ICTC DH – RAJAMEHINDRAVARAM	977		0	0	104	90				
109	WP	Dist1	FEB 18	ICTC RMC KAKINADA	807		0	0	44	44				

Sno	Province	District	Month	Name of the Facility	Total tested during the month	Total clients received post-test and results	Total ANC positives during the month	Total ANC Positives linked to ART during the month	Total Non ANC positives during the month	Total Non ANC clients linked to ART during the month	Total Positives (ANC+Non ANC)	Positivity Rate (%)	Total Positive clients linked to ART(ANC+Non ANC)	Linkage loss during the month
110	WP	Dist1	FEB 18	ICTC TBIDH GGH KAKINADA	739		0	0	27	24				
111	WP	Dist3	FEB 18	SA AH - BAPATLA (P&V)	951		0	0	14	14				
112	WP	Dist3	FEB 18	SA AH - NARSARAOPET (P&V)	1352		1	1	31					
113	WP	Dist3	FEB 18	SA CHC - CHILAKALURIPETA (V)	618		1	1	13	13				
114	WP	Dist3	FEB 18	SA CHC - MACHERLA (V)	35		0	0	7	6				
115	WP	Dist3	FEB 18	SA CHC - VINUKONDA (I)	332		0	0	10	8				
116	WP	Dist3	FEB 18	SA DH - TENALI (V)	625		0	0	45	39				
117	WP	Dist3	FEB 18	SA GGH - GUNTUR (V)	1521		0	0	66	76				
118	WP	Dist3	FEB 18	SA TBH - GUNTUR (I)	358		0	0	14	13				
119	WP	Dist2	FEB 18	SA AH - GUDIVADA (P&V)	1068		2	2	124	12				
120	WP	Dist2	FEB 18	SA AH - NUZIVEEDU (P&V)	890		4	2	18	12				
121	WP	Dist2	FEB 18	SA CHC - NANDIGAMA-OLD (V)	574		0	0	15	15				
122	WP	Dist2	FEB 18	SA CHC - RAJIVNAGAR-U (V)	692		0		16	13				
123	WP	Dist2	FEB 18	SA DH - MACHILIPATNAM (V)	668		12		25	24				
124	WP	Dist2	FEB 18	SA GGH - SIDDHARTHA MEDICAL COLLEGE (V)	19842		0	0	91	71				
125	WP	Dist2	FEB 18	SA GGH - VIJAYAWADA (P)	1089		1	1	14	14				
126	WP	Dist2	MAR 18	SA AH - GUDIVADA (P&V)	1316		0	0	15	15				
127	WP	Dist2	MAR 18	SA AH - NUZIVEEDU (P&V)	1146		2	2		37				
128	WP	Dist2	MAR 18	SA CHC - NANDIGAMA-OLD (V)	532		1	1	18	9				
129	WP	Dist2	MAR 18	SA CHC - RAJIVNAGAR-U (V)	384		0	0	15	15				
130	WP	Dist2	MAR 18	SA DH - MACHILIPATNAM (V)	835			0	23	23				
131	WP	Dist2	MAR 18	SA GGH - SIDDHARTHA MEDICAL COLLEGE (V)	2297			0	104					
132	WP	Dist2	MAR 18	SA GGH - VIJAYAWADA (P)	1421		4	5	11	11				
133	WP	Dist3	MAR 18	SA AH - BAPATLA (P&V)	808		1	1	9	9				
134	WP	Dist3	MAR 18	SA AH - NARSARAOPET (P&V)	1334		1	1	26	21				
135	WP	Dist3	MAR 18	SA CHC - CHILAKALURIPETA (V)	531		0	0	13	13				
136	WP	Dist3	MAR 18	SA CHC - MACHERLA (V)	352		0	0	5	3				
137	WP	Dist3	MAR 18	SA CHC - VINUKONDA (I)	513		0	0	9	6				
138	WP	Dist3	MAR 18	SA DH - TENALI (V)	679		0	0	27	24				
139	WP	Dist3	MAR 18	SA GGH - GUNTUR (V)	1598		0	0	78	71				
140	WP	Dist3	MAR 18	SA TBH - GUNTUR (I)	390		0	0	20	18				
141	WP	Dist1	MAR 18	ICTC AMALAPURAM	823		1	1	17	17				
142	WP	Dist1	MAR 18	ICTC RAMACHANDRAPURM	605		1	1	12	12				
143	WP	Dist1	MAR 18	ICTC TUNI	677		0	0	11	11				
144	WP	Dist1	MAR 18	ICTC PEDDAPURAM	477		1	1	12	12				
145	WP	Dist1	MAR 18	PPTCT GGH – KAKINADA	926		2	2	5	5				
146	WP	Dist1	MAR 18	ICTC RMC KAKINADA	767		0	0	36	36				
147	WP	Dist1	MAR 18	ICTC TBIDH GGH KAKINADA	570		0	0	18	17				
148	NP	Dist4	MAR 18	Dr.R.N.Cooper	1518	1259	0	0	17	7				
149	NP	Dist4	MAR 18	Shatabdi Govandi	553	518	2	2	16					
150	NP	Dist4	MAR 18	KEM Hos.	3322	3544	1	1	42	27				
151	NP	Dist4	MAR 18	LTMGH	2009	1782	1	1		23				
152	NP	Dist4	MAR 18	J.J. Hos.	3042	2934	12	0	65	38				
153	NP	Dist4	MAR 18	Nair Hos.	1700	1652	0	0	32	25				
154	NP	Dist4	MAR 18	G.T. Hospital	742	740		0	10	1				
155	NP	Dist4	MAR 18	H. Bhagawati Hospital	1908	2233	1	1	23	3				
156	NP	Dist4	MAR 18	Siddhartha Nagar Hospital	591	555	0	0	6	8				
157	NP	Dist4	MAR 18	Rajawadi Hospital	1463	1332	1	3	8	5				
158	NP	Dist5	MAR 18	A F M C, PUNE VCTC (SDLIC2725000220983)	1059	1048	0	0	18	8				
159	NP	Dist5	MAR 18	BHARATHI HOSPITAL & MEDICAL COLLEGE (SDLIC2725000201912)	403	403	0	0	5	3				
160	NP	Dist5	MAR 18	BHOSARI HOSPITAL (SDLIC2725000202045)	521	513	0	0	6	6				
161	NP	Dist5	MAR 18	BJMC PUNE VCTC	1030	1030	0	0	64	50				
162	NP	Dist5	MAR 18	CHEST & DISTRICT HOSPITAL AUNDH (SDLIC2725000201797)	729	729	1	1	18	17				
163	NP	Dist5	MAR 18	DR.D Y PATIL MEDICAL COLLEGE & HOSPITAL (SDLIC2725000202047)	328	328	0	0	3	3				
164	NP	Dist5	MAR 18	MAHARASHTRA AROGYA MANDAL (SDLIC2725000201911)	441	441	0	0	9	6				

Sno	Province	District	Month	Name of the Facility	Total tested during the month	Total clients received post-test and results	Total ANC positives during the month	Total ANC Positives linked to ART during the month	Total Non ANC positives during the month	Total Non ANC clients linked to ART during the month	Total Positives (ANC+Non ANC)	Positivity Rate (%)	Total Positive clients linked to ART(ANC+Non ANC)	Linkage loss during the month
165	NP	Dist5	MAR 18	MIMER MEDICAL COLLEGE (SDLIC2725000202046)	787	442	0	0	1	1				
166	NP	Dist5	MAR 18	RAJIV GANDHI HOSPITAL (KHANSAHEB PESTANJI MATERNITY HOME) (SDLIC2725000216075)	53	53	0	0	0	0				
167	NP	Dist5	MAR 18	RH MANCHAR (SDLIC2725000202042)	498	508	0	0	3	3				
168	NP	Dist5	MAR 18	SDH INDAPUR (SDLIC2725000220633)	185	185	34	0	2	2				
169	NP	Dist5	MAR 18	SILVER JUBILEE GOVT RURAL HOSPITAL,BARAMATI (SDLIC2725000216081)	271	271	0	0	12	12				
170	NP	Dist5	MAR 18	SMT KASHIBAI NAVALE HOSPITAL PPP (SDLIC2725000222763)	1498	1551	1	1	185	14				
171	NP	Dist5	MAR 18	SONAWANE HOSPITAL PMC (SDLIC2725000201798)	463	450	1	1	6	3				
172	NP	Dist5	MAR 18	YASHWANTRAO CHAVAN MEMORIAL HOSPITAL (SDLIC2725000203522)	990	994	0	0	39	38				
173	NP	Dist6	MAR 18	V S GENERAL HOSPITAL I (SDLIC2721001414877)	271	271	0	0	6					
174	NP	Dist6	MAR 18	MEENATAI THAKARE THANE(SHIVAJI NAGAR) (SDLIC2721001420585)	146	146	0	0	5	5				
175	NP	Dist6	MAR 18	DR A G JOSHI HOSPITAL (SDLIC2721001421951)	146	146	0	0		3				
176	NP	Dist6	MAR 18	RAJIV GANDHI MEDICAL COLLEGE, KALVA (SDLIC2721001414883)	1045	1019	0	0	25	24				
177	NP	Dist6	MAR 18	SUTIKAGRUH HOSPITAL (SDLIC2721001421955)	462	452	1	1	4	4				
178	NP	Dist6	MAR 18	RUKMINIBAI HOSPITAL (SDLIC2721001416388)	667	663	0	0	18	18				
179	NP	Dist6	MAR 18	SHASHTRINAGAR GENERAL HOSPITAL (SDLIC2721001416392)	397	398		0	7	9				
180	NP	Dist6	MAR 18	CENTRAL HOSPITAL ULHASNAGAR (SDLIC2721001401795)	762	807	1	1	35	34				
181	NP	Dist6	MAR 18	BALKUM (CHHAYA HOSPITAL, ABERNATH) (SDLIC2721001416961)	705	705	0	0	6	6				
182	NP	Dist6	MAR 18	HEALTH CENTRE BHAYANDAR-W (SDLIC2721001420583)	540	538	0	0	10	10				
183	NP	Dist6	MAR 18	HEALTH CENTRE MIRA ROAD (SDLIC2721001421947)	576	544	0	0	9	9				
184	NP	Dist6	MAR 18	I G M HOSPITAL, BHIWANDI (SDLIC2721001416389)	622	619	1	1	30	30				
185	NP	Dist6	MAR 18	NAVI MUMBAI (GEN.HOSPITAL),VASHI (SDLIC2721001401796)	1239	1332	0	0	29	28				
186	NP	Dist6	MAR 18	RAJMATA JIJAU MATA VAL RUGNALAYA,AIROLI (SDLIC2721001416959)	0	0	2	2	4	4				
187	NP	Dist6	MAR 18	PAD. DR D Y PATIL MEDICAL COLLEGE & HOSPITAL, NERUL,NAVI MUMBAI (SDLIC2721001414879)	589	586	3	2	4	4				
188	WP	Dist1	APR 18	ICTC AMALAPURAM	708		0	0	14	13				
189	WP	Dist1	APR 18	ICTC RAMACHANDRAPURM	525		1	1	11	8				
190	WP	Dist1	APR 18	ICTC TUNI	658		0	0	8	8				
191	WP	Dist1	APR 18	ICTC PEDDAPURAM	447		0	0	16	15				
192	WP	Dist1	APR 18	ICTC DH – RAJAMEHINDRAVARAM	936		0	0	803	62				
193	WP	Dist1	APR 18	ICTC RMC KAKINADA	690		0	0	46	42				
194	WP	Dist1	APR 18	ICTC TBIDH GGH KAKINADA	602			0	37	37				
195	WP	Dist3	APR 18	SA AH - BAPATLA (P&V)	813		1	1	13	13				
196	WP	Dist3	APR 18	SA AH - NARSARAOPET (P&V)	1368		1	1		29				
197	WP	Dist3	APR 18	SA CHC - CHILAKALURIPETA (V)	6713		1	1	18	18				
198	WP	Dist3	APR 18	SA CHC - MACHERLA (V)	379		0	0	6	6				
199	WP	Dist3	APR 18	SA CHC - VINUKONDA (I)	286		2	2	13	10				
200	WP	Dist3	APR 18	SA DH - TENALI (V)	528		0	0	39	35				
201	WP	Dist3	APR 18	SA GGH - GUNTUR (V)	1582		0	0	86	53				
202	WP	Dist3	APR 18	SA TBH - GUNTUR (I)	300		0	0	18	15				
203	WP	Dist2	APR 18	SA AH - GUDIVADA (P&V)	1147		0	0	18					
204	WP	Dist2	APR 18	SA AH - NUZIVEEDU (P&V)	909		1	1	12	12				
205	WP	Dist2	APR 18	SA CHC - AVANIGADDA (V)	275		0	0	5	4				
206	WP	Dist2	APR 18	SA CHC - GANNAVARAM (V)	344		0	2	4	4				
207	WP	Dist2	APR 18	SA CHC - JAGGAIAHPET-OLD (V)	335		2	2	10	10				
208	WP	Dist2	APR 18	SA CHC - KAIKALUR (V)	232		0	0	3	3				
209	WP	Dist2	APR 18	SA CHC - MYLAVARAM-OLD (V)	283		1	1	12	11				
210	WP	Dist2	APR 18	SA CHC - NANDIGAMA-OLD (V)	441		0	0	16	16				
211	WP	Dist2	APR 18	SA CHC - RAJIVNAGAR-U (V)	56		0	0	21	19				
212	WP	Dist2	APR 18	SA CHC - TIRUVURU-OLD (V)	201		0	0	7	7				
213	WP	Dist2	APR 18	SA CHC - VUYYURU (V)	213		12	0	2	2				
214	WP	Dist2	APR 18	SA DH - MACHILIPATNAM (P)	453		0	0	0	0				
215	WP	Dist2	APR 18	SA DH - MACHILIPATNAM (V)	776		0	0	20	19				
216	WP	Dist2	APR 18	SA GGH - SIDDHARTHA MEDICAL COLLEGE (V)	1744		0	0	85	78				
217	WP	Dist2	APR 18	SA GGH - VIJAYAWADA (P)	119		2	2	7	7				
218	WP	Dist2	APR 18	SA PHC - KANKIPADU (Previously NP)	526		0	0	0	3				
219	WP	Dist2	APR 18	SA PVT-MC - PINNAMANENI INSTITUTE OF MEDICAL SCIENCES (I)	472		0	0	5	3				

Sno	Province	District	Month	Name of the Facility	Total tested during the month	Total clients received post-test and results	Total ANC positives during the month	Total ANC Positives linked to ART during the month	Total Non ANC positives during the month	Total Non ANC clients linked to ART during the month	Total Positives (ANC+Non ANC)	Positivity Rate (%)	Total Positive clients linked to ART(ANC+Non ANC)	Linkage loss during the month
220	WP	Dist2	APR 18	SA UPHC - RUDRAPAKA (I)	216		0	0	24	0				
221	WP	Dist1	MAY 18	ICTC AMALAPURAM	843		0	0	12	11				
222	WP	Dist1	MAY 18	ICTC RAMACHANDRAPURM	582		0	0	11	9				
223	WP	Dist1	MAY 18	ICTC TUNI	695		4	4						
224	WP	Dist1	MAY 18	ICTC PEDDAPURAM	573			0	16	11				
225	WP	Dist1	MAY 18	ICTC DH – RAJAMEHINDRAVARAM	1717		2	2	90	66				
226	WP	Dist1	MAY 18	ICTC RMC KAKINADA	733				43	39				
227	WP	Dist1	MAY 18	ICTC TBIDH GGH KAKINADA	473				29	24				
228	WP	Dist2	MAY 18	SA AH - GUDIVADA (P&V)	1159		0	0	11	10				
229	WP	Dist2	MAY 18	SA AH - NUZIVEEDU (P&V)	871		2	2	25	23				
230	WP	Dist2	MAY 18	SA CHC - NANDIGAMA-OLD (V)	420		0	0	12	12				
231	WP	Dist2	MAY 18	SA CHC - RAJIVNAGAR-U (V)	375		0	0	8	8				
232	WP	Dist2	MAY 18	SA DH - MACHILIPATNAM (V)	808		0	0	36	34				
233	WP	Dist2	MAY 18	SA GGH - SIDDHARTHA MEDICAL COLLEGE (V)	1821		0	0	89	75				
234	WP	Dist2	MAY 18	SA GGH - VIJAYAWADA (P)	1288		3	3	12	12				
235	WP	Dist3	MAY 18	SA AH - BAPATLA (P&V)	828		2	2	4	4				
236	WP	Dist3	MAY 18	SA AH - NARSARAOPET (P&V)	1559		0	0	32	32				
237	WP	Dist3	MAY 18	SA CHC - CHILAKALURIPETA (V)	770		0	3	9	9				
238	WP	Dist3	MAY 18	SA CHC - MACHERLA (V)	359		0	0	44	4				
239	WP	Dist3	MAY 18	SA CHC - VINUKONDA (I)	289		23	0	1	1				
240	WP	Dist3	MAY 18	SA DH - TENALI (V)	6410		0	0	50	53				
241	WP	Dist3	MAY 18	SA GGH - GUNTUR (V)	1925		0	0	77	50				
242	WP	Dist3	MAY 18	SA TBH - GUNTUR (I)	386		0	0	23	21				

7. Basic Programmatic Analysis – Measures & Methods

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Basic Measures

- Sum/Total
- Sub-totals
- Percentage
- Distribution
- Min-Max
- Cut-off based
- Top & bottom
- Quartiles
- Average/ Mean
- Median

Computing new variables

- Recoding text to numeric variables
- Converting continuous to categorical variables
- Computing new variables – age groups, percentages

Indicator Estimation

- Number
- Percentage
- Rate
- Ratio

Useful Excel Functions

- Sort & Filter
- Replace
- Go to
- Pivot tables
- V-look Up
- Conditional Formatting
- Remove Duplicates
- Sum, Average, Median, Quartiles

Levels of Measurement

- **Nominal Variable**
  - Categorical/ Discrete data
  - Distinct groups
  - No ordering/ Order is not meaningful
  - Gender, Occupation, Marital status, Colour, Yes/No,
- **Ordinal Variable**
  - Categorical/ Discrete data
  - Distinct groups
  - Ordering meaningful
  - Distance/gap between two level in the order is not meaningful
  - Stages of disease, Severity of disease, Mild-Moderate-Severe, Low-Medium-High, Education, Ranks
- **Interval Variable**
  - Numeric/ Continuous data
  - Distance/gap between two levels can be measured & is meaningful; Subtraction is possible between values
  - No absolute zero/ No Common Reference Point; Zero is not meaningful
  - Temperature, Age groups, Distance b/w two points
- **Ratio Variable**
  - Numeric/ Continuous data
  - Absolute zero is meaningful
  - Levels can be expressed as ratio or number of times of one another
  - Height, Weight, Prevalence rate

Summary Measures

Level of Measurement	Type	Measures of Central Tendency	Measures of Dispersion	Other Tests
Nominal	Categorical/ Discrete	Count/ Frequency, Mode	Distribution	Chi-square, T-test
Ordinal	Categorical/ Discrete	+ Median	+ Percentiles, Quartiles, Inter-quartile range	Chi-square, T-test
Interval	Continuous/ Numeric	+ Mean	+ Standard Deviation	One-way ANOVA
Ratio	Continuous/ Numeric	+ Ratio	+ Coefficient of Variation	Regression

Exercise 6

- Estimate the basic measures from the NSACP dataset you have cleaned in DQA exercise
- Convert continuous to categorical variable
- Estimate four indicators

8. Basic & Advanced Functions in MS Excel

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Overview

- Understanding Excel Layout
- Sheet Visualisation Functions
- Dealing with Rows & Columns
- Cell Formatting
- Cell Navigation
- Paste Special
- Dataset Functions
- Analytic Functions
- Basic Formulae
- Advanced Formulae
- Advanced Operations

Understanding Excel Layout

- Three views – Normal, Print, Pagebreak
- Sheets
- Ribbon & Tabs

Sheet Visualisation Functions

- Adding headers
- Colour cells
- Freeze panes
- Naming & Renaming sheets

Dealing with Rows & Columns

- Inserting Rows & Columns
- Deleting Rows & Columns
- Copying & Pasting Rows & Columns
- Adjusting Column Widths & Row Heights

Cell Formatting

- Understanding Cell formats – General, Number, Date
- Wrap Text
- Adding borders

Cell Navigation

- Using arrow buttons to navigate
- Using Ctrl & Shift buttons with arrow buttons to select cells
- Go To Dialogue Box & Keyboard Shortcut (Ctrl+G)

Paste Special

- Paste values
- Paste formulas
- Paste column widths
- Paste formats
- Transpose
- Link Cells

Dataset Functions

- Sort
- Filter
- Conditional Formatting – Blanks, Duplicates, Text/ Numeric Condition
- Remove Duplicates
- Find & Replace
- Increasing & Decreasing Decimal values

Analytic Functions

- Writing Formulas
- Using Brackets & Commas
- Locking Formulas
- Relative Reference

Basic Formulae

- Count, CountA, Countblank, Countif, Countifs
- Sum, Sumif, Sumifs
- Other Fundamental operations – Subtract, Multiply, Divide
- Percentage & Distribution
- Min-Max
- Cut-off based
- Top & bottom
- Quartiles
- Average/ Mean
- Median

Advanced Formulae

- If function
- Nested If functions – And, Or
- Left, Right, Mid functions
- Join function

Advanced Operations

- Data migration using Vlookup
- Analysis using Pivot Tables
- Creating Graphs
- Data Validation

9. Commonly Used Formulae in MS Excel

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Overview

- General Rules
- Number Functions
  - Count
  - Sum
  - Min-Max
  - Quartiles
  - Mean/Average
  - Median
- Text Functions
  - Join
  - Left, Right & Mid
- Logical Functions
  - And
  - Or
- Special Functions
  - If
  - Vlookup

General Rules of Writing Formulae

- Formula starts with '=' followed by the formula name followed by opening bracket
- Formula can be typed in capital or small letters. Excel automatically converts it into capitals.
- Every formula has a defined syntax/ structure/ components, that should be followed
- Parts of syntax are separated by commas
- Formula always ends with bracket
- No. of opening brackets and closing brackets should be equal
- Numbers in formulae can be entered directly
- Text in formulae is entered within double quotation marks – "OK"
- Cell reference is entered in Alphanumeric code – A3, B5
- Cell range is entered with a colon between two cell references – A3:B5
- There should not be any blank spaces in formulae

Count Functions

- 5 types of Count Functions
  - Count: Counts only numbers
    - Formula: =Count(A2:A30)
  - CountA: Counts all non-blank cells – numbers & text
    - Formula: =CountA(A2:A30)
  - Countblank: Counts blank cells
    - Formula: =Countblank(A2:A30)
  - Countif: Counts the cells that satisfy a condition
    - Formula: =Countif(Range,Criteria). Eg: =Countif(A2:A30,">5")
  - Countifs: Counts the cases that satisfy more than one condition in the same or different cells
    - Formula: =Countifs(Range1,Criteria1,Range2,Criteria2…..).
    - Eg: =Countifs(A2:A30,">5", A2:A30,"<10")

Sum Functions

- 3 types of Sum Functions
  - Count: Gives total of the numbers in the range
    - Formula: =Sum(A2:A30)
  - Sumif: Adds the cells that satisfy a condition
    - Formula: =Sumif(Range,Criteria,Sum_Range).
    - Eg: =Sumif(A2:A30,">5",B2:B30)
  - Sumifs: Adds the cases that satisfy more than one condition in the same or different cells
    - Formula: =Sumifs(Sum_range,Criteria_Range1,Criteria1,Criteria_Range2,Criteria2…..)
    - Eg: =Sumifs(A2:A30,A2:A30,">5", A2:A30,"<10")

Min, Max, Quartile & Percentile Functions

- Min: Gives the minimum number in the range
  - Formula: =Min(A2:A30)
- Max: Gives the maximum number in the range
  - Formula: = Max(A2:A30)
- Quartile: Gives the number at the defined quartile in the range
  - Formula: =Quartile(Range,Quart); Eg: =Quartile(A2:A30,2)
  - Quart values: 0-Min, 1-1st quartile, 2 – median, 3 – 3rd quartile, 4 – Max)
- Percentile: Gives the number at the defined percentile in the range. Percentile should be given in decimal value (E.g. 50th percentile – 0.5, 90th percentile – 0.9)
  - Formula: =Percentile(Range,k); Eg: =Percentile(A2:A30,0.5)

Mean/ Average & Median

- Average: Gives the average/ mean of the in the range
  - Formula: =Mean(A2:A30)
- Median: Gives the middle number/ 50th percentile in the range
  - Formula: = Median(A2:A30)

Text Functions

- Join Function: Physically joins the characters and creates a text string
  - Formula: =A2&" is OK"
- Left: Extracts the defined number of characters from the left of a text string
  - Formula: =Left(Cell Reference/ Text,No. of characters)
  - E.g. =Left(A2,2); =Left("Sri Lanka",3)
- Right: Extracts the defined number of characters from the right of a text string
  - Formula: =Right(Cell Reference/ Text,No. of characters)
  - E.g. = Right(A2,2); =Right("Sri Lanka",3)
- Mid: Extracts the defined number of characters from the Middle of a text string
  - Formula: =Mid(Cell Reference/ Text,Start_Num,No. of characters)
  - E.g. = Mid(A2,2,3); =Left("Sri Lanka",4,5)

IF Function

- One of the most powerful & most productive function in MS Excel
- If Function: Returns a predefined result if a condition is fulfilled and another predefined result if the condition is not fulfilled
- Multiple If conditions, upto 32, can be nested inside one another, to evaluate a series of conditions and give results
- If function has three parts – Condition, Value if True, Value if False.
- Condition can be specified using any other function in Excel
- Value if True and False can be a number, text (enclosed in double quotation marks), blank (double quotation marks without anything in between), a cell reference, a formula or another IF function.
- Used to convert one data into the other (Text to Numbers, Numbers to Categories, coding, etc.)
- Formula: =IF(Condition,Value if True,Value if False)
- E.g. =IF(A3>0,"OK","Error")

Examples of IF Function

- =IF(A3>0,"OK","ERROR")
- =IF(A3=0,"",B3/A3)
- =IF(OR(A3="NORTH",A3="SOUTH"),"NS","EW")
- =IF(A3<15,"<15",IF(AND(A3>=15,A3<=24),"15-24","25 & ABOVE")))

VLOOKUP Function

- Used to bring data of same cases from a different dataset or different part of the same dataset.
- Useful in combining two different datasets
- Need at least one common field – Lookup Value
- Lookup value should be in the first column of the lookup range
- Column number is relative to the first column of the lookup range; Not the column number in the worksheet
- Match types: True – Approximate Match; False – Exact Match
- =VLOOKUP(LOOKUP VALUE,LOOKUP RANGE,COLUMN NO.,MATCH TYPE)
- =VLOOKUP(A3,SHEET1!A2:P100,4,FALSE)

10. HIV/AIDS Specific Analysis from Programme Data

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VHS VOLUNTARY HEALTH SERVICES WE SERVE THE PEOPLE WHO SERVE YOU

Outline

- Epidemic Analysis
- Progress Analysis
- Performance Analysis
- Cascade & Cohort Analysis

Epidemic Analysis

- HIV & STD Positivity rates
  - Levels
  - Trends
  - Differentials by demographic & risk characteristics
- Size of beneficiaries
- Profile of beneficiaries
  - STD Clinic attendees
  - PLHIV
- Vulnerabilities/ Risk Behaviours among KP
  - Typology, Partner volume, condom use, STI uptake

Progress Analysis

- Progress against targets
  - Levels
  - Trends
  - Differentials by province/ district
  - Differentials by other characteristics
- Gaps in Achievement
- Different Denominators

Performance Analysis

- Performance Indicators
  - Levels
  - Trends
  - Differentials by province/ district
  - Differentials by other characteristics
- Best & worst performing units
- Performance Quartiles
- Performance Quadrants

Cascade & Cohort Analysis

- Testing and treatment cascade
- Prevention cascade
- PLHIV cohort analysis
  - Incidence of Opportunistic infections
  - Mortality rates
  - Survival analysis
- Positive Pregnant Women & Exposed Baby cohort analysis

11. Data Triangulation

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Data Triangulation

- Concept & Principles
- What it does?
- When to use?
- Advantages
- Questions for Data Triangulation
- Data Sources
- Data Quality Assessments
- Methods of analysis
- Hypothesis-building/ Epidemiological Framework

Data triangulation

Data Triangulation is an Analytical Approach that synthesizes data from multiple sources, to improve the understanding of a public health issue and guide programmatic decision-making to address the issue.

- By putting different bits of information from different sources into a meaningful framework, it explains and improves the understanding of HIV/AIDS scenario in the district.
- By providing answers to vital programme questions, it helps in taking effective decisions for planning and implementation of HIV prevention & control efforts

Basic Premise

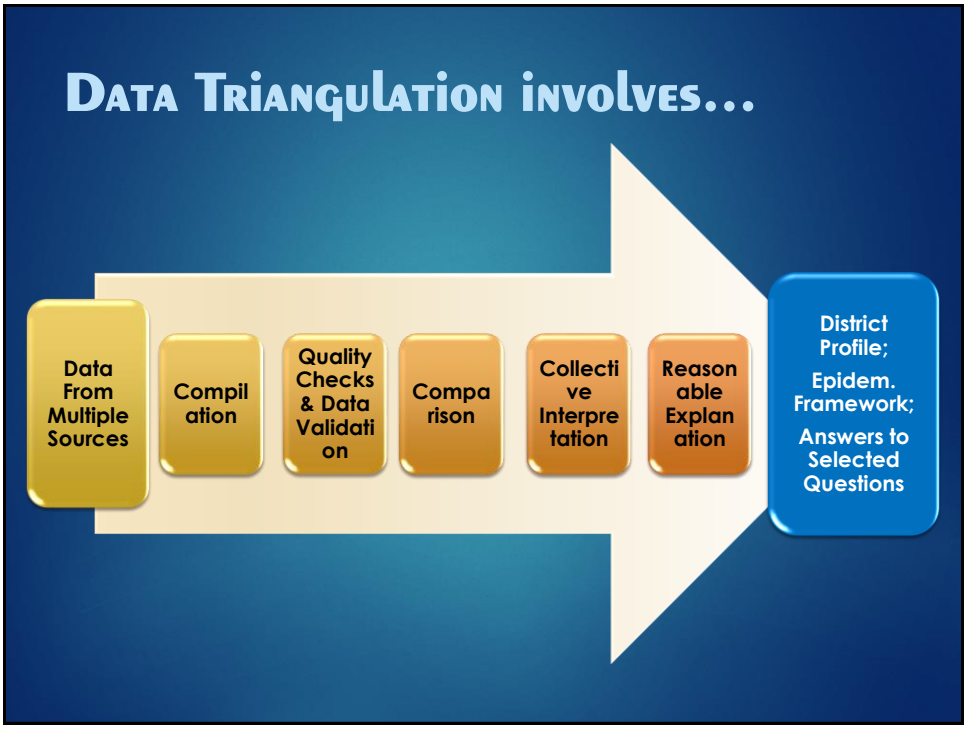
- Works on already available evidence; No primary data collection involved
- Exploratory in nature (arrive at plausible hypothesis); May not always give definitive conclusions with statistical rigour
- Better suited for broader issues at population/ community/ region level; May not work well for micro-issues
- Always guided by a specific question to be answered
- Need a basic understanding of epidemiological considerations
- Element of judgment is critical

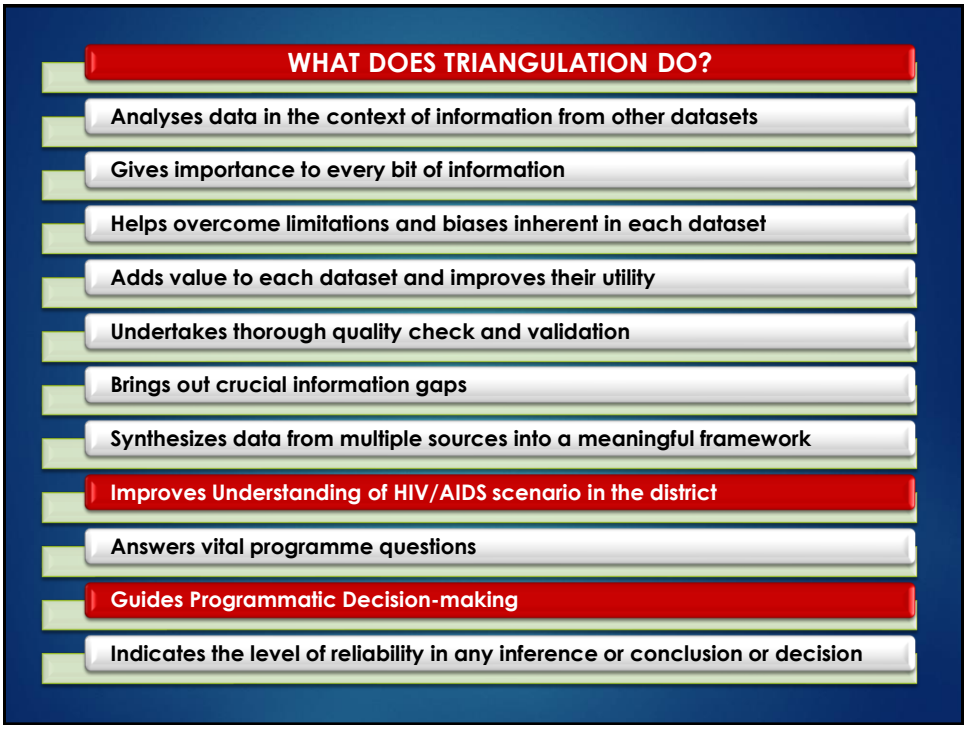
Integrates diverse information

- Info. on diverse subjects/ thematic areas
- Data of different types (quantitative, qualitative, anecdotal, referential, experiential, etc.)
- Data generated through different methodologies (survey, evaluation, rapid assessment, OR etc.)
- Data generated for different purposes (Surveillance, monitoring, academic research etc.)
- Data generated by different sources (health, non-health)
- Data applicable to different populations/ regions
- Data with different time reference

Basic principle of data triangulation

To analyse and interpret a dataset in the light of information emerging from other datasets so that the synthesis offers a better understanding of the issues than what will be inferred from a single dataset


DATA TRIANGULATION INVOLVES...
Data From Multiple Sources
Compil ation
Quality Checks & Data Validati on
Compa rison
Collecti ve Interpre tation
Reason able Explan ation
District Profile;
Epidem. Framework;
Answers to Selected Questions


WHAT DOES TRIANGULATION DO?
Analyses data in the context of information from other datasets
Gives importance to every bit of information
Helps overcome limitations and biases inherent in each dataset
Adds value to each dataset and improves their utility
Undertakes thorough quality check and validation
Brings out crucial information gaps
Synthesizes data from multiple sources into a meaningful framework
Improves Understanding of HIV/AIDS scenario in the district
Answers vital programme questions
Guides Programmatic Decision-making
Indicates the level of reliability in any inference or conclusion or decision

When to use Triangulation Approaches?

- When data is plentiful but dissimilar
- When data is scanty
- When the quality of data is not optimal
- When quick answers are needed
- When there are no resources for rigorous research
- When broad policy / strategy level decisions are to be made
- Before planning any major data collection activities

Research and triangulation

Research analysis	Triangulation analysis
• Focus on statistical analysis • Designed to provide data that can be generalized • Variables from a single dataset • Focus on internal validity: *"Did A cause B to change among group C?"* • Emphasis on generating the highest scientific rigor of data for interpretation • Long delay between data collection and presentation of results	• May or may not use statistics.Use of statistical analysis will depend on available data • Variables from multiple datasets • Focus on external validity: *"Can observed effects in group C be attributed to the larger population as well?"* • Emphasis on the "best possible" interpretation of existing data for policy and programme decision-making • Quick turnaround between secondary data capture and presentation of results

Source: HIV Triangulation Resource Guide, CDC/UNAIDS

Meta-analysis and triangulation

- Meta-analysis: combines rigorous scientific data of similar quality and design to conduct statistical analyses.
- Triangulation: seeks to make use of data from diverse sources and study designs, and incorporates judgments, findings and interpretations on each data source's limitations.

Advantages of Triangulation

- Can be easily employed in programme settings
- Can be applied to virtually every programme question, but more appropriate while seeking answers to broad questions on programme and policy
- Helps in identifying more data sources
- Helps in bringing out information gaps
- Provides new insights and helps in generating new hypotheses
- Less resource intensive (relatively)
- Makes best possible use of available evidence; a good alternative to new data generation

Questions for triangulation

- Importance: The question should address a current and pertinent issue. The answer to that question should significantly influence your understanding of and response to the HIV/AIDS situation.
- Actionability: The results of the data-triangulation process should be useful to make improvements in HIV/AIDS programme.
- Answerability (Data availability & Feasibility): The data required for answering the question should be available in usable format. At least, it should be possible to extract it from existing reports or documents. Primary data collection of any form is not a part of data triangulation.
- Appropriateness: Finally, it is important to ascertain whether triangulation is the appropriate method to use to answer the question. Could the question be better answered by any other research methods of analysis, an expert panel, or another type of study?

HELPS ANSWERING CRITICAL PROGRAM QUESTIONS…

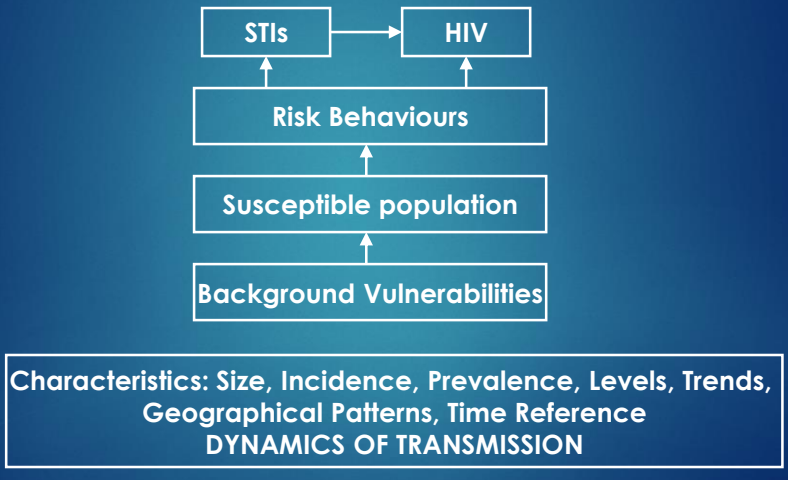
- Is the HIV epidemic increasing, decreasing or stable in the state/ district?
- In a state, which districts should be prioritised for focused attention?
- In a district, what are the specific areas that programme should focus its resources and efforts? *What are the drivers of HIV epidemic?*
- What are the important programme gaps in the district?
- Are the current prevention efforts positioned for maximum impact?
- Is the current mix of interventions appropriate for averting new infections? Or will a different mix result in more averted infections?

Start with a broad set of questions. Revisit and Refine the questions at every step

Steps in Data Triangulation

- Step 1: Understanding Thematic Areas & Questions for Triangulation
- Step 2: Review of Data Sources and Assessment of Data Availability in the District
- Step 3: Decision on Questions to be answered for the district
- Step 4: Compilation of Secondary Data
- Step 5: Data Quality Assessments
- Step 6: Data Validation, Adjustments & Filling Data Gaps
- Step 7: Preparation of Data Tables with clean data for analysis
- Step 8: Data Analysis, Interpretation and Inferences; Describe Thematic Areas
- Step 9: Data Triangulation (Hypotheses Building; Answer Triangulation Questions)
- Step 10: Preparation of Draft Reports
- Step 11: Discussion & Consultation with stakeholders & Local experts on draft reports
- Step 12: Finalisation of conclusions and recommendations
- REVISIT QUESTIONS AT EVERY STEP

Epidemiological Considerations



Data Quality Assessment

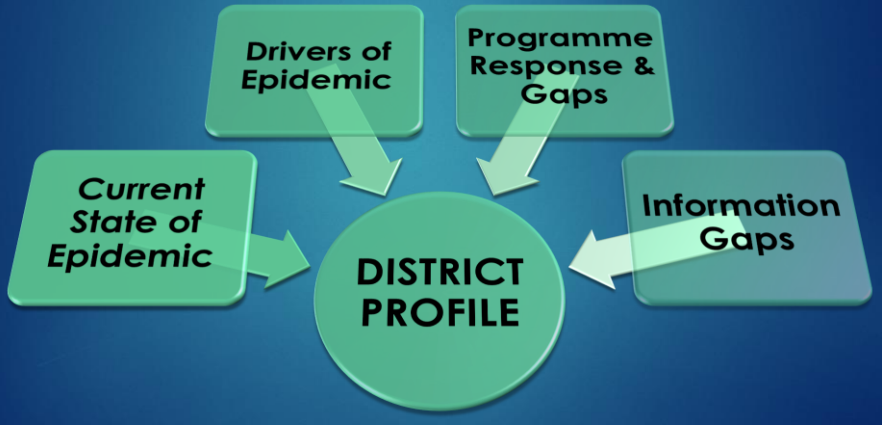
1. Review of methodology, design, data collection processes, possible biases
2. Validity and Reliability
3. Completeness
4. Correctness
5. Consistency

Followed by cleaning, adjustments and validation

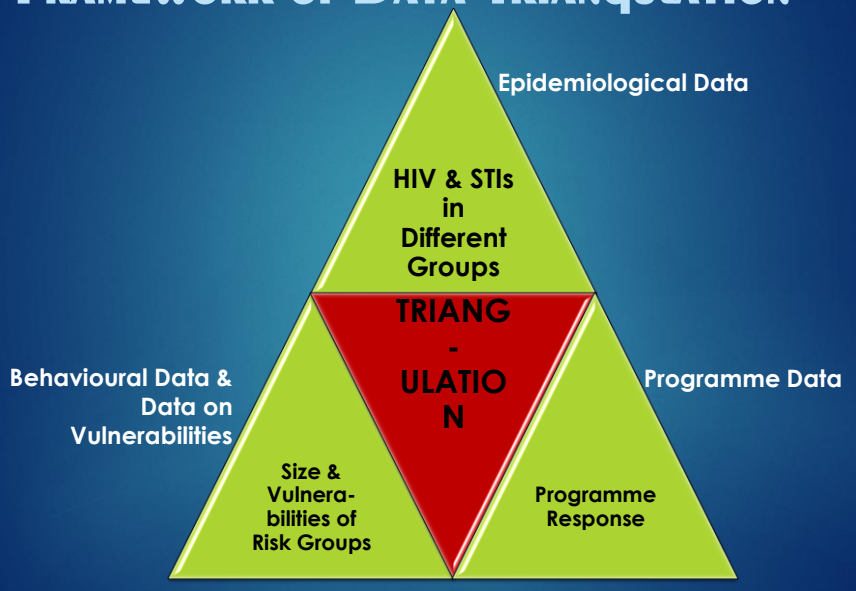
Framework for Data Triangulation

Components of Data Triangulation	What it Does?	Guiding Elements	Action To Do	Output
Descriptive Analysis	Describes (What? Who? When? Where?)	Themes	Analyse Data & Describe the Themes	Descriptive Section of Report
Triangulation	Explains (How? Why?)	Questions	Triangulate Data & Answer the Questions	Synthesis Section of Report

Framework for Descriptive analysis



Framework of Data Triangulation



Triangulation of info on same data element from different sources

- HIV Prevalence among GP: Compare HSS ANC data with PPTCT data.
- HIV Prevalence among HRG: Compare HSS HRG data with ICTC General Clients data and TI NGO data (wherever available)
- STI Prevalence and Trends: Compare data on STIs from STI Clinics with that collected in behavioural surveys, STI treatment data from TI projects and VDRL positivity rates from ANC HSS.
- Size of HRG: Compare mapping data of HRG with TI Needs Assessment and coverage data of HRG, both refer to the size of HRG.
- Risk Behaviours: Compare behavioural survey data from BSS/ IBBA with the behavioural data from TI Projects, wherever available.
- Risk Behaviours: Compare routes of transmission reported from ICTC with that from ART registration data
- Size of PLHA: PLHA density from ART registration data with that from Positive People Networks & ICTC detections

Triangulation of information on different data elements

- Triangulation in time plane
- Triangulation in geographical plane
  - It establishes the relationships between different data elements.
  - It helps in identifying the blocks/ talukas where the HIV epidemic or its vulnerabilities are high and hence require priority attention under the programme.
  - Thirdly, it shows the gaps in the programme response in terms of geographic coverage i.e. talukas/ blocks where there is need but there are no interventions.

Triangulating time data

- Timing of HIV trends among different groups
- HIV trends among different groups and see if the information available on HRG size at different points in time has any relation with HIV trends.
- HIV trends among different groups and see if the trends in STIs correlate with HIV trends.
- HIV trends among different groups and see if the data on risk behaviours and changes in risk behaviours correlate with HIV trends.
- HIV trends among different groups and mark the timing of start of interventions in the district (preventive interventions, counseling & testing and ART) and see if any message emerges out.
- If any information is available on any other data element such as migration, establishing industries or factories, starting of Trans-shipment locations or truck halt points, etc., they should be captured in the time trend analysis and examined for any meaningful insights.

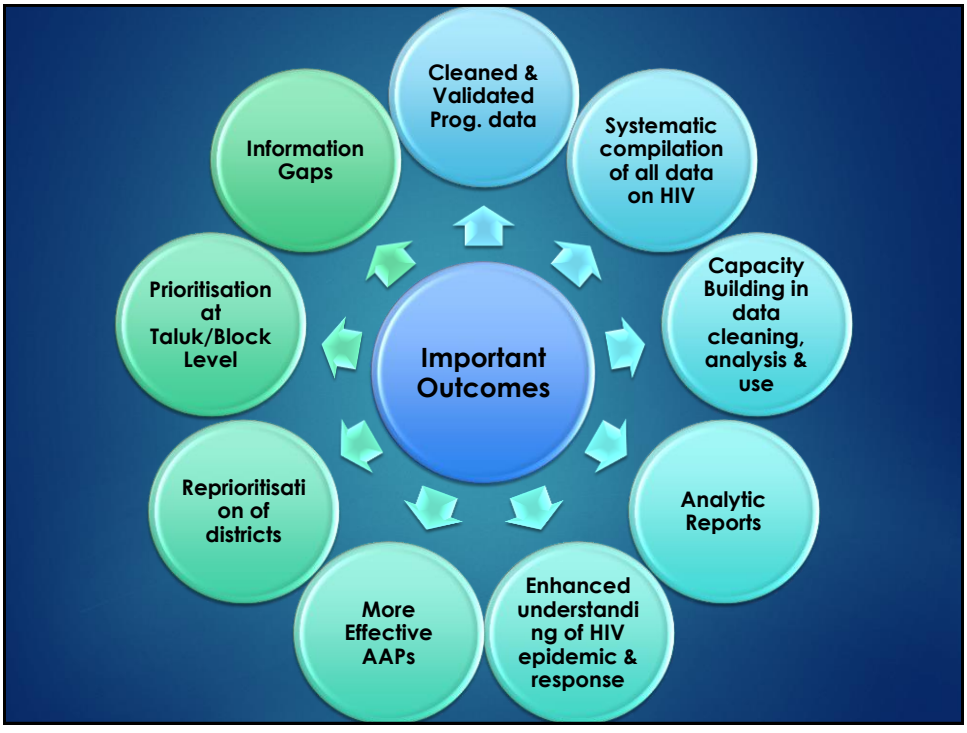
Triangulating geographic patterns

- HIV Positivity among pregnant women from PPTCT data
- HIV Positivity among general clients from ICTC data.
- High Risk Group mapping data
- ART Registration data
- Presence of programme interventions (TIs, ICTCs, PPTCT centres, ART centres, LACs, CCCs, PLHA networks)
- Sources of out-migration & Destinations of in-migration

Building an Epidemiological Framework…

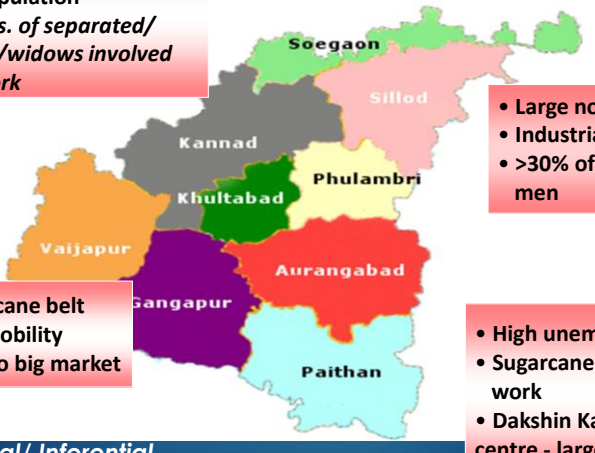
- Starting with contextual factors and background vulnerabilities
- Understanding the presence of risk groups, their distribution and profiles
- Understanding the risk behaviour patterns among HRG and GP
- Understanding the important vulnerabilities such as STIs and Migration
- Describing how all these interact and cause HIV epidemic
- Understanding the HIV/AIDS scenario and patterns in different risk groups and general population
- Appreciation of possible future vulnerabilities and direction of epidemic

Making a story out of it…



District Vulnerabilitites explain the epidemic scenario in Aurangabad Dt., Maharashtra

- Large out migration of single men
- Tribal population
- *Large nos. of separated/ deserted/widows involved in sex work*

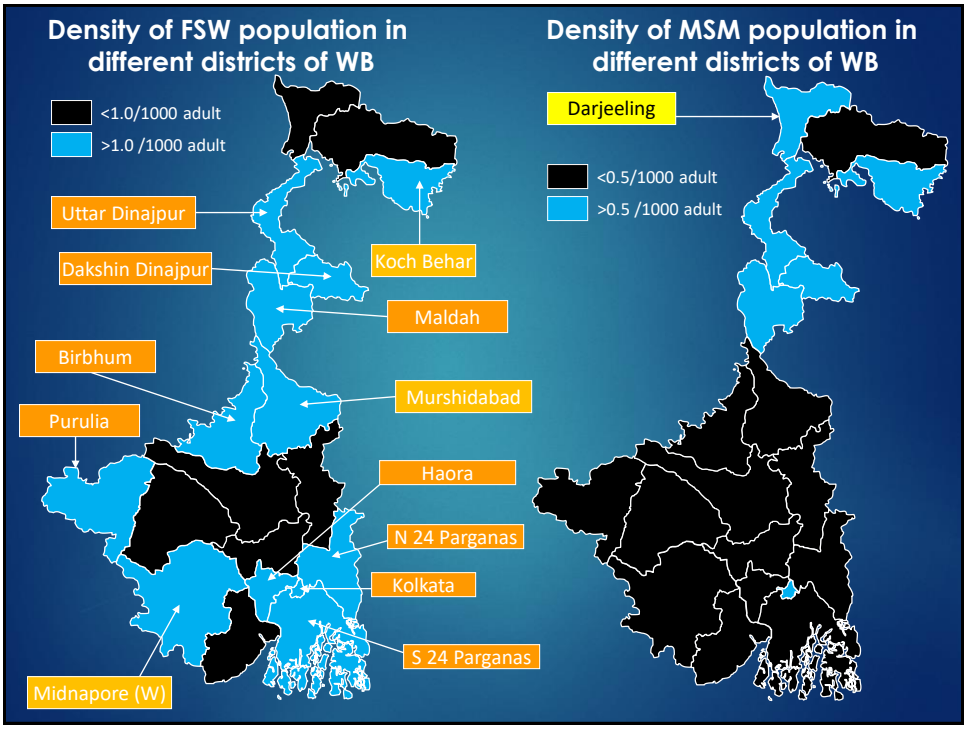


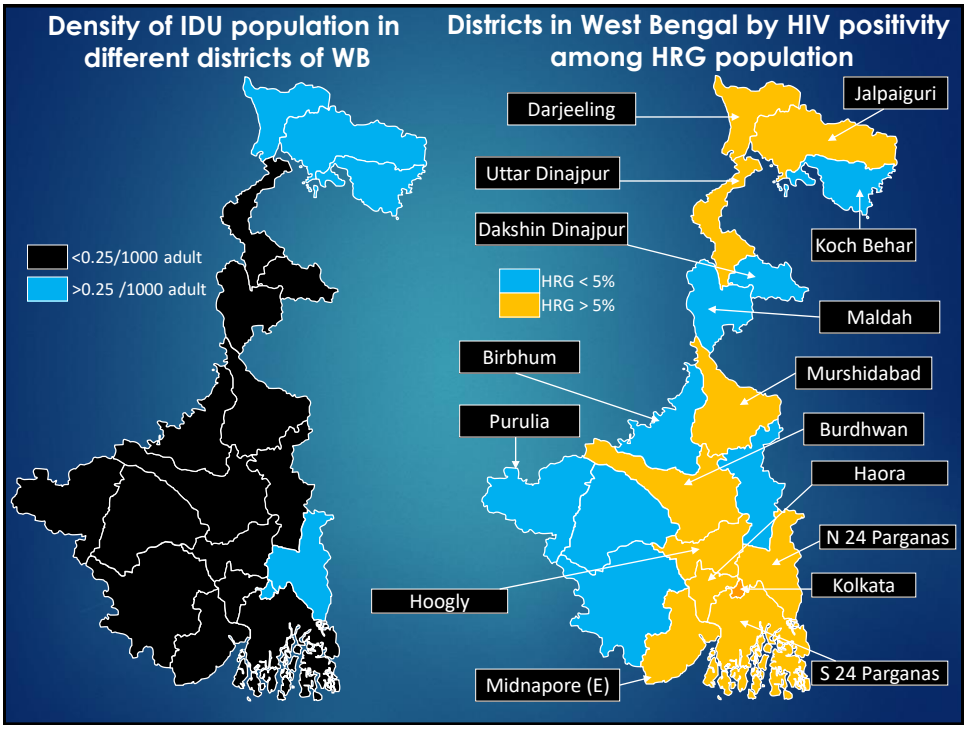
- Large no. of in-migrants
- Industrialization & sex work
- >30% of in-migrants - single men

- Rich sugar cane belt
- Seasonal mobility
- Proximity to big market

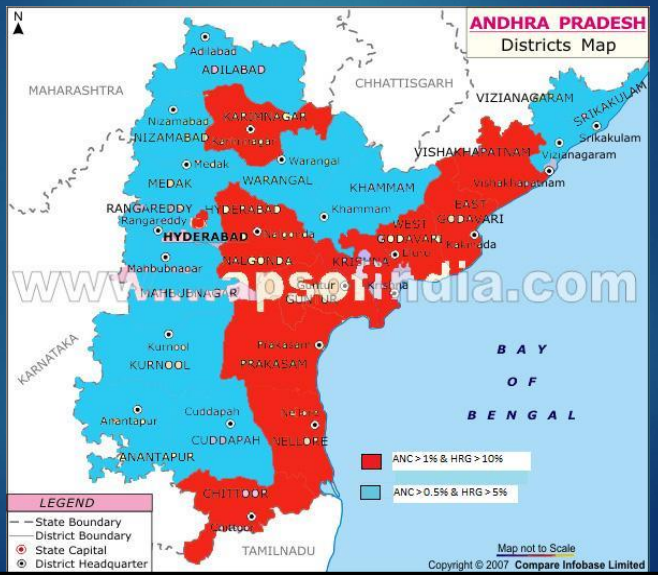
- High unemployment
- Sugarcane factories – seasonal work
- Dakshin Kashi – Pilgrimage centre - large no. of visitors

Anecdotal/ Inferential…
Evidence-based…

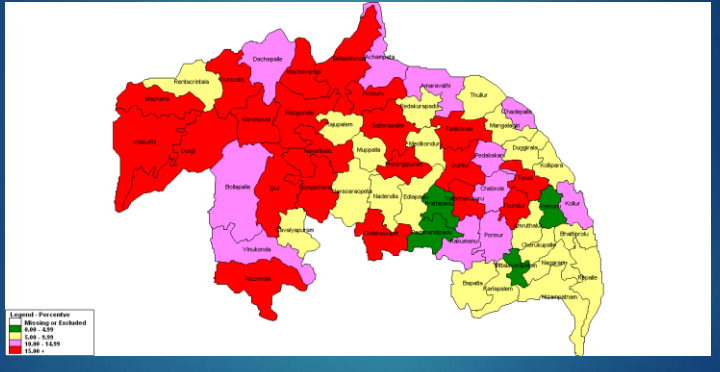

Density of FSW population in different districts of WB
<1.0/1000 adult
>1.0 /1000 adult
Uttar Dinajpur
Dakshin Dinajpur
Koch Behar
Maldah
Birbhum
Murshidabad
Purulia
Haora
N 24 Parganas
Kolkata
S 24 Parganas
Midnapore (W)
Density of MSM population in different districts of WB
Darjeeling
<0.5/1000 adult
>0.5 /1000 adult

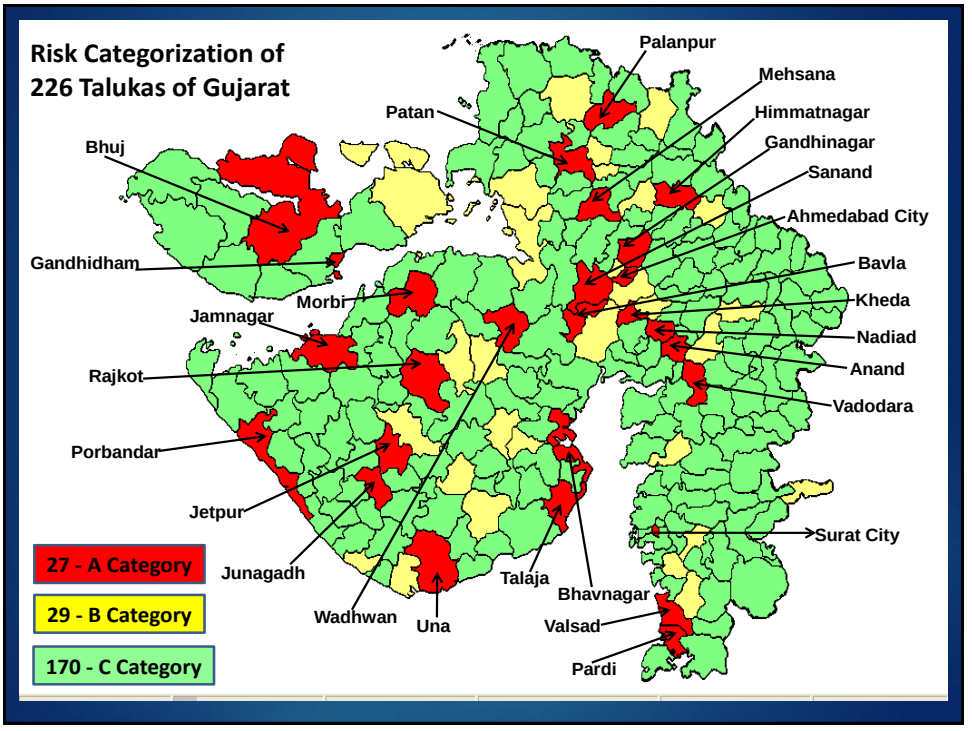

Density of IDU population in different districts of WB
<0.25/1000 adult
>0.25 /1000 adult
Districts in West Bengal by HIV positivity among HRG population
Darjeeling
Jalpaiguri
Uttar Dinajpur
Dakshin Dinajpur
Koch Behar
HRG < 5%
HRG > 5%
Maldah
Birbhum
Murshidabad
Purulia
Burdhwan
Haora
N 24 Parganas
Kolkata
Hoogly
Midnapore (E)
S 24 Parganas

Example of District Prioritisation in AP using Multiple datasets



Mandal Wise VCTC Positivity, Guntur Dt., AP





Data Triangulation brings out the Programme gaps & Information Gaps…

VCTC - 2008	Men	Women	Total	PPTCT 2008	Size/TI coverage FSW	Size/TI coverage MSM	PLHA registered
Hanumakonda	Sample very small			1.70	270		657
Warangal	24.8	29.6	26.7	0.68	117	1655	1234
Wardhannapet	3.7	2.5	3.3	0.66	262	268	101
Janagaon	9.0	9.0	9.0	0.59	752	542	164
Mahabubabad	4.4	6.7	5.3	0.52	850	1147	155
Narsampet	3.6	5.6	4.2	0.41	260	211	
Parkal	6.6	4.5	5.6	0.39	522	906	135
Mullugu	2.5	6.0	3.5	0.36			
Cheryal	4.3	4.1	4.2	0.18			
Station-Ghanpur	5.6	3.0	4.4	0.00			109
Eturunagaram	1.4	3.2	1.9	0.00			
Gudur	2.2	1.7	1.9	0.00			
Chityal	1.1	1.3	1.2	0.00			
Warangal	6.6	8.0	7.0				
Bhopalpally					745	453	
Thorur					271	273	111
Keesamudram					276	256	
Nekonda					254	266	
Hasanbharathy					332		180
Dharmasagar					51		147
Atmakur							121

Needs to initiate prevention program

No information on HIV prevalence

12. Communication of Data Analysis Results

Ministry of Health Sri Lanka | National STD/AIDS Control Programme | Voluntary Health Services

Outline

- Scientific Abstracts
- Scientific Papers/ Articles
- Policy Briefs
- Handouts
- Posters
- Reports
- Presentation

Scientific Abstracts

- Short and Summarised; 250-300 words
- For conference and journal publications
- Features
  - Objectives
  - Methods
  - Results
  - Conclusion
  - Key words

Scientific Paper/ Article

- Published in Peer Reviewed Journals
- Detailed version of the data sources, methods and results
- Sections
  - Introduction
  - Methods
  - Results
  - Discussion – The most important section; Author's contribution
  - References
  - Acknowledgements
  - Conflict of Interest statement
  - Tables & Figures

Policy Briefs

- Targeted at Policy makers, senior level programme managers & experts
- Key take home messages to be very crisply worded and adequately highlighted; Should be at the top or on the cover page
- Extract of only the top-line, high priority findings
- Presented in a visually appealing manner for quick consumption; Graphs, Infographics, Pictures; Not in tables; Very limited text
- Very limited info on methods, if required
- Not more than 2 pages
- Sometimes done as wall chart
- Colourful and elegantly designed

Handouts/ Folders

- Larger than policy briefs; Smaller than Papers; Much smaller than reports
- 4-8 pages; designed as a folder/ folder chart/ booklet
- Targeted at key stakeholders – prog managers, communities, field staff, academicians, experts
- Easy to mail and circulate
- Presents the topline findings and other key findings
- Graphs & Tables
- Brief paragraphs or text on the intro/background/ methods may be included
- Conclusions/ Key take home messages highlighted
- Contact details

Posters

- Large sized visual presentation of the methods and results
- Mostly use graphs, maps, infographics; Limited text
- Targeted towards academic dissemination & conferences
- May also be used as standees during official meetings/ programme events
- Sections:
  - Introduction
  - Methods
  - Results – Major section
  - Conclusions

Presentations

- Using MS Power-point or Prezi
- Based on target audience
  - Detailed or Brief
  - Text or Graphics or both
  - Academic or Policy
  - High impact presentations vs Discussion-oriented
  - Interesting or boring
- Clean, less crowded slides (not more than 8 lines)
- Can embed videos and animation

Reports

- More detailed and elaborate
- Complete details of the project, process, outcomes and discussion
- Sections and chapters
- Detailed presentation of results in text, tables & graphics forms
- Foreword/ Preface, List of abbreviations, List of tables, List of figures
- Detailed references and acknowledgements
- Targeted at universal audience
- Final and complete consolidation of all the process and the results

13. Data Use for Decision Making - A Systems Approach

Ministry of Health Sri Lanka | National STD/AIDS Control Programme | Voluntary Health Services

Successful control of HIV epidemic needs..

- Effectiveness in delivery of services at every level….targeting the right thing
- Understanding local scenario and factors responsible
- District specific action plans with identified focus areas
- Customised choice of programme interventions
- Contextualised emphasis on programme components
- EFFECTIVE & TIMELY USE OF DATA AT ALL LEVELS

DATA USE - An approach, a mindset

- Developing the approach of using data for decision-making and programme planning at all levels
- Inculcating, among programme managers, a habit of looking at data regularly
- Encouraging simple analytical methods that anyone can employ
- Emphasizing the importance of local knowledge and contextual understanding
- Capacity-building of state & district level institutes/personnel for sustainability

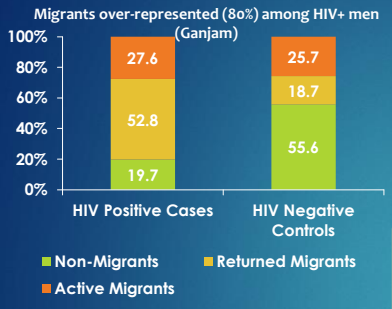
A long term process...

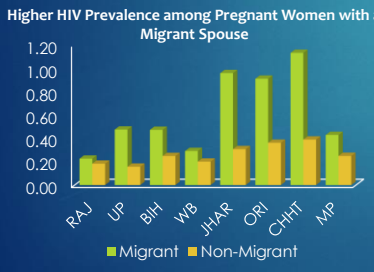
- Trigger the interest in data analysis and use
- Make programme managers work on data of their own state/district & reflect upon the insights
- Expose them to real time examples that demonstrate the use of data for decision making & programme planning
- Develop guidelines & tools to assist them in data use
- Develop HR plans that sustain the interest & facilitate data use as an on-going process
- SYSTEM STRENGTHENING FOR DATA USE

AIDS Response: An Evidence-led Approach

Some examples

Evidence-driven recasting of Migrant Strategy





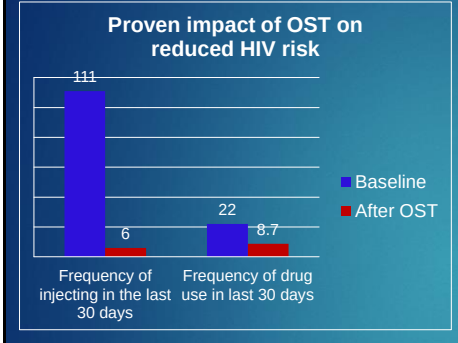
Linkages between source & destination

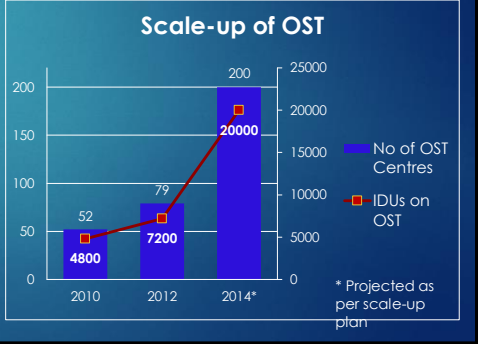
…resulting in re-scoping and reshaping of Migrant Strategy under NACP-III

- Adding transit and source interventions to migrant package
- Tracking migrant routes for better focus
- Reverse mapping to identify sources and address spouses of migrants

*Source: Population Council Study -- **Reference:** Saggurti N, Mahapatra BB, Swain SN, Jain AK. Male out-migration and sexual risk behavior in India: Is the place of origin critical for HIV prevention programs?. BMC Public Health. 2011. **11**:S6; NACO HIV Sentinel Surveillance*

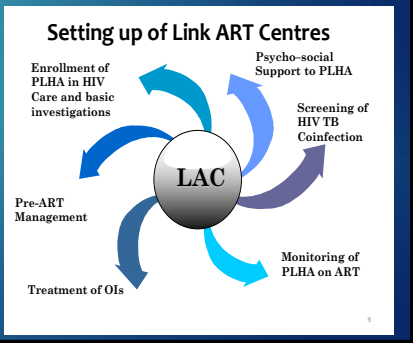
Roll out of OST Programme



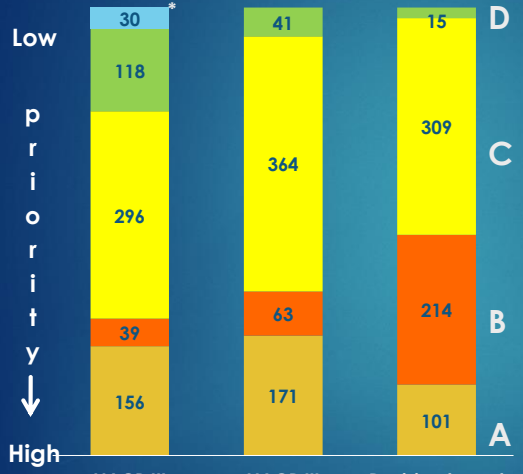


Setting up of Link ART Centres

OR to study Factors affecting ART Adherence



District Prioritisation through Data Triangulation



- Multiple data sources
- Rigorous quality checks and data cleaning

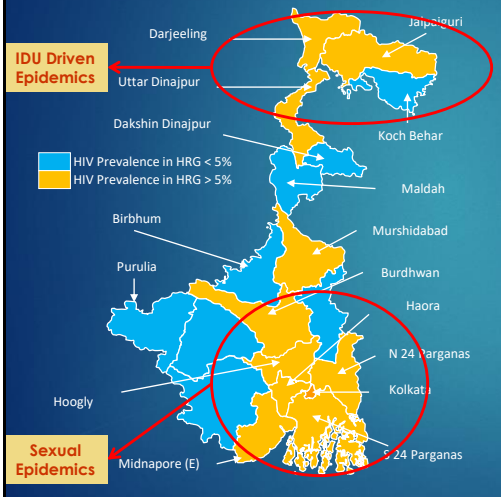
- Better Quality Data
- Weightage for Vulnerability

Source: Draft Framework for Recategorisation of Districts, NACO

* = new districts

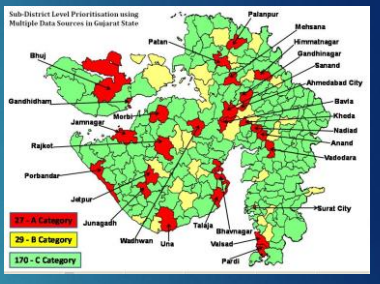
Refining Focus Based on Evidence

Focusing on HRG in West Bengal based on Epidemic Potential

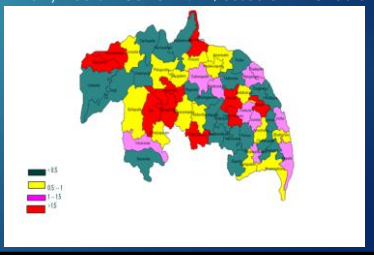


Sub-District Level Focus based on HIV burden and Vulnerabilities

Priority Blocks in Gujarat based on Data Triangulation



Priority Blocks in Guntur Dt. AP, based on PPTCT data

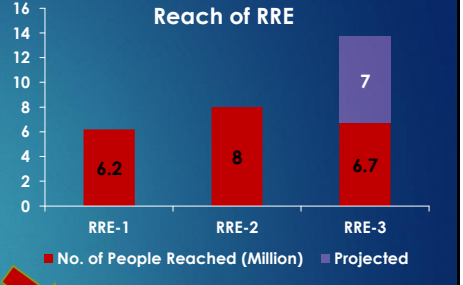


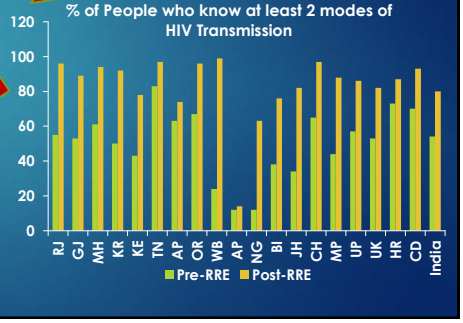
Lessons from Evaluation ... Red Ribbon Express

PHASE-1 (2007)
- Focus on outreach
- No services

PHASE-2 (2010)
- Addition of services (Counseling & testing)
- Convergence with the health system
- External Evaluation

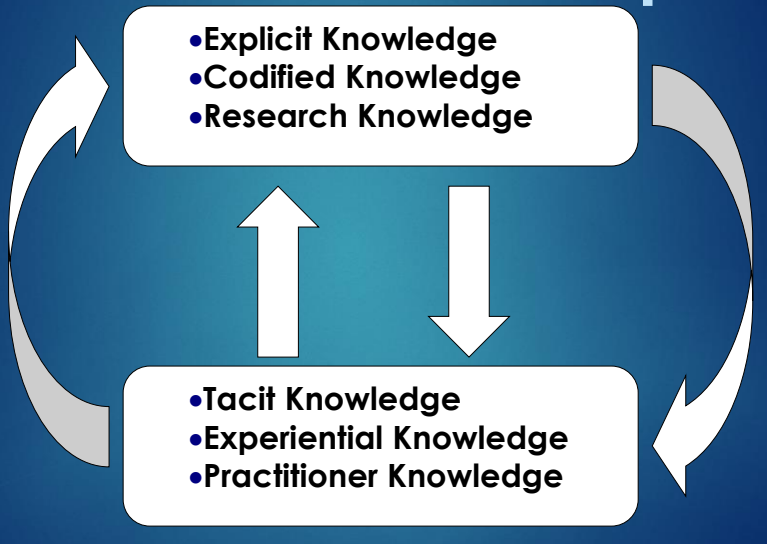
PHASE-3 (2012)
- Focus on Knowledge of HIV Services & Vulnerabilities – Youth & Migration
- Convergence and outreach strengthened; Greater focus on mid-media, folk arts, bus campaigns etc.
- Day-to-day monitoring, Pre-Post & Concurrent Evaluation





Knowledge Management

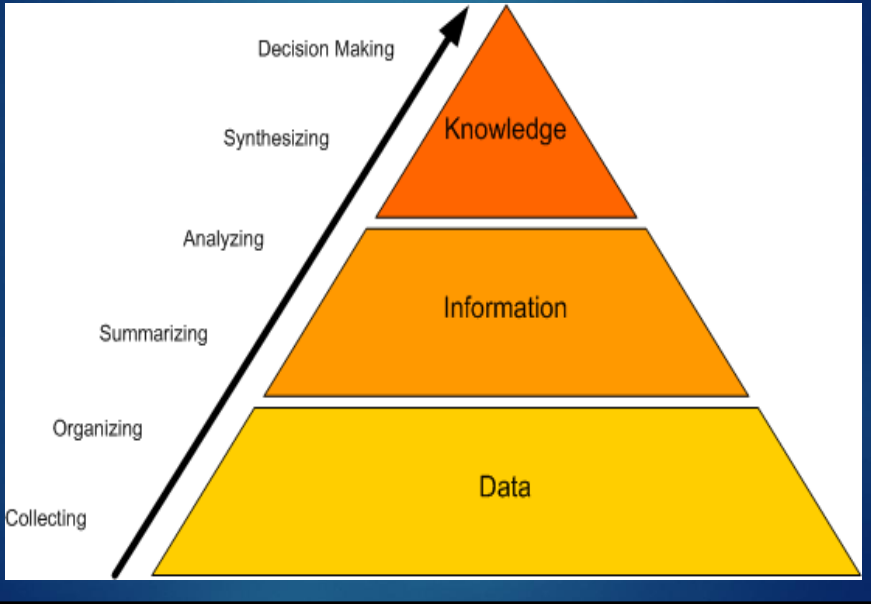
Two main forms of Knowledge



Four components of Knowledge Management

- Creation of Knowledge – Data Analysis (Converting data into information & knowledge)
- Collection and storage of Knowledge from other sources – Data Archiving
- Sharing of knowledge – Dissemination and Communication of Knowledge Products to stakeholders
- Translation of Knowledge – Data Use for programmatic action

Knowledge Creation



Knowledge Collection and Archiving

- While the former refers to creating knowledge from data generated under the programme, this component refers to
  - Collection and documentation of knowledge generated by other sources (communities, partners, academia, etc.)
  - Collection and synthesis of tacit or experiential knowledge from beneficiaries, service providers and other stakeholders
- Data Archives and Repositories
- Experience sharing mechanisms

Knowledge Sharing and Dissemination

- Passive dissemination through website, knowledge hub, data-centre, e-libraries, newsletters, reports and publications, etc.
- Active dissemination through seminars/ workshops/ conferences, discussion forums, e-learning tools etc.
- Addressing needs of wider audience - lay man, programme manages, policy makers, academia & researchers

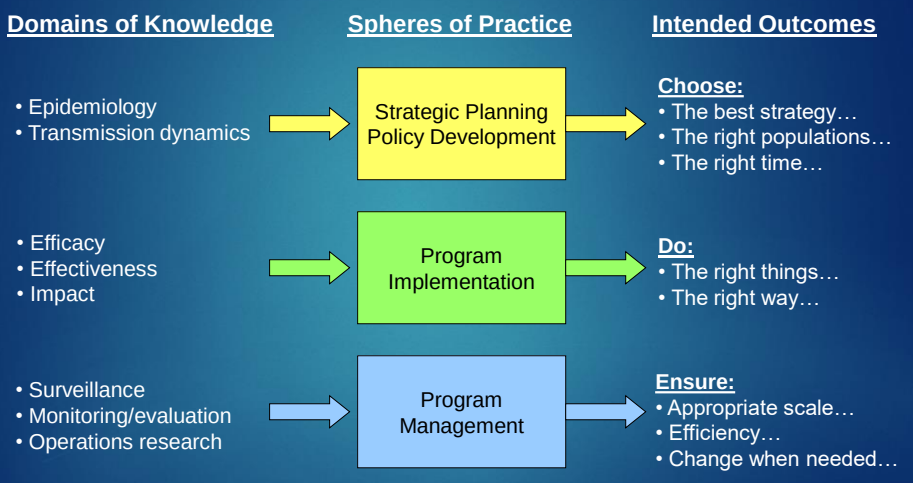
Knowledge Translation (Ensuring Effective Data Use)

- Effective use of knowledge for programmatic decision-making and policy or strategy formulation
- Starts with understanding programme requirements for evidence and analysis and identifying key questions to be answered
- Generating/compiling/synthesising body of knowledge to answer the questions and communicating them to the programme
- Taking programmatic decisions based on the evidence

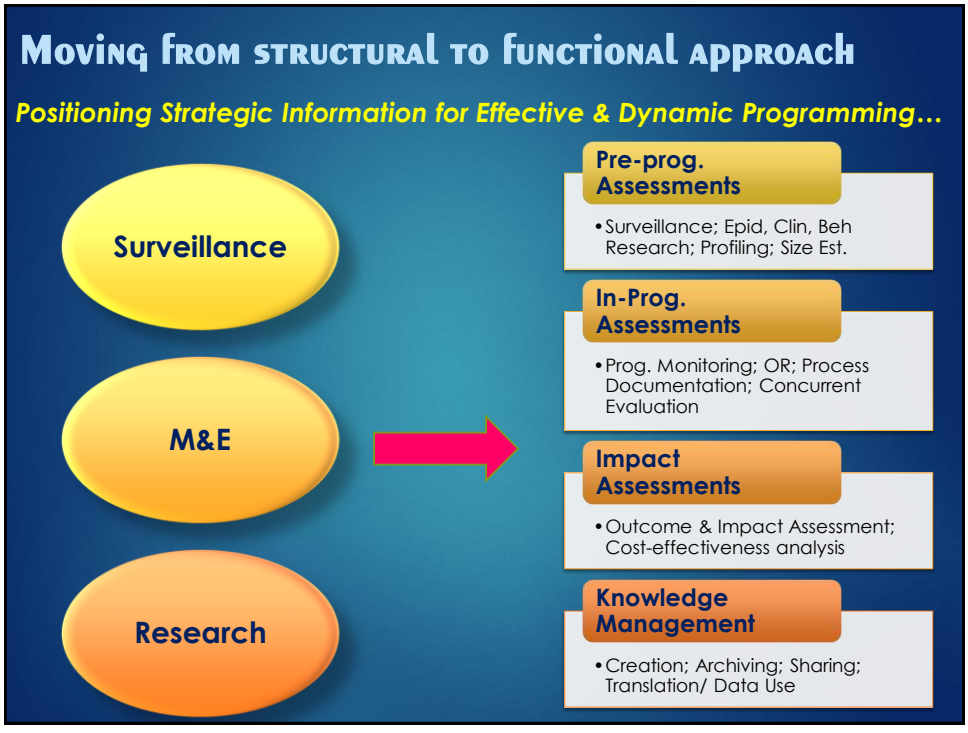
Need for System Strengthening for Effective Knowledge Management

- Not as a one-time activity, but as a systematic approach of
  - Shaping perceptions and promoting 'data use for decision-making' as a regular practice
  - Standardisation of methods & tools
  - Capacity building of prog. Managers & data teams
  - Devising mechanisms for collection of knowledge from other sources
  - Developing strategy for timely communication of analytical outputs
  - Building institutional resource pools
  - Promote scientific writing within the programme

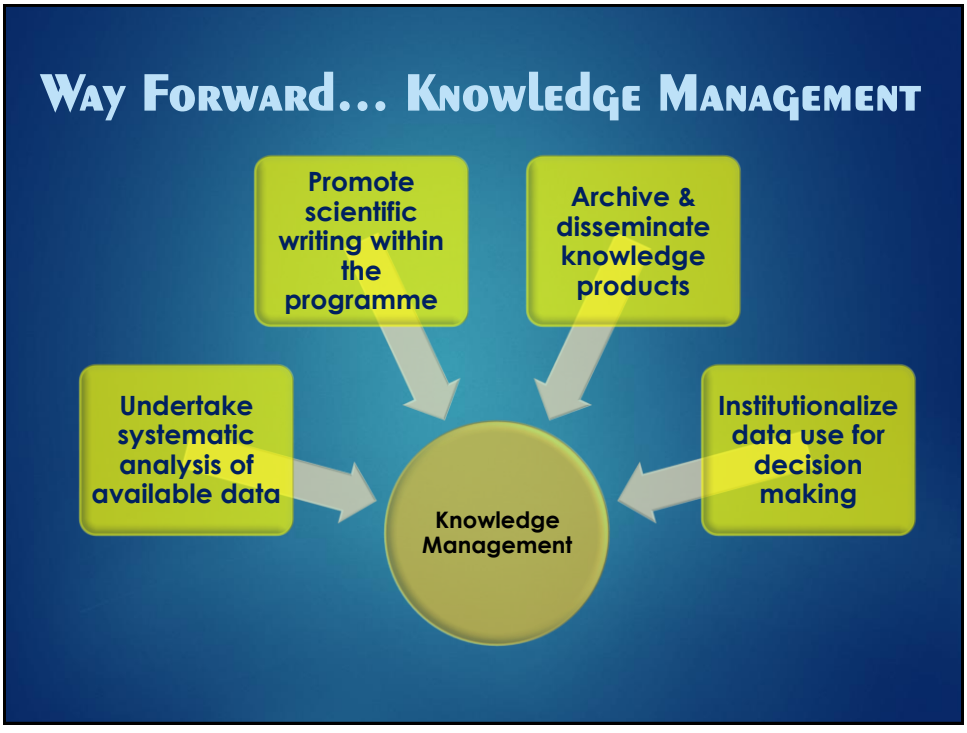
TRANSLATING KNOWLEDGE INTO PRACTICE – A SIMPLE MODEL FOR PUBLIC HEALTH



Way forward


Moving from structural to functional approach
Positioning Strategic Information for Effective & Dynamic Programming…
Surveillance
M&E
Research
Pre-prog. Assessments
• Surveillance; Epid, Clin, Beh Research; Profiling; Size Est.
In-Prog. Assessments
• Prog. Monitoring; OR; Process Documentation; Concurrent Evaluation
Impact Assessments
• Outcome & Impact Assessment; Cost-effectiveness analysis
Knowledge Management
• Creation; Archiving; Sharing; Translation/ Data Use


Way Forward… New Evidence
Strengthen Surveillance & MIS for real time monitoring
Commission studies to generate Country specific data
Integrate HIV into general health & social surveys
Initiate Process Documentation of Programme Approaches
Build evaluation into every programme initiative


Way Forward... Knowledge Management
Promote scientific writing within the programme
Archive & disseminate knowledge products
Undertake systematic analysis of available data
Institutionalize data use for decision making
Knowledge Management

14. How to Present the Data/Information

Ways to Present Data

- Text
- Tables
- Figures and illustrative graphs

Text

Text

Use text whenever there are small amounts of data to be summarized.

“During 2018, there is 60% increase in clinic attendance compare to last year”

Tables

Tables

Tables are arrangements of numbers or words in columns and rows that display data or relationships

STIs reported from STD clinics during 2018 by sex

Diagnosis	Male		Female		Total	
	No.	%	No.	%	No.	%
Genital herpes	1268	26%	1813	32%	3081	29%
Non-gono. infections	886	18%	1956	35%	2842	27%
Genital warts	1416	29%	1156	21%	2572	25%
Syphilis*	552	11%	263	5%	815	8%
Gonorrhoea	206	4%	76	1%	282	3%
Trichomoniasis	11	0%	42	1%	53	1%
Other STIs	528	11%	322	6%	850	8%
Total STIs	4867	100%	5628	100%	10495	100%

* All types of syphilis

Tables

- **Allow presentation of exact values**
- **Present large amounts of data in a small space**

STIs reported from STD clinics during 2018 by sex

Diagnosis	Male		Female		Total	
	No.	%	No.	%	No.	%
Genital herpes	1268	26%	1813	32%	3081	29%
Non-gono. infections	886	18%	1956	35%	2842	27%
Genital warts	1416	29%	1156	21%	2572	25%
Syphilis*	552	11%	263	5%	815	8%
Gonorrhoea	206	4%	76	1%	282	3%
Trichomoniasis	11	0%	42	1%	53	1%
Other STIs	528	11%	322	6%	850	8%
Total STIs	4867	100%	5628	100%	10495	100%

* All types of syphilis

Tables

- **Table number**
- **Title** describes the contents
- **Field** is the space filled by rows and columns
- **Column headings** describe column contents

STIs reported from STD clinics during 2018 by sex

Diagnosis	Male		Female		Total	
	No.	%	No.	%	No.	%
Genital herpes	1268	26%	1813	32%	3081	29%
Non-gono. infections	886	18%	1956	35%	2842	27%
Genital warts	1416	29%	1156	21%	2572	25%
Syphilis*	552	11%	263	5%	815	8%
Gonorrhoea	206	4%	76	1%	282	3%
Trichomoniasis	11	0%	42	1%	53	1%
Other STIs	528	11%	322	6%	850	8%
Total STIs	4867	100%	5628	100%	10495	100%

* All types of syphilis

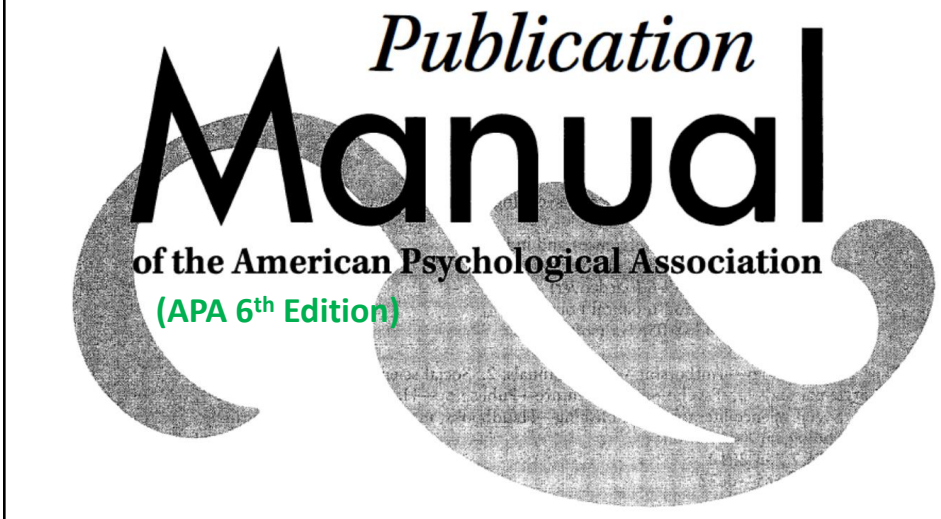
An example for a scientific paper

Table X

Proportion of Errors in Younger and Older Groups

	Younger			Older		
Level of difficulty	*n*	*M* (*SD*)	95% CI	*n*	*M* (*SD*)	95% CI
Low	12	.05 (.08)	[.02, .11]	18	.14 (.15)	[.08, .22]
Moderate	15	.05 (.07)	[.02, .10]	12	.17 (.15)	[.08, .28]
High	16	.11 (.10)	[.07, .17]	14	.26 (.21)	[.15, .39]

Note. CI = confidence interval.



Figures and graphs

Type of Charts/Graphs

- Bar Charts
- Pie Charts
- Line Charts

1. Bar Charts

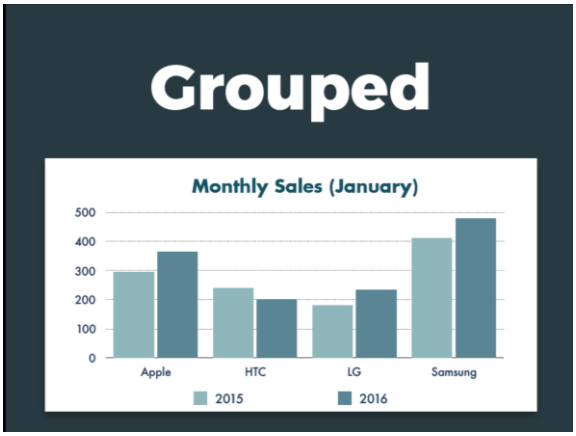

Bar Charts
display and compare discrete categories of data


Bar Charts
the height of each bar represents a numerical value

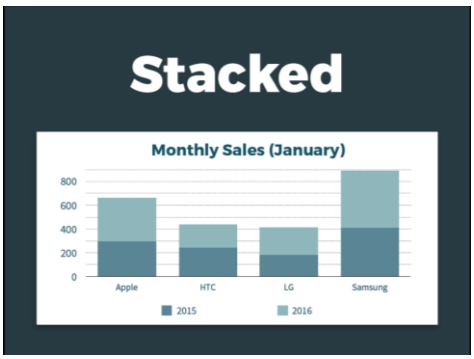
Bar Charts

Bar Charts

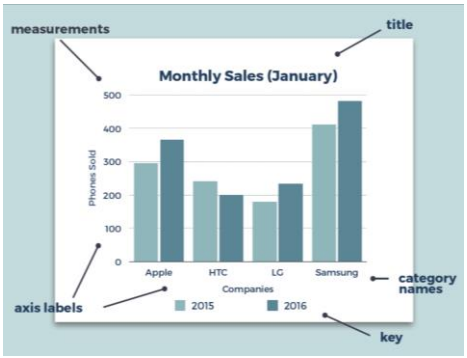
Bar Charts



Bar Charts



Parts of a Bar Chart

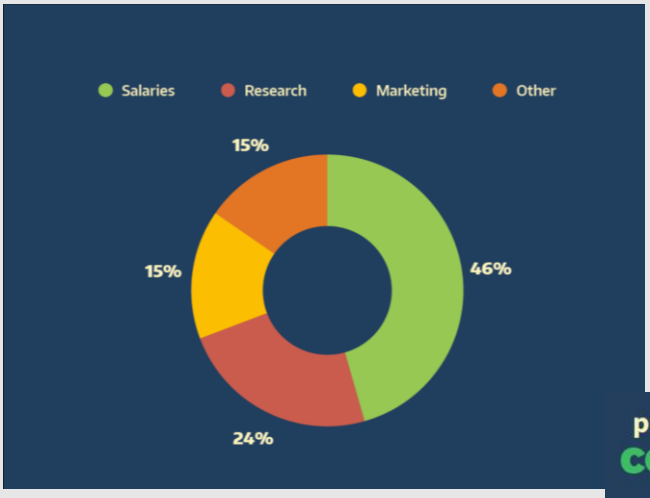


Pie Charts

Pie Charts

a **circular** graph that shows the **relative contribution** different categories make to an **overall total**

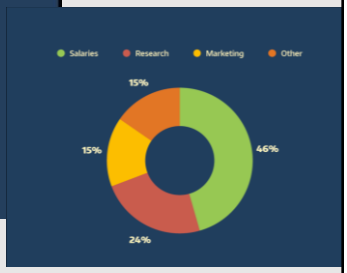
Pie Charts



pie charts only show **comparison** not change

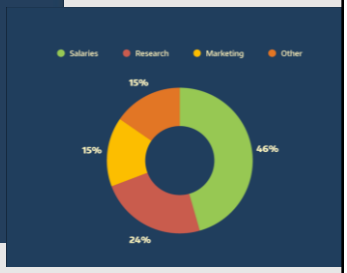
each **wedge** represents a category's contribution

Pie Charts



pie charts show **proportion** which is measured in **percentages** or **fractions**

Pie Charts



too many wedges can be confusing

pie charts are good for displaying data for around 6 categories

small categories can be grouped together as 'other'

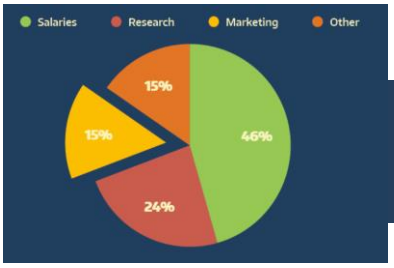
Pie Charts

arrange sectors clockwise in order of size

'other' should always be last

Pie Charts

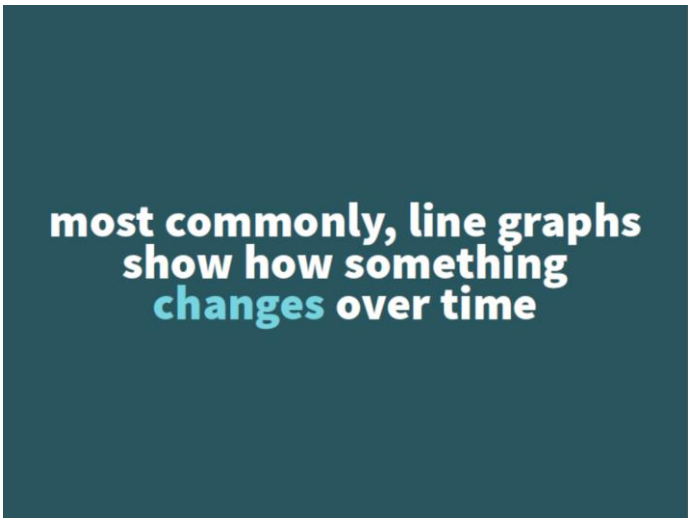
make sure segments are easily **distinguishable**



'other' should always be **last**

1. Pie Charts

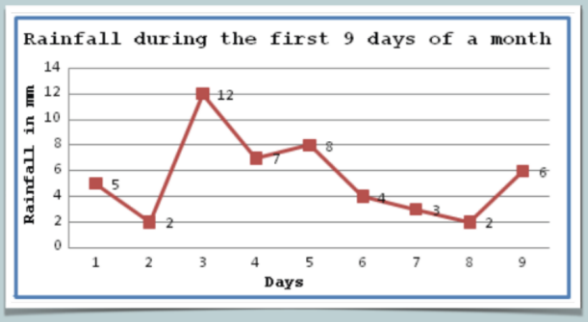
3. Line Charts


most commonly, line graphs
show how something
changes over time


They're useful for identifying
patterns and trends in data

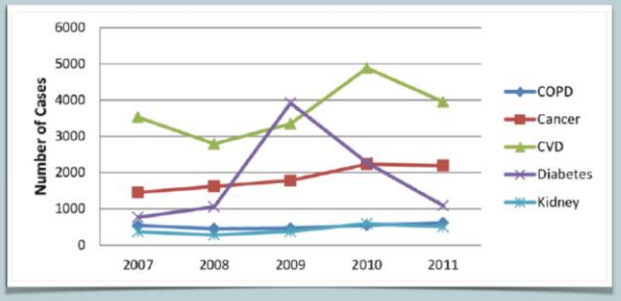
Types of line graphs

Basic



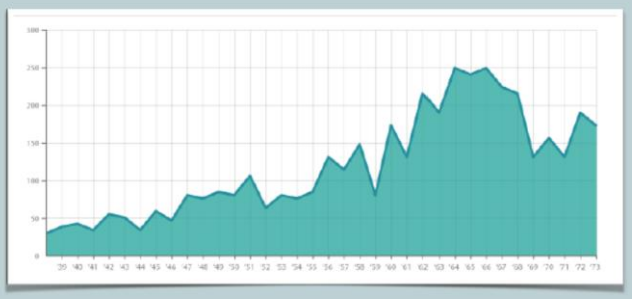
used to display one trend or theme

Multiple



used to display multiple relationships

Area Graph



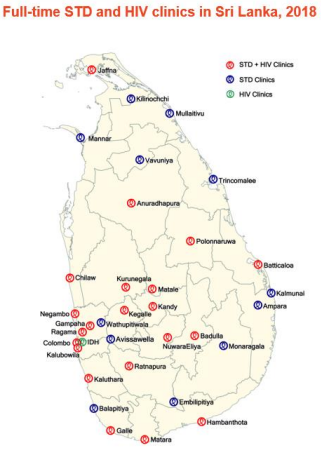
makes it easier to highlight certain points

Multiple Area

shows the relationship between sum totals

GIS

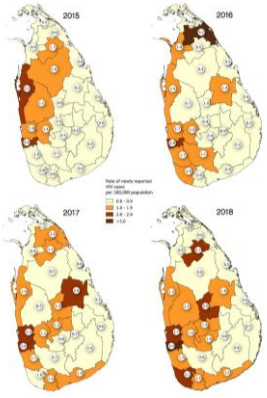
(Geographical information system)



GIS

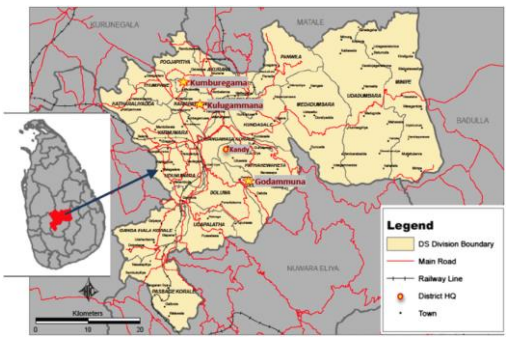
(Geographical information system)

Rate of newly reported HIV infections/ 100,000 population reported in 2015, 2016, 2017 and 2018



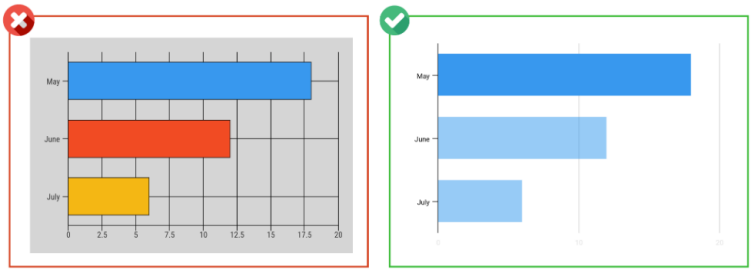
GIS

(Geographical information system)



Keep it Simple

Don’t introduce unnecessary complexity



15. Introduction to BIOSTATISTICS in Clinical Research

Role of Biostatistician

- Protocol development
- Study Implementation
  - Data Analysis
  - Data Management
- Report/Manuscript writing

What is Biostatistics?

It is an application of Statistics to the analysis of biological and medical data.

It is the field of study concerned with the collection, organization, summarization and analysis and interpretation of data.

Sources of Data

Routinely Kept Records

Surveys

Experiments

External sources

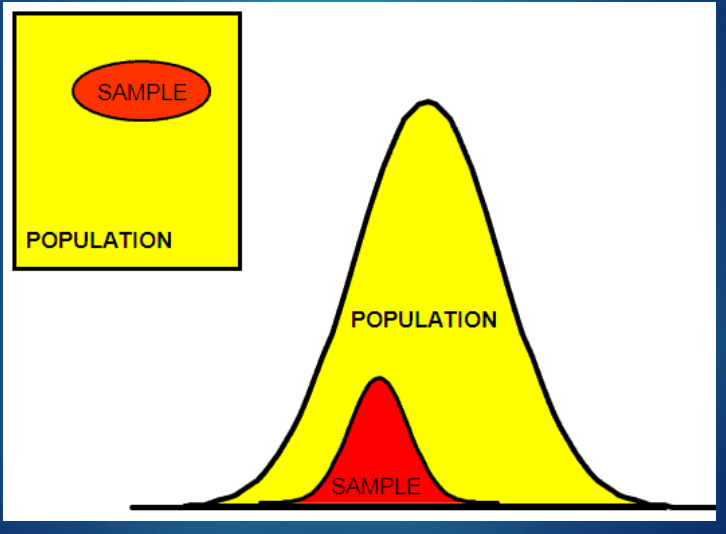
Population

It is the complete collection of data to be studied and it contains all the subjects and variables of Interest.

Sample

A sample is a subset of the population.

Diagrammatic Representation



Variable

A Variable is a charecteristic which differs from a person, place or things etc

Quantitative Variable

Measurements made on quantitative variables convey information regarding amount.
Ex – height. Weight, age

Qualitative Variable

Measurements made on qualitative variables convey information regarding attribute.
Ex – color of the eye

Measurement and Measurement Scales

It is the assignment of numbers to objects or events according to a set of rules.

Nominal Scale

Ordinal Scale

Interval Scale

Ratio Scale

Sampling

It is the Process of collecting information from the sample.

Types of Sampling

- Simple Random sampling
- Stratified Sampling
- Systematic Sampling
- Multistage Sampling
- Cluster sampling

Types of statistics

Descriptive Statistics

It describes or summarize the parameters of the population.

Inferential Statistics

It is the procedure by which we reach a conclusion about a population on the basis of information from the sample.

Descriptive Statistics

Tabulations

Diagrams

Graphs

Measures of central Tendency

Measures of dispersion

Measures of Association

Tables

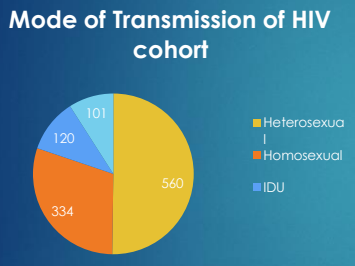
	Column heading	Column heading
Row identifier		
Row identifier		

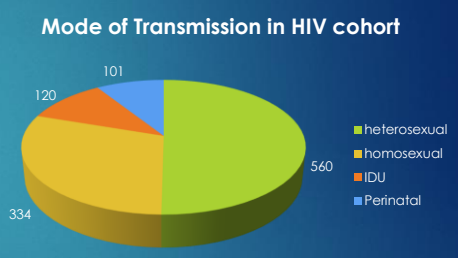
Footnote: Details of the abbreviations given in the body of the table.

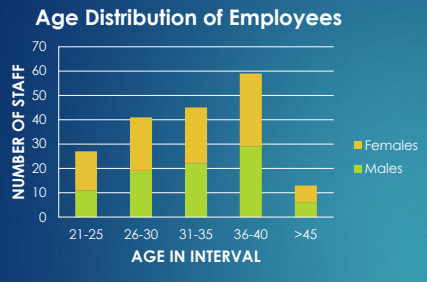
The table should be Complete and Accurate.

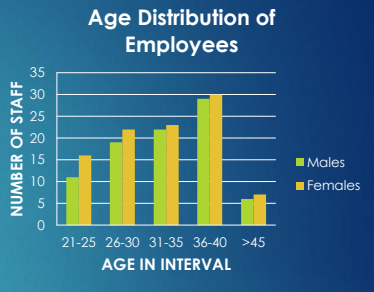
No repetition with texts or figures.

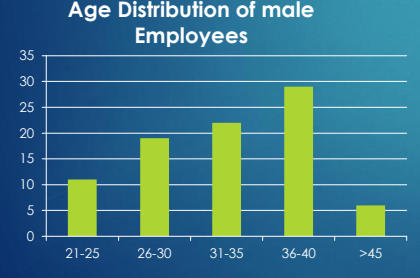
Pie diagrams

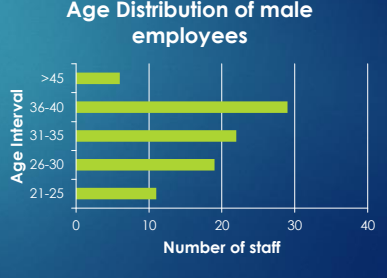












Other Diagrams and Graphs

Line graphs

Box Plots

Dot Plots

Histogram

Bubble plots

Pyramids

Basic Measures of Central Tendency

1. Mean
2. Median
3. Mode
4. Geometric Mean
5. Harmonic Mean
6. Weighted Mean
7. Truncated mean

Basic Measures of Dispersion

Range

Inter Quartile Range

Standard deviation

Measures of Association

Correlation (refers to a linear relationship between two variables or quantities)

Cross Tabulations (chi square tests)

Inferential Statistics

Types of statistical Inference

Estimation

Hypothesis Testing

Estimation Types

Point Estimation

Interval Estimation: Confidence Intervals

Estimation

Objectives ———> Non – Comparative

Example,

Objective: To determine the prevalence of Rheumatic heart disease in children age 5 – 15, India.

Summary Measure: % of children diagnosed with RHD based on an echocardiogram screening.

Example

Population: 18, 000 children between 5 – 15 years of age
Sample: 2, 400 children (Multistage random sample)

No. of children diagnosed with RHD: 72
Point Estimate: 72/2400 = 3%

Interval Estimation: 95% confidence Interval (2.2 – 3.8%)

Interpretation: We have 95% confidence that the prevalence of RHD is between 2.2% and 3.8%

Hypothesis Testing

Objectives ———> Comparative

Example,

Objective: To determine the effectiveness of HPV vaccine in the prevention of the transmission.

Study Design: Placebo controlled RCT

Hypothesis: The new HPV vaccine is effective for the prevention of transmission.

Summary measure: % of people infected with HPV in a two year period.

Components and Steps

Statistical Hypotheses

Level of significance

Statistical Test

Decision Rule

Data Analysis Decision

Interpretation

Statistical Hypotheses

Null Hypothesis (H0): Hypothesis of no difference.

Alternate Hypothesis (H1): An alternate statement about the
null hypothesis (one sided or two sided).

H0: $\mu 1 = \mu 2$

H1: $\mu 1 \neq \mu 2$ or H1: $\mu 1 \geq \mu 2$ or H1: $\mu 1 \leq \mu 2$

Example

H0: % of people infected with HPV is the same in both the groups. (New vaccine is ineffective).

H1: % of people infected with HPV is lower among the people vaccinated with new HPV vaccine than with the placebo (New vaccine is effective)

Level of Significance

The Level of significance (α) is the probability of rejecting the null hypothesis.

Possible action	True	False
Fail to reject H0	Correct action (1-α)	Type II error (β)
Reject H0	Type I error (α)	Correct action (1-β)

(1-α) – Confidence Interval

(1-β) – Power of the test

Decision based on the data

P value: Probability of obtaining the results as observed with the collected data under the assumption that H0 is true.

The P value depends on the statistical tests such as

Z –test
t – test
Chi square test
Model based test etc.

Decision Rule

If p-value < a, we do not reject H0

If p-value > a, we reject H0

Example

H0: % of people infected with HPV is the same in both the groups. (New vaccine is ineffective).

H1: % of people infected with HPV is lower among the people vaccinated with new HPV vaccine than with the placebo (New vaccine is effective)

Level of significance (a) = 0.05

Decision

% of people infected with HPV in a new vaccine group is 3%

% of People infected with HPV in a placebo group is 10%

Statistical test used
Z –test (Test for proportions)

P-value – 0.012

Decision
0.012<0.05, we reject H0. The vaccine is effective.

Statistical tools used for data analysis

- Analysis of Variance
- Simple Linear Regression
- Multiple linear Regression
- Logistic Regression
- Chi square tests
- Non Parametric Tests
- Vital Statistics

Simple and Multiple Linear Regression

Regression analysis is widely used for prediction.

It is used to obtain the linear relationship between a dependent variable and one or more independent variables.

Logistic Regression

When the dependent variable is a dichotomous variable, we can make use of logistic regression.

Chi square Test

Types of Chi – Square Tests

- Tests of goodness of fit
- Tests of Independence
- Tests of homogeneity
- Fisher's Exact Test
- Relative Risk and Odds Ratio
- Survival Analysis

Relative Risk

It is the ratio of risk of developing a disease among subjects with the risk factor to the risk of developing the disease among subjects without the risk factor.

Example

Subjects with and without the risk factor who became cases of preterm labor.

Risk factor	Cases of preterm labor	Non cases of preterm labor	Total
Extreme exercising	22	216	238
Not exercising	18	199	217
Total	44	415	455

Relative Risk = (22/238)/(18/217) = 1.1

This data indicates that the risk of experiencing preterm labor when a woman exercises heavily is 1.1 times as great as it is among women who do not exercise at all.

Odds Ratio

The odds for success are the ratio of probability of success to the probability of failure.

Example

Subject classified according to obesity status and mother's smoking status during pregnancy

Smoking status	Cases (obese)	Non cases (non obese)	Total
Smoked	64	342	406
Not smoked	68	3496	3564
Total	132	3838	3970

Odds ratio = (64/342)/(68/3496) = 9.6

The obese children are 9.6 times as likely as non obese children to have had a mother who smoked throughout the pregnancy.

Survival analysis

It is a branch of statistics which studies the amount of time that it takes before a particular event occurs.

The event may be death or time to failure or time to occurrence of an event.

Analytic Techniques

- Life Table Analysis
- Distribution plots
- Kaplan Meier product limit estimator
- Regression Models

Vital Statistics

Branch of statistics concerning the important events In human life such as births, deaths, marriages, and migrations.

Death Rates and Ratios

Measures of Fertility

Measures of Morbidity

Death Rates

- Annual crude death rate
- Annual specific death rate
- Standardized death rate
- Maternal mortality rate
- Infant mortality rate
- Neonatal mortality rate
- Fetal death rate
- Perinatal mortality rate

Measures of Fertility

- Crude birth rate
- General fertility rate
- Age specific fertility rate
- Standardized fertility rate

Measures of Morbidity

- Incidence Rate
- Prevalence Rate
- Case – fatality Ratio
- Immaturity Ratio
- Secondary attack rate

Incident Rate

It is the ratio of total number of new cases of a specific disease during a year to the total population.

Prevalence Rate

It is the ratio of total number of cases, new or old at a point in time to the total population at that point in time.

This rate is especially useful in the study of chronic diseases, but it may also be computed for acute diseases.

Case fatality Ratio

This ratio is useful in determining how well the treatment Programme for a certain disease is succeeding.

It is the ratio of total number of deaths due to a disease to the total number of cases due to the disease.

Data Management

- Data Capture
- Data Transcription
- Data Transfer
- Data Entry
- Data Cleaning
- Storage of hard copies of CRFs
- Storage of electronic data
- Data coding
- Dataset creation
- Data backup and recovery

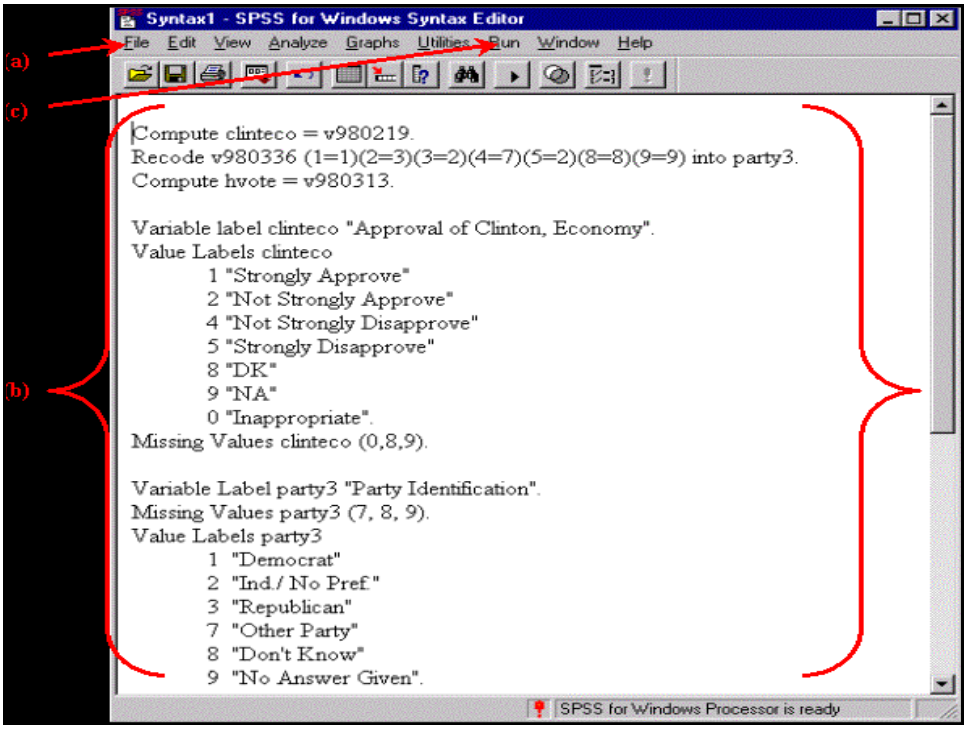
16. Statistical Package for the Social Sciences (SPSS)


Introduction – Why Use SPSS

- SPSS has been around since the late 1960s.
- SPSS is the statistical package most widely used by social scientists
- Of the major packages it seems to be the easiest with less coding
- One can use it with either a Windows point-and-click approach or through syntax (i.e., writing out of SPSS commands.). Each has its own advantages, and the user can switch between the approaches.

Syntax

- Originally, SPSS was written like a programming language. Users wrote SPSS syntax (often on a mainframe computer and even with key-punch cards) that performed the tasks they wanted.
- In SPSS-Windows, users can still use syntax by using the syntax editor.
- They would open the syntax window by clicking on *File*, dragging down to *New*, and choosing *Syntax*;
- Type the SPSS syntax that they want to run;
- Click on *Run* and drag down to *All*. (Alternatively, if they want to run only a few commands, they would highlight those commands, click on *Run*, and drag down to *Selection*.)



Choosing Appropriate Scales & measures

- There are many different ways of collecting data
- We need to measure output or performance on some objective criteria
- In choosing appropriate scales, need to aware of reliability and validity
- The reliability of a scale indicates how free it is from random error

Reliability and Validity

- Reliability means consistency
- Nunnally (1978) recommends a minimum level of .7 Cronbach Alpha value
- Validity refers to the degree to which it measures what it is supposed to measure.
- Research design → Objectives and goals of study → Questionnaire construction → code book preparation → data management and mining → selecting appropriate statistical tools and techniques to explore data towards achieving research goals and objectives.

Scales of Measurement

- The variables are of two types: quantitative and qualitative, and the measurement of variables or levels of measurement are of four types. They are nominal, ordinal, interval and ratio (N,O,I,R)
- Levels of measurement are very important and serve as a basis for which statistical tests are permitted on a given set of data and type of research.

Preparing a codebook

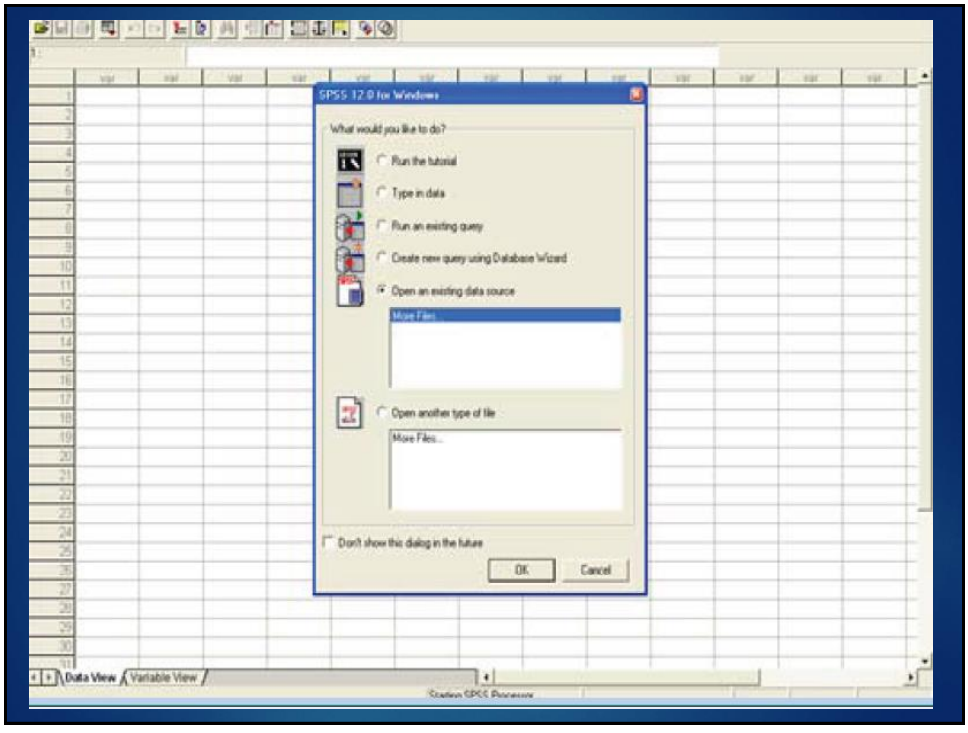
- Preparing the codebook involves deciding (and documenting) how you will go about
- Defining and labeling each of the variable
- Assigning numbers to each of the possible responses
- You should have a unique ID for each respondent or case
- Keep minimum on open ended questions

Rules for naming of variables

- Must begin with a letter (not a number)
- Cannot include period, blanks or other characteristics like (!, ?, *)
- Cannot include words used as commands by SPSS (all, ne, eq, to, by, or, gt, and, not, with etc)
- Cannot exceed 64 characters (SPSS V12) or 8 characters for earlier versions of SPSS

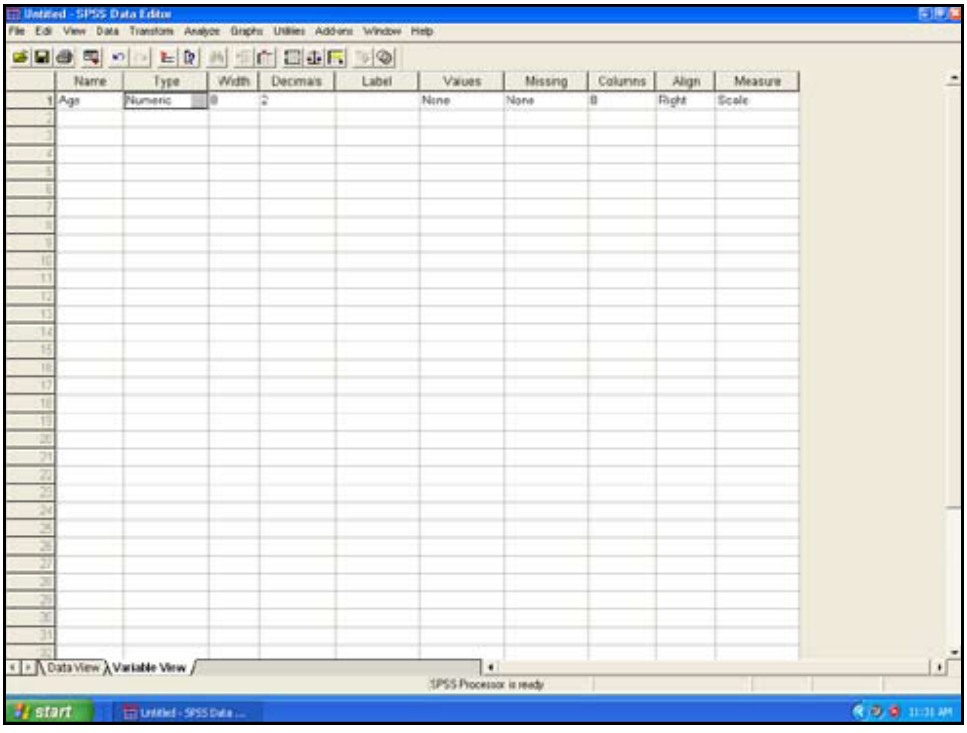
Opening an existing data file

- Like opening a word document, double click on SPSS .sav data file
- You will get an untitled blank spreadsheet like Microsoft Excel data sheet
- You can open an existing data file by using Open menu and then pick Data
- The data file will open which has data view and variable view
- You can click on save to save your file



Creating a data file and entering data

- Open an untitled data file
- Click on the 'Variable View" tab at the bottom of the Data Editor
- Variable view has 'Name', Type, width etc appear at the top of each column. Each row corresponds to a variable.
- Let us say you type "Age" under variable name, the other values appear as they are
- set by default. You can change all the values as per your requirements.



Variable Names

- First you need to decide on names for each of your variables
- Follow the rules involved in creating variable names
- Our tip: Avoid using symbols in variable names…..that way you won't need to remember which ones are okay and which one are not
- Specify correct variable type, column width, number of decimal points .

Defining variables and value labels

- Variable LABELS and Value Labels can be used to make output easier to understand.
- A variable label is used to descriptively label the variables
- Its use makes the output easier to read and can be very useful if the output is used over a long time period. Each label cannot exceed 60 characters.

Value Labels

- A Value label is used to descriptively label the values of a variable.
- Example of a value label:
- Lets do a value label for a variable named community

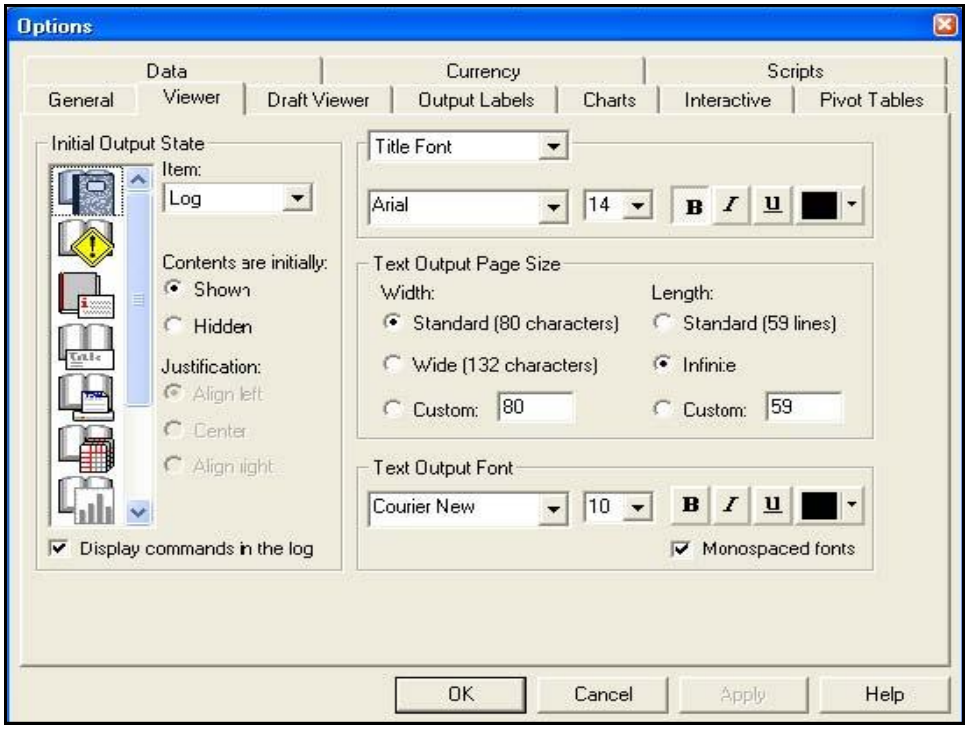
  A value of 1 refer to Asian, 2 refer to African American, 3 means Hispanic and so one
- For Age Category let us say 1 = 0-10 years; 2 = 11 -20; 3=20-45; 4=45-65 etc

Missing data

- The presence of missing data is very common in any kind of research. A research will come across dead mice, sick children, non-compliant household head, unfilled forms, lost samples and so on.
- We can't really do anything about it other than planning for it while creating the data file
- You can enter a special code like '8 for refusal' or '9=don't know'; or you can leave the item blank
- A good thing about SPSS is that any blank or period (".") is considered missing data unlike some other software programs which consider a blank as zero.

Changing the SPSS option

- It is always a good idea to check the SPSS options. Especially when you are working in a computer lab, students will change these options which can dramatically influence how the program appears. You can reset the option
- To reset the option, click on edit menu and select Option. You will see this option window



Data Entry using Excel

- You can create your data file in Excel and import it to SPSS
- Warning: Remember, Excel can deal with 256 columns of data and if the data is likely to be larger, it is probably easier to set it up in SPSS
- Once you import the data in SPSS, you have to do variable and value labeling. The variable name must conform to the SPSS rules for naming variables.
- Make sure you save the file with .sav extension

Screening and cleaning the data

1. Checking for errors: Check each of your variable by doing a descriptive statistics (frequencies)
2. Check minimum and maximum values and declare missing values appropriately

17. Key Learnings & Next Steps

MINISTRY OF HEALTH SRI LANKA

NATIONAL STD/AIDS CONTROL PROGRAMME

VOLUNTARY HEALTH SERVICES
VHS
WE SERVE THE PEOPLE WHO SERVE YOU

Key Learnings

Hands-on Training on

- Understanding datasets, components, structure & database management principles
- Variables & Indicators – Types and how to manage
- Data Quality Assessment & Adjustments using Excel
- Data Management using SPSS
- Exposure to Cohort Database using MS Access
- Data Triangulation
- Communication of Data Analysis Results

Next Steps…

- Work more on NSACP datasets
- Explore Excel functions and options
- Use SPSS with survey data
- Identify topics for analysis and commission
- Practice…. Practice…. & Practice….

Thank you

National STD/AIDS Control Programme (NSACP), Sri Lanka & The Voluntary Health Services (VHS), India

Supported by Centers for Disease Control and Prevention (CDC/DGHT-India) (VHS-CDC Project)