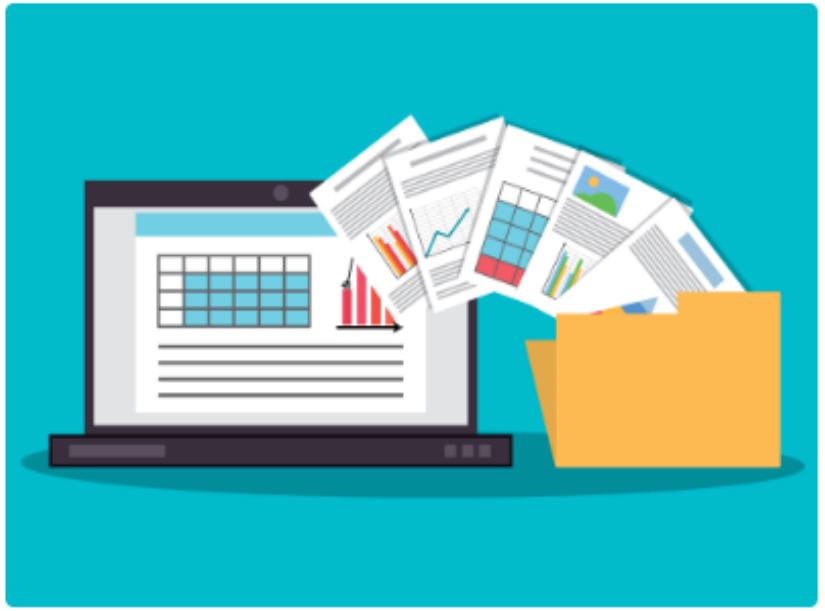
Plan for Data Management & Data Analysis



Dr. Myat Htut Nyunt

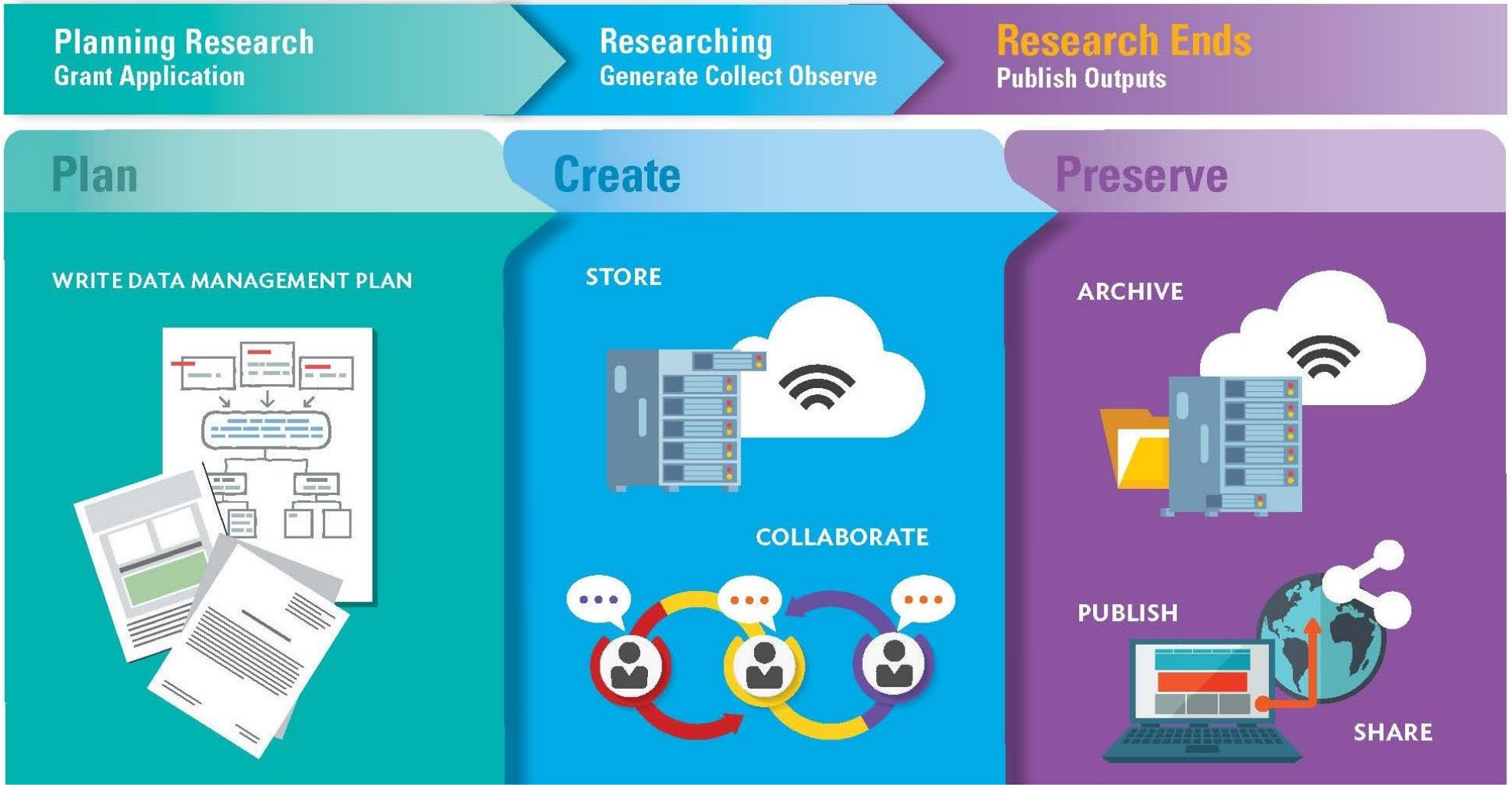
MBBS, MMedSc (Microbiology), DAP&E (IMR, Malaysia), PhD (Parasitology, KNU, Korea)
Deputy Director/Head, Bioinformatics Division, Department of Medical Research
myathtutnyunt@mohs.gov.mm
HP: +959-43106659

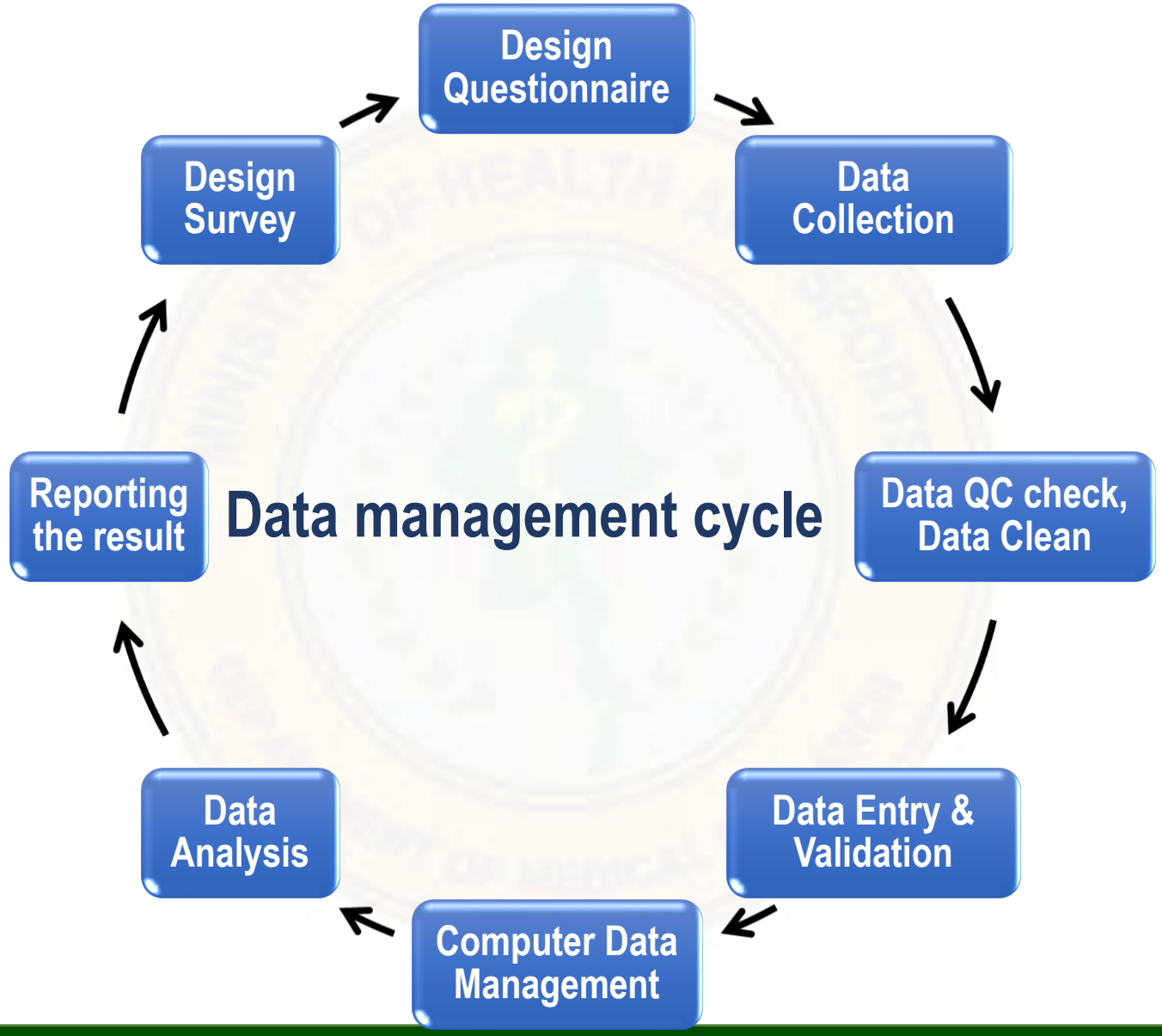
Learning Objectives

At the end of this presentations, the participants should be able

- ✓to understand the important of data management and statistical analysis plan.
- ✓to include the clear description on the plan of data management and statistical analysis in their proposal.
- ✓to criticize and/or review the research proposal for appropriateness of data management plan and statistical analysis plan.

Research Data Management (RDM)




Design Questionnaire
Data Collection
Data QC check, Data Clean
Data Entry & Validation
Computer Data Management
Data Analysis
Reporting the result
Design Survey
Data management cycle


Good Preparation
(Accuracy, Timeliness, Completeness, Conciseness, Relevancy
Good Quality of DATA
Good Quality of Research Result

Plan for

Data Management

1. Data Security
2. Data Sharing
3. Data Repositories/registration
4. Data Retention and Destruction
5. Data Backup
6. Data Policy

Data Analysis

- Plan for entry, validation, data-clean & analysis
- Plan for data presentation (dummy table)

1 Security

To keep sensitive research data secure and confidential

1. Technical safeguards
2. Physical safeguards
3. Good work practices

1 Protection from Viruses and Other Malicious Codes

- Always use -approved antivirus software.
- **Immediately** stop using any computer or software you suspect is infected
  - **Immediately** isolate the computer from any network.
  - Do not reboot the system (viruses are triggered to propagate upon system reboot)
  - If it appears that a negative activity is occurring, the system must be shut off and left off until a clean Antivirus boot media is used to clean the system
  - Employees not authorized to attempt recovery and restoration must not remove the suspected software themselves, but must contact a qualified IT Specialist
  - Only -approved software and tools may be used to attempt recovery from infection with a virus or other malicious code
- If a non- technician is called to work on non- owned equipment, use caution to protect the information.
- If a hard drive or other storage medium that contains research data becomes infected, never surrender or swap it with an outside party

Data Sharing

3 Data Repositories

Gene and protein sequences submission

Nucleotide sequences should be deposited in one of the three major collaborative databases: **EMBL, GenBank**, or **DDBJ**. Protein sequences should be deposited with **UniProt.**

NCBI http://www.ncbi.nlm.nih.gov/genbank/submit/

ENA European Nucleotide Archive http://www.ebi.ac.uk/ena/submit

DDBJ DNA Data Bank of Japan http://www.ddbj.nig.ac.jp/submission-e.html

UniProt http://www.uniprot.org/help/submissions

- De-identified data should be shared so that others can verify your conclusions or analysis
- Sharing of personal patient information is NOT a good practice as noted in Privacy sections earlier.

3

Clinical Trial registration

ClinicalTrials.gov
A service of the U.S. National Institutes of Health

Search

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

Home > Find Studies > Search Results > Study Record Detail

Trial record 1 of 1 for: NCT02708199

Previous Study | Return to List | Next Study

Molecular Assessment of Drug Resistance Markers in Asymptomatic Malaria Cases and Malaria Antibody Kinetic

This study has been completed.

Sponsor:
Department of Medical Research, Lower Myanmar

Collaborator:
Kangwon National University

Information provided by (Responsible Party):
Myat Htut Nyunt, Department of Medical Research, Lower Myanmar

ClinicalTrials.gov Identifier:
NCT02708199

First received: March 5, 2016
Last updated: March 9, 2016
Last verified: March 2016
History of Changes

Full Text View | Tabular View | No Study Results Posted | Disclaimer | How to Read a Study Record

ANZCTR
Australian New Zealand Clinical Trials Registry
The ANZCTR is an online registry of clinical trials being undertaken in Australia, New Zealand and elsewhere.

Thai Clinical Trials Registry
www.clinicaltrials.in.th

MHRR
Myanmar Health Research Registry

HOME | REGISTRY | KNOWLEDGE SHARING | UNIVERSITY RESEARCH | RESEARCH FUNDING | MEMBERSHIP | FAQs

Sri Lanka
Clinical Trials Registry
Managed by the Sri Lanka Medical A

PACTR
Pan African Clinical Trials Registry

4 Data Use (Transfer) Agreements

A data use agreement is a written contract that defines the following:

- What data may be used
- How data may be used
- How data will be stored and secured
- Who may access data
- Legal authority under privacy for access to data
- Disposition of data after the research has been terminated
- Actions required if data are lost or stolen

4 Data Retention and Destruction

- You must retain research data in accordance with local and IRB policies, protocol sponsor guidelines, or Privacy Act system of records notice, whichever is most restrictive.
- Once the required retention period has lapsed, the data may be destroyed using a method that will render them unreadable, undecipherable and irretrievable.

- *Note: This pertains to both and non- owned computer equipment and storage devices.*

- *Note: Pushing the delete button is not sufficient to permanently delete data.*

5 Backups

- You must backup essential data and software at regular intervals and treat backups and archives according to their security classification.

- You also must securely store any backups containing sensitive research data.

Note: Do not store original sensitive research data on laptops or portable media.

- Thumb drives can store considerable data and are easy to misplace or lose.
  - Use physical locks
  - When in an uncontrolled environment, follow “clear desk” practices for media to reduce the risk of unauthorized access to, loss of, and/or damage to the sensitive research information

6 Data Policy and ownership of Data

- Who owns the data that is generated?
  - Patient?
  - Institution?
  - Funder?
  - Investigator?
  - Publisher?

Plan for Data Analysis

*Data Quality Check in every steps.

Steps

- Before data collection (Questionnaires or data collection tool development, Pre-test, Interviewer training, ...)
- Data Collection

(Precoded-Proforma/Questionnaires, checklists, measurement, observations, mobile data collections tools [ODK, Kobo Toolbox, Google Form, Microsoft Form, Survey Monkey...])

- Data Clean
- Data Check/validate (MS Excel, Epidata, ...)
- Data Entry (MS Excel, Epidata, SPSS, PSPP, R, ...)
- Data Analysis (SPSS, GraphPad Prism, EpiInfo, Atlas-Ti, NVivo, RQDA,...)
- Data Visualization (ArGIS, QGIS, GraphPad Prism, MS Excel, yEd, EdrawMax, ...)

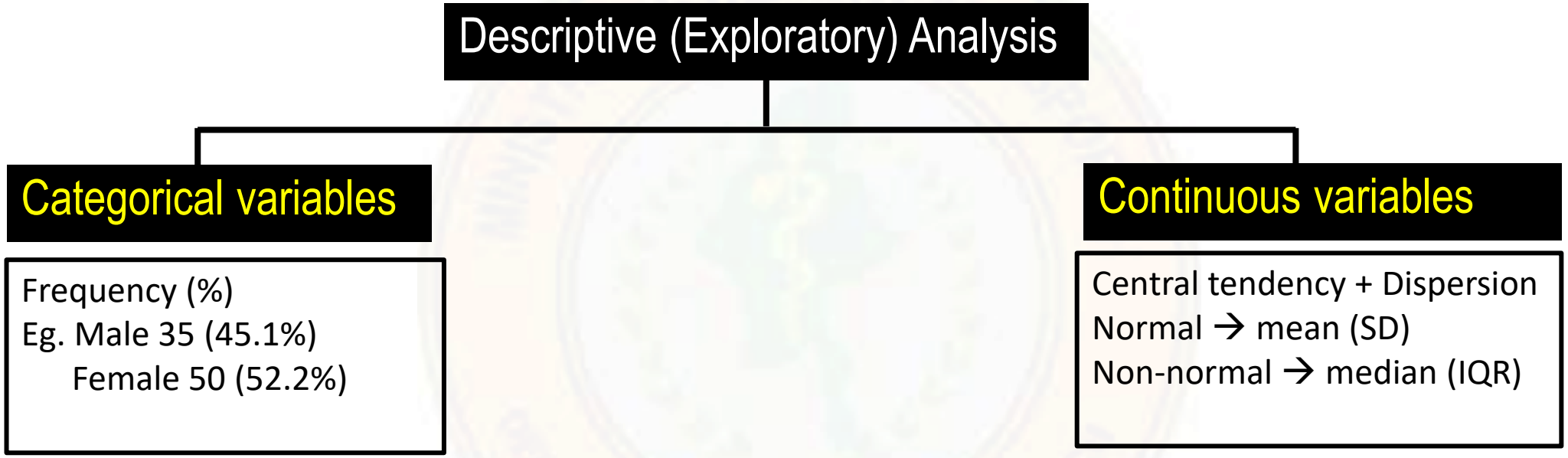
Plan for Data Analysis

To decide the data analysis plan, **THREE** considerations,

1. Objectives & Hypothesis
2. Study Design
3. Variable collected

Choice of statistical tests

1) **Descriptive (exploratory) analysis**
2) Inferential statistics (Hypothesis testing)



Choice of statistical tests

1) Descriptive (exploratory) analysis
2) **Inferential statistics (Hypothesis testing)**

Inferential Analysis (compare)

Categorical variables

Chi-square
Fishers' exact (expected count less than 5)
Paired→ McNemar's test

Continuous variables

	Normal Distribution	Non-normal Distribution
	Parametric Tests	Nonparametric tests
Paired Samples	Paired T test	Wilcoxon Signed Rank Test
Two independent samples	Independent T test	Mann-Whitney U test
More than 2 groups	ANOVA	Kruskal-Wallis test

Choice of statistical tests

1) Descriptive (exploratory) analysis
2) **Inferential statistics (Hypothesis testing)**

Association/correlation

- to quantify association between an independent (predictor) variable and continuous dependent (outcome) variable.
- Pearson correlation (normal distribution) or Spearman correlation (non-normal distribution or ordinal variables)
- Regression (continuous variable) or logistic regression (binary dependent variable)

Strength of association

- between two binary predictor and outcome variable.
- Odds Ratio (Case-control study)
- Relative Risk (Cohort study)

Choice of statistical tests

1) Descriptive (exploratory) analysis
2) **Inferential statistics (Hypothesis testing)**

Agreement

- Kappa agreement (no standard test, multiple measures on same population)

Reliability

- Cronbach's alpha, split half reliability

Time to event

- Survival analysis (Kaplan-maier analysis)
- Proportional hazards analysis (Cox regression)

Diagnosis Accuracy

- Accuracy (Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, Likelihood ratio)
- Receiver operating characteristic (ROC) curve

Qualitative Data Analysis

Steps

- Data collection (triangulation)
- Coding
- Identifying themes, patterns, relationships
- Summarizing

Content analysis	process of categorizing verbal or behavioural data to classify, summarize and tabulate the data.
Narrative analysis	reformulation of stories presented by respondents taking into account context of each case and different experiences of each respondent.
Discourse analysis	analysis of naturally occurring talk and all types of written text.
Framework analysis	advanced method that consists of several stages such as familiarization, identifying a thematic framework, coding, charting, mapping and interpretation.
Grounded theory	starts with an analysis of a single case to formulate a theory. Then, additional cases are examined to see if they contribute to the theory.

Data Analysis Plan (Sample)

- All data will be coded, edited and entered to computerized system using EpiData (version 4.6). Categorical data will be described in parentage with frequency table. Continuous data will be described in central tendency and dispersion. Data analysis will be carried out by SPSS ver 25.
- Chi-square test will be used as a test of association for categorical data. If the expected count is less than 5, Fisher's exact test will be used with 95% Confident Interval.
- -------
- Qualitative analysis will be done for transcription and transfer to NVivo software (ver. 12). Matrix will be prepared for each themes of interest. After reading computer output of the themes and full content of transcript, findings will be written.

Common errors and how to fix

Step	Problem	How to prevent
Data collection	• Missing data - No response - Missing - Filling error - Skip pattern error • Single/Multiple response mixing	- Compulsory mark in mobile collection - Pretest - "No response" should be included in the form - Manual check for competency of data - Training before data collection - "Multiple response possible" should be mentioned in multiple response questions
Data Entry	• Transposition (39-93 in entry) • Spelling error • Coding error • Entry error	- Double entry - Coding rather than direct-typing - Validation by data entry software - Proper coding and value label in software
Data Analysis	• Poor knowledge on statistics • Unfamiliar to software	- Learn - Consult with statistician

References

- http://www2.le.ac.uk/services/research-data/rdm/what-is-rdm
- Biostatistics (World Health Organization)
- Medical Statistics: a textbook for the health sciences / Michael J. Campbell,
- David Machin, Stephen J. Walters. 4th edition.
- Statistics for Epidemiology, Nicholas P.Jewell
- Handbook of Applied Statistics in Pharmacology (Katsumi Kobayashi and K. Sadasivan Pillai)
- Introduction to Biostatistics (Robert R Sokal and F James Rohlf)
- Essentials of Medical Statistics (Betty R. Kirkwood and Jonathan A.C. Sterne)
- Medical Statistics Made Easy (M Harris and G Taylor)


Any Questions