Statistics – Role of a statistician, programming

Module 12 Topic 6

Role of Biostatistician

Defining Research Question

- Reasonable & Worthwhile

Defining Hypothesis

- Null hypothesis - H0
- Alternate hypothesis - H1
- Based upon:
  - Study Design
  - Observed Differences
  - Incidence & prevalence of said condition/outcome

Planning of the study?

- Study Design
  - Superiority /Non-inferiority
  - Open /blind
  - Controlled /uncontrolled /Stratified
  - To Minimize bias
  - To enable appropriate data collection for analysis
- Sample Size Appropriate
  - Adequate to answer research question
  - Sample size should be the minimum required to detect a desirable difference
  - Larger size more precise estimates
  - Funds? Time? Ethical issues?

Planning of the study (contd)

- Assessment parameters
  - Which?
    - Objective Vs Subjective
  - How often?
    - Frequency
  - How many?
    - Too many parameters – Failure of conclusion

Randomization

- Random allocation of intervention
  - RCTs – Gold standard in CTs
  - Randomization – Crucial in RCTs
  - Stratified randomization
  - Block randomization
  - Randomization Code
  - Randomization Ratio
    - 1:1 , 1:2, 1:3, 1:1:2
  - Serial assignment /chronological order
  - Web based central randomization

Data Collection

- Designing of CRF
  - Log
  - Visit-wise
  - Forward / Backward
  - Open / Closed
  - e-CRF / e-Data Capture [Web based trials]
- To standardize capture of relevant data
- To facilitate efficient & complete data recording, processing, analysis & reporting

Data entry in CRF

- Minimum text — Use of strings
- Options for tick-mark answers — Unambiguous
- Numerical data entry — Emphasis on decimals

- Clinical Trial Database
  - Use of PC based software
    - FoxPro, Oracle, dBASE etc.
  - Single/Double/Triple data entry
  - Data retrieval

Data analysis

- Data Processing
  - Tests for normality, skewness & homogeneity of variances
  - Data arrangement
  - Use of PC Based software
- SAS, STATA, SPSS, NCSS, Add-ons (Excelstat, Analyze-IT)

Data Analysis

Test whether data fits into any of the known mathematical model:

"Goodness of Fit test"

Find out relation between two different aspects of one population (blood pressure & heart rate)

or

one aspect of 2 different populations

(blood pressure in young and old)

If exists, nature of relation:

Correlation analysis

Data Analysis

How close it is?

Given one set of values can we predict other?

:Regression analysis

Establish confidence limit

If you repeat the study, how close the statistic calculated by you would lie wrt population parameter

Interpretation & Conclusions

Statistical significance ≠ biological/ clinical significance

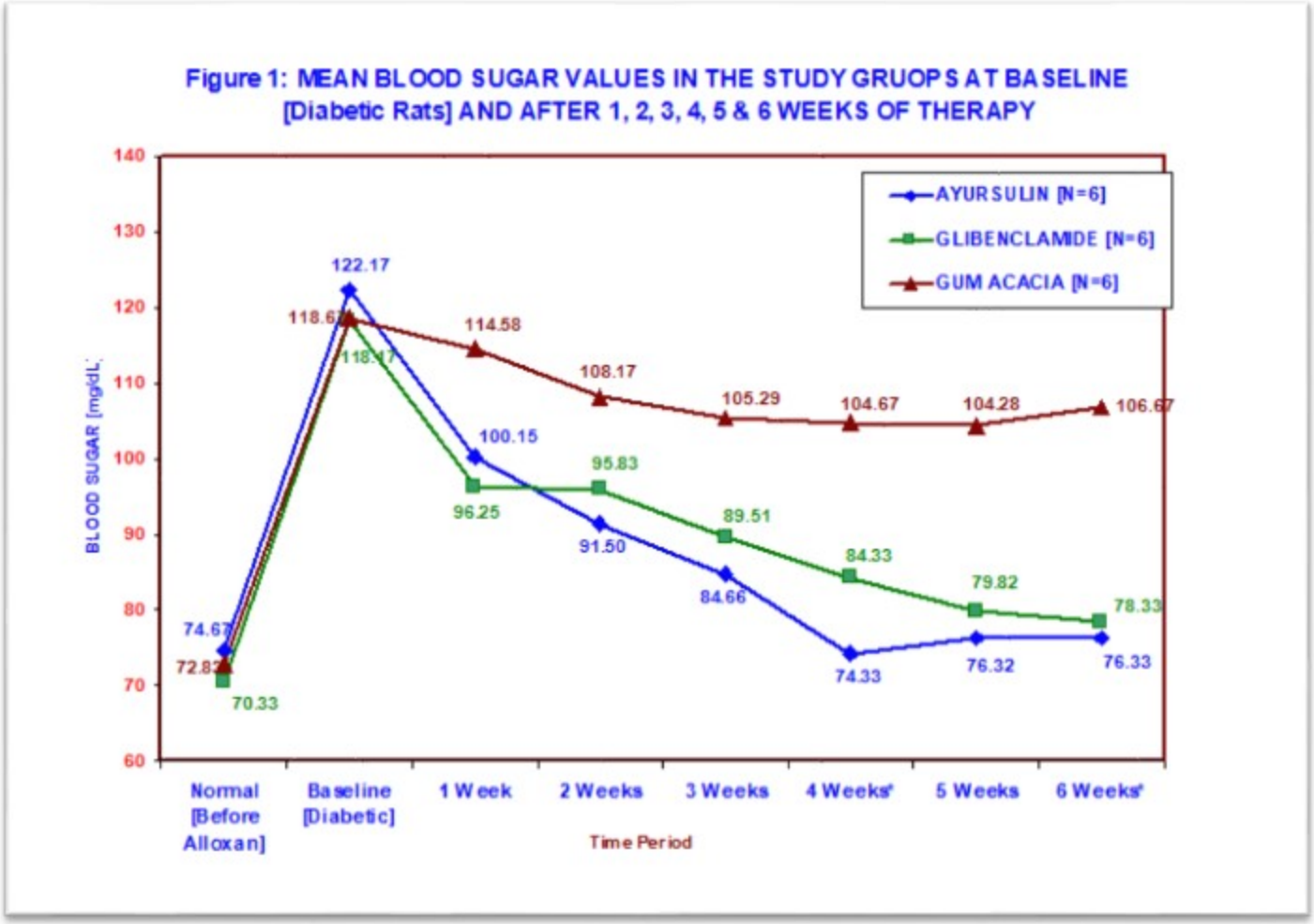
In-vitro significance ≠ in-vivo significance

To derive reasonable conclusions

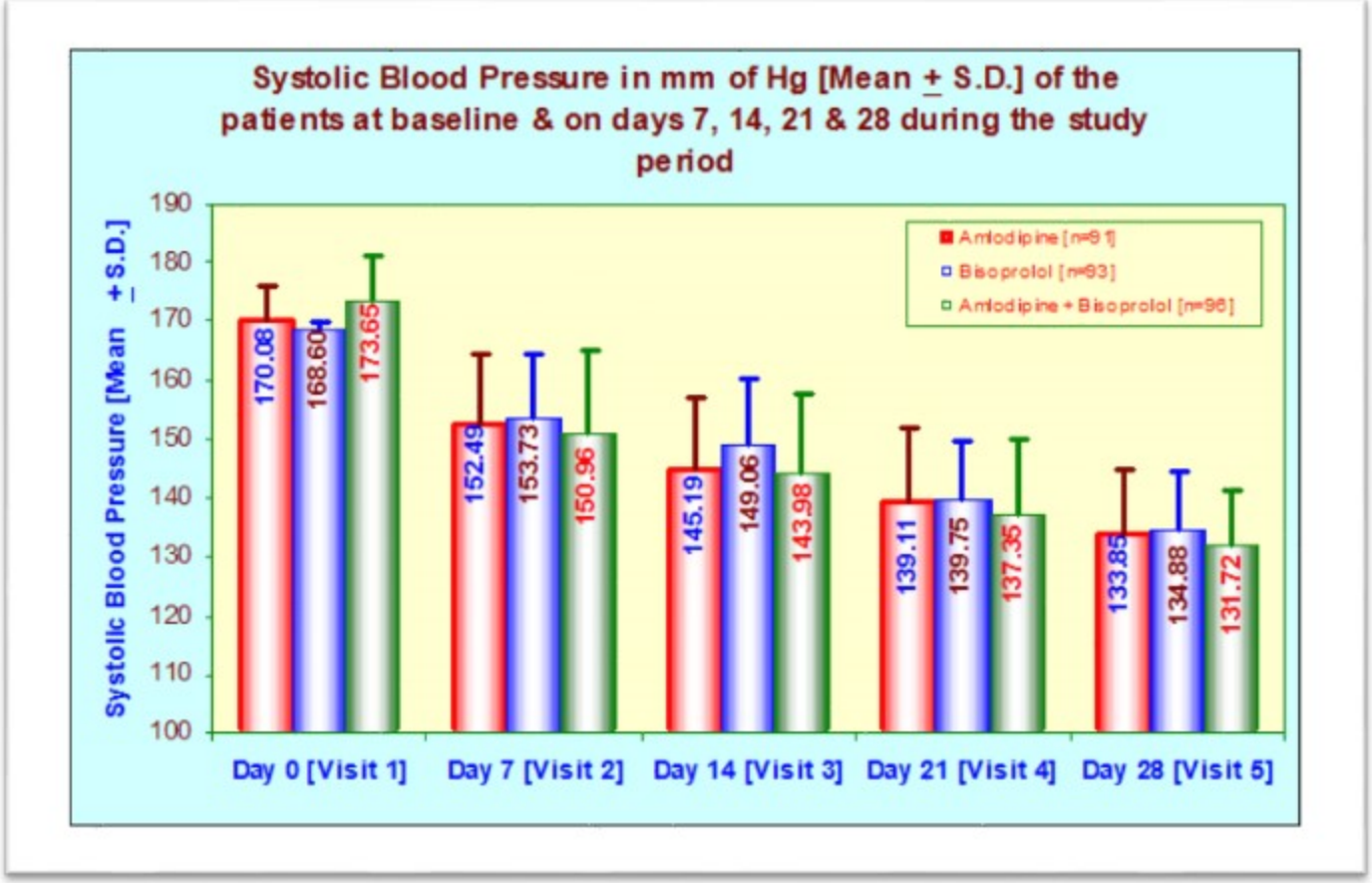
Data Presentation

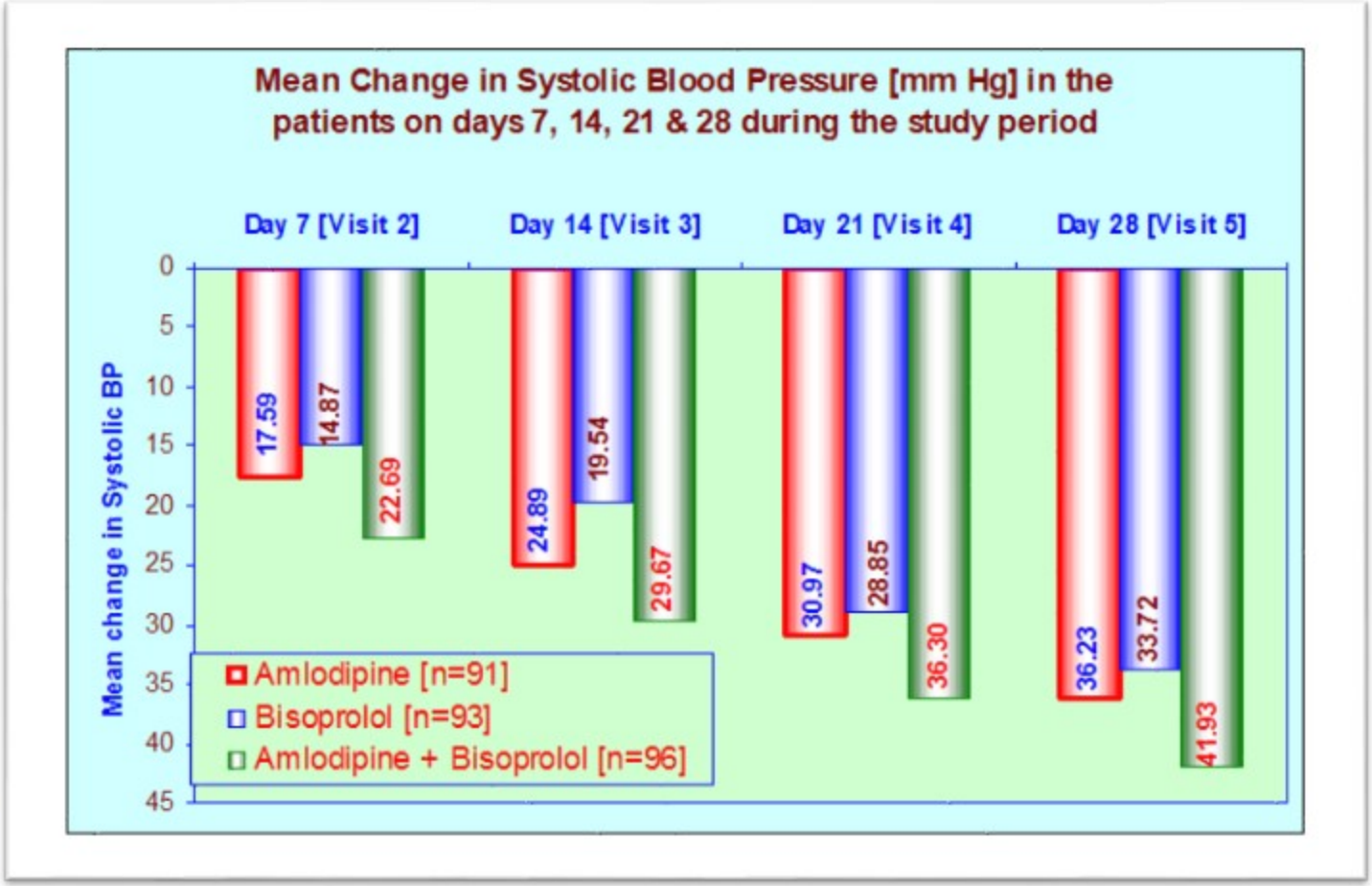
- Tabulation
  - Simple table
  - Frequency distribution table
- Charts & Diagrams
  - Qualitative data
    - Histogram
    - -Freq. plygon
    - -Line diagram
    - -Frequency Curve
  - Quantitative data
    - Bar diagram
    - Pie /sector diagram
    - Scatter plot
    - Box plot
    - Pictogram

Line Diagram

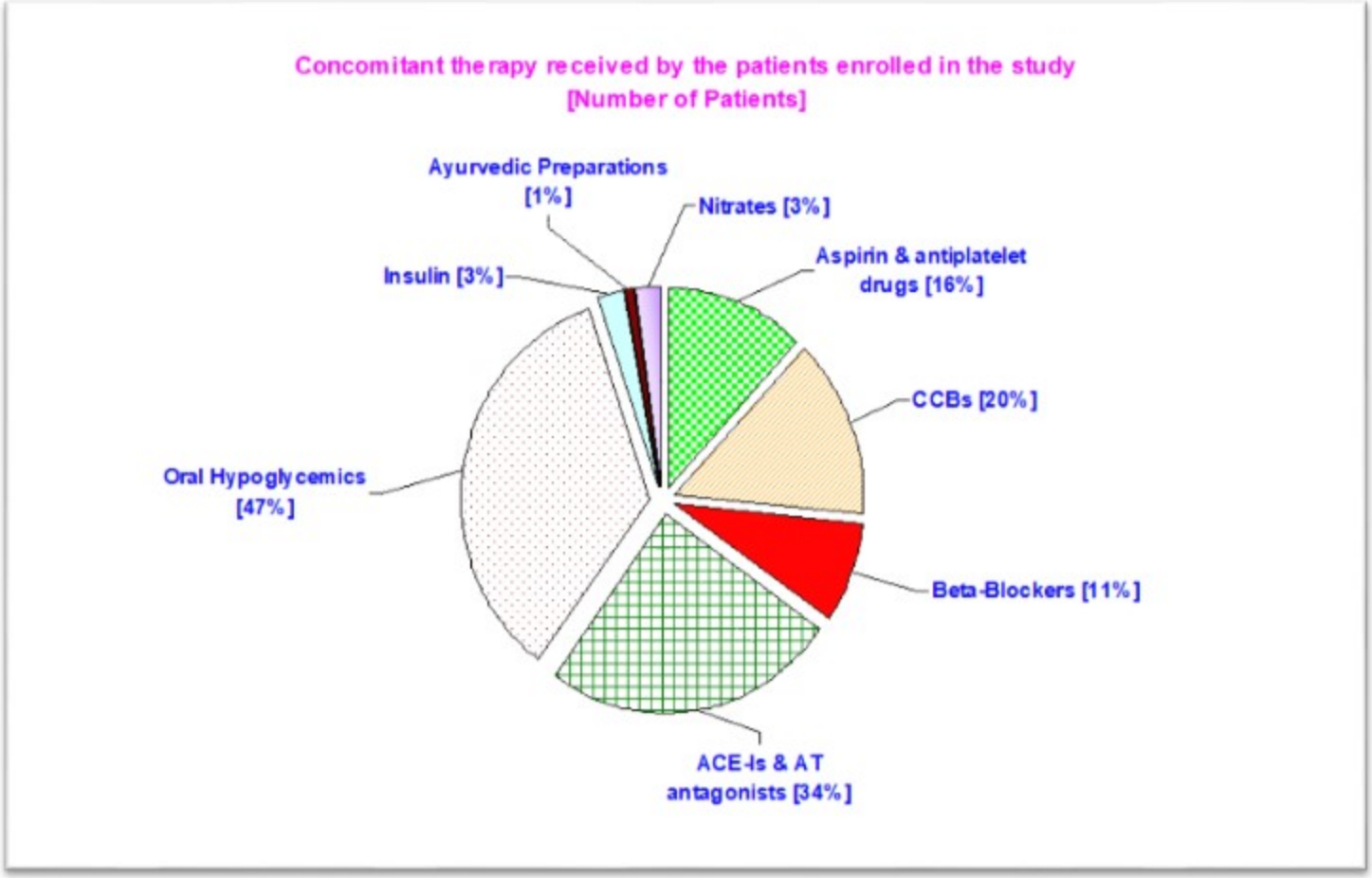


Multiple Bar Diagram

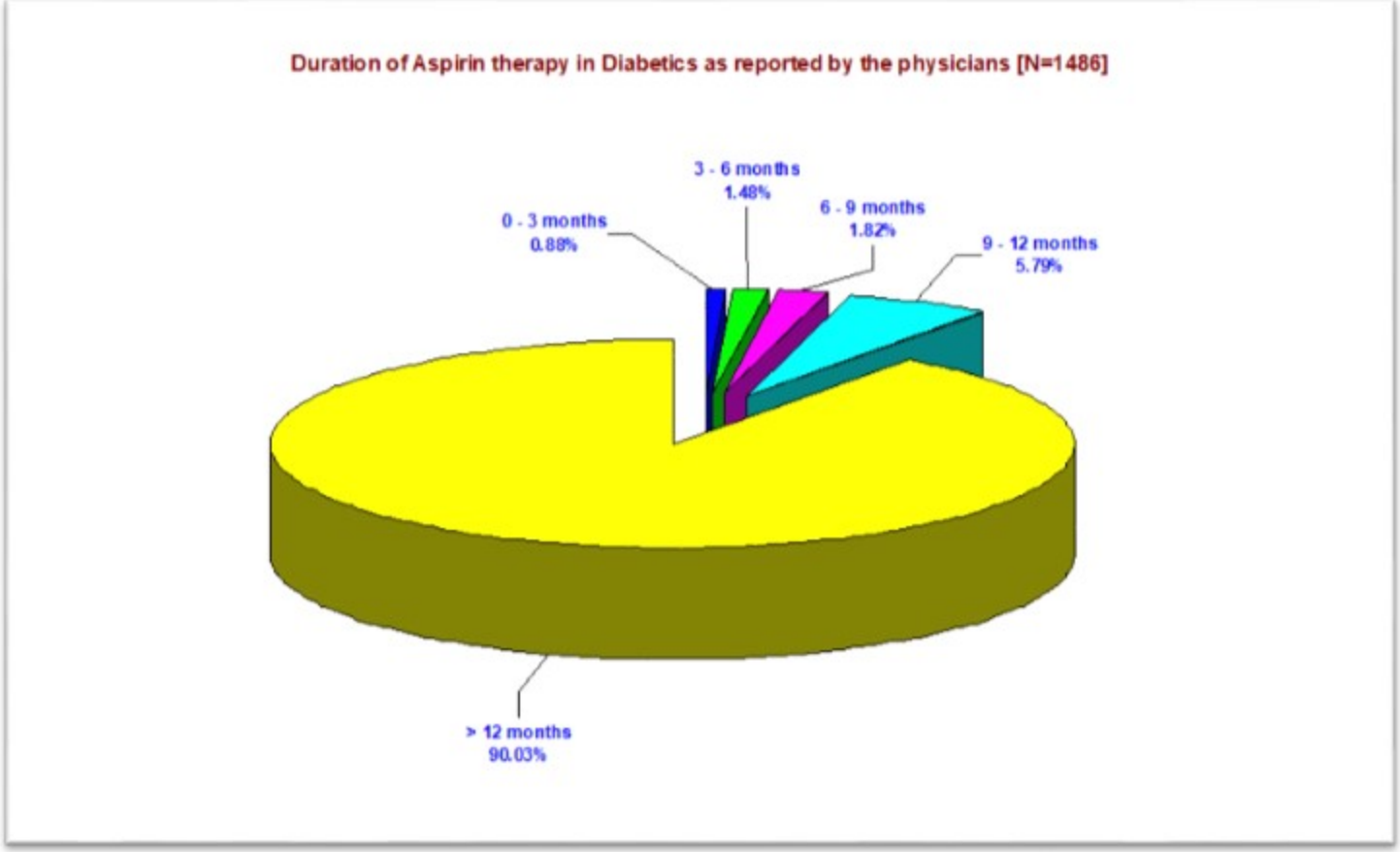



Mean Change in Systolic Blood Pressure [mm Hg] in the patients on days 7, 14, 21 & 28 during the study period
Day 7 [Visit 2]
Day 14 [Visit 3]
Day 21 [Visit 4]
Day 28 [Visit 5]
Mean change in Systolic BP
0
5
10
15
20
25
30
35
40
45
17.59
14.87
22.69
24.89
19.54
29.67
30.97
28.85
36.30
36.23
33.72
41.93
Amlodipine [n=91]
Bisoprolol [n=93]
Amlodipine + Bisoprolol [n=96]

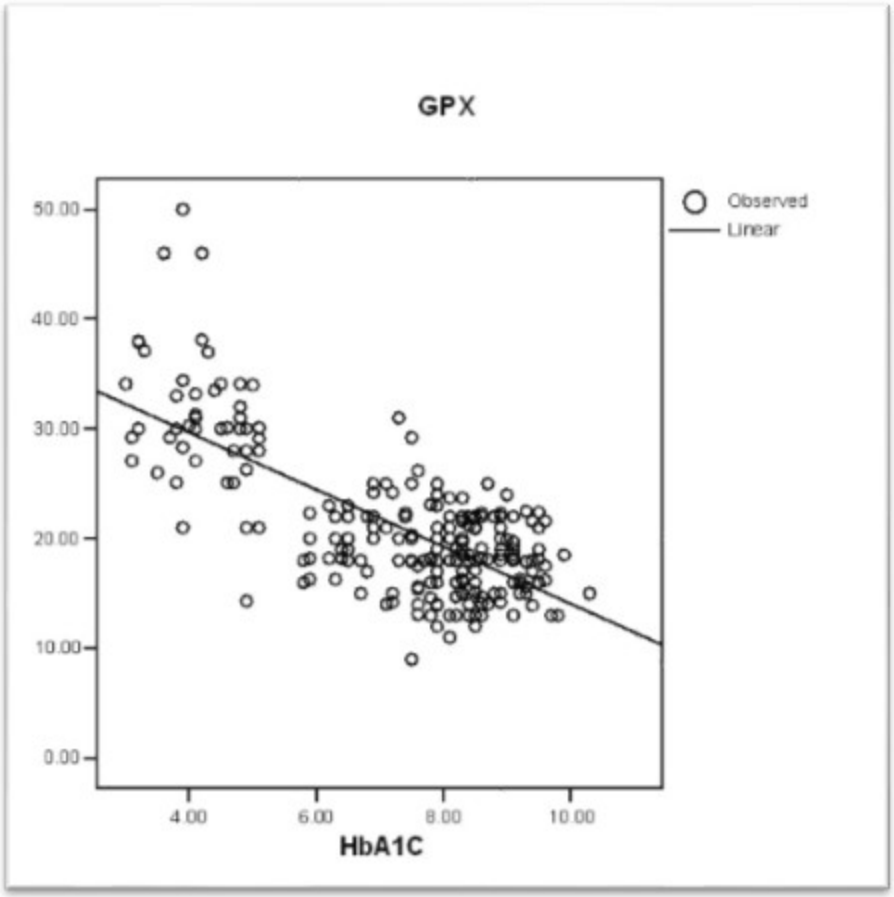
Pie Diagram


Concomitant therapy received by the patients enrolled in the study
[Number of Patients]
Ayurvedic Preparations [1%]
Nitrates [3%]
Aspirin & antiplatelet drugs [16%]
Insulin [3%]
CCBs [20%]
Oral Hypoglycemics [47%]
Beta-Blockers [11%]
ACE-Is & AT antagonists [34%]

3-D Pie Diagram



Scatter Plot



Statistical Programming

- When a CRF is being annotated, statistical programmers map it to SDTM
- SDTM is Standard Data Tabulation Model
- He creates the final Analysis Data Sets
- He creates Dummy Tables Figures Listings
- He programmes the data to fall into these Tables
- All this is done during data base build stage
- Finally analysisand statistical report generation

Role of statistician vs trial stages

- Statistician has a role at various stages in a trial
- Initiation and data base build stage
  - Inputs in protocol creation
  - CRF creation
  - Dummy TFLs
  - Statistical analysis plan
- Trial conduct- programming for creation of ADS and TFLs
- Trial close - Creation of ADS, TFLs, analysis and stats report

References

- Aggarwal YP, Statistical Methods, concepts, application, computation. Sterling publishers (1986)
- Varalakshmi V et al. Statistics – Higher Secondary, First Year. Tamil Nadu Textbook Corporation 2005
- Langley R, Practical Stats. The Chauser press 1968
- Barkan H, Annals of Cardiac Anaesthesia (2015)18:74
- Dunn OJ, Clark VA, Basic stats, a primer for biomed sciences. 4th Ed, 2009,John Weiley and sons inc.
- Harris M, Taylor G. Medical Statistics made easy. 1st Ed, Martin Dunitz,2003
- Krousel wood M, clinicials guide to stats part 1 The Ochsner Journal (2006) 6:2,68
- Kadam P ,Sample size calculation . International J of Ayurveda (2010) 1:55-57
- Fox N, Hunn A, MathewsN, Sampling and sample size calculation. The NIHR RDS EM / YH (2009)
- Available at http://www.biostathandbook.com/kruskalwallis.html
- Jaykaran. Journal of pharmaceutical negative results.(2010)1:61
- NayakBK. HazraV. Indian J Ophthalmol. (2011) 59 (85–86).
- http://wise.cgu.edu/choosemod/test61.html