

Dr SB Bhattacharyya

MBBS, MBA

Head – Health Informatics, TCS

Ex President, IAMI

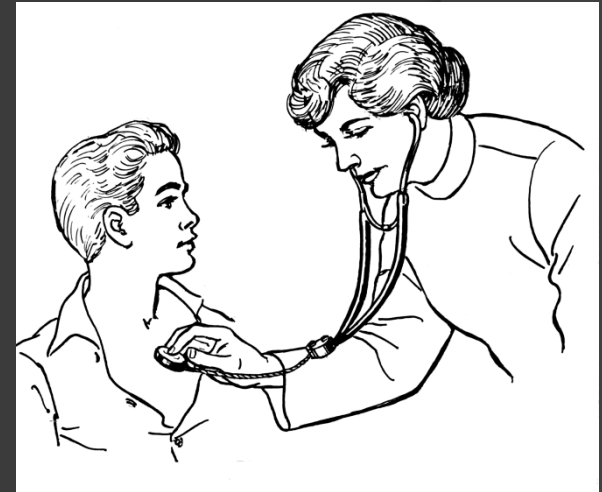
Ex Hony. Secretary, IMA Faridabad

Member, EHR Standards Committee, MoH&FW, GoI

CLINICAL DATA ANALYTICS

Stethoscope

- Stethoscopes are often considered as a symbol of the doctor's profession, as doctors are often seen or depicted with stethoscopes hanging around their necks.
- A 2012 research demonstrated that among several icons of medical profession, the stethoscope had the highest positive impact on the perceived trustworthiness of the practitioner.



Stethoscope

“That is will ever come into general use, notwithstanding its value, is extremely doubtful because its beneficial application requires much time and gives a good bit of trouble, both to the patient and the practitioner because its hue and character are foreign and opposed to all our habits and associations. There is something even ludicrous in the picture of a gray physician proudly listening through a long tube applied to the patient’s thorax.”

– London Times, 1834

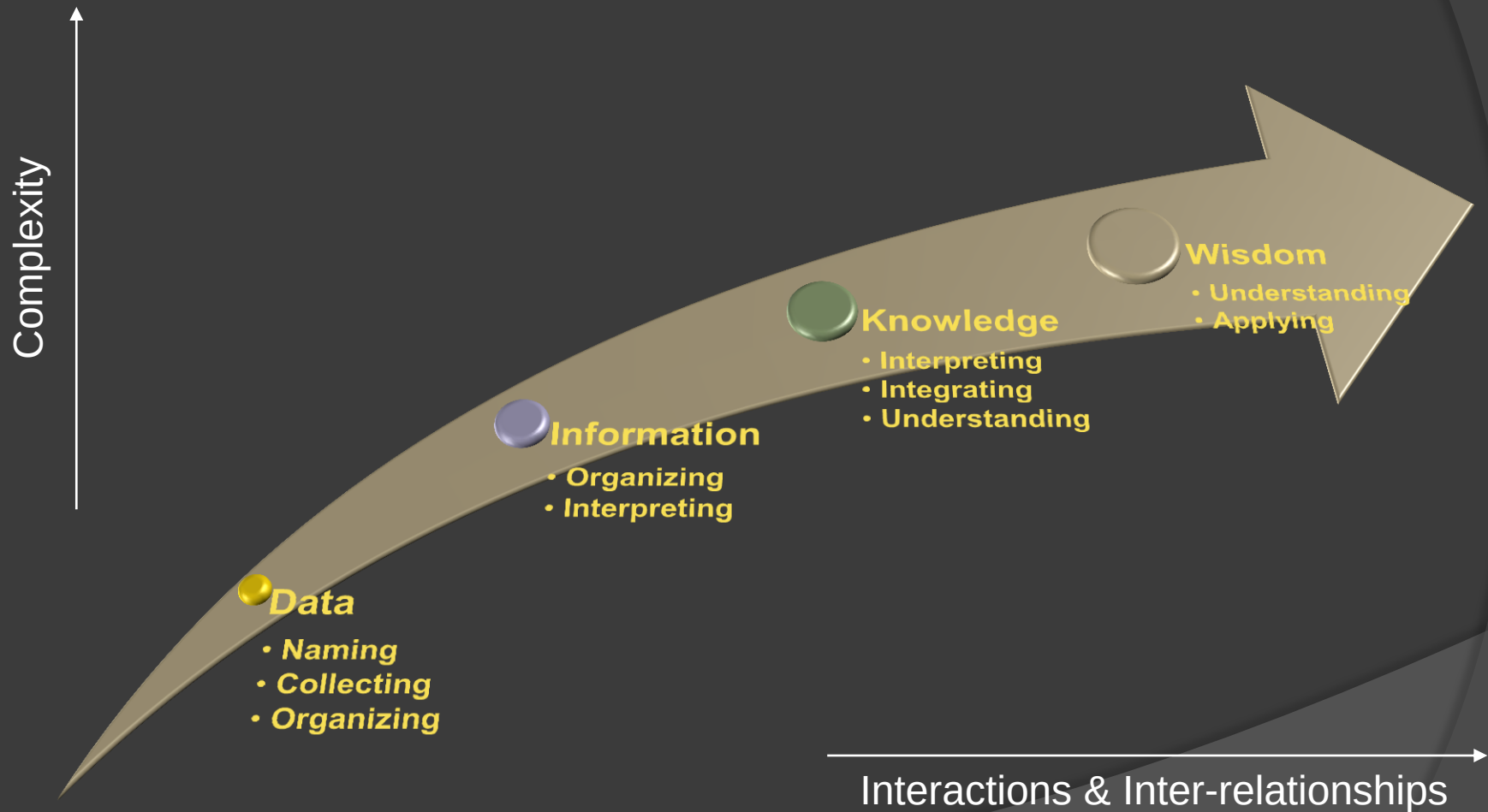


“If it were not for the great
variability among individuals,
medicine might as well be a
science and not an **art**”

“The **good** physician treats the
disease; the **great** physician
treats the **patient** who has the
disease.”

Sir William Osler, 1892

Nelson's Data to Wisdom



Hertfordshire Records :: DOHAD

- Meticulous birth records were maintained throughout Hertfordshire County, UK, from 1911 onwards through the efforts of a dedicated and visionary midwife, Ethel Margaret Burnside
- Through linking records of births with health in later life by a research team headed by Dr David Barker led to the development of the fetal origins hypothesis termed DOHAD (developmental origins of health and diseases)

Clinical Science is Empirical

- The word **empirical** denotes information gained by means of observation, experience, or experiment.
- **Empirical data** is data that is produced by an experiment or observation.
- As opposed to theoretical that depends on hypotheses

Medical Records

Data Volume & Costs

- ⦿ Around 80 MB of data (4 MB text & 76 MB imaging) per patient is generated
- ⦿ Storage costs < \$ 2.00 per patient for 7 years
 - Dr John Halamka, MD, MS
CIO, Beth Israel Deaconess Medical Center

Nota Bene

- ⦿ Statistics is confusing unless one understands the numbers and what they actually mean making them open to misinterpretation
- ⦿ There is always a chance of over-analysis leading to analysis paralysis
- ⦿ It is important to ask the right questions and re-frame them intelligently

Nota Bene

- ⦿ Running the analytics is all science - mostly mathematics
- ⦿ Interpreting the results is all art derived from knowledge and wisdom
- ⦿ It is possible to predict with a reasonable degree of accuracy (~95%) the most likely outcome under a given set of circumstances

Clinical data is...

- ⦿ Highly multivariate with many important predictors and response variables
- ⦿ Temporally correlated (longitudinal, survival studies)
- ⦿ Costly and difficult to obtain
- ⦿ Historical in nature

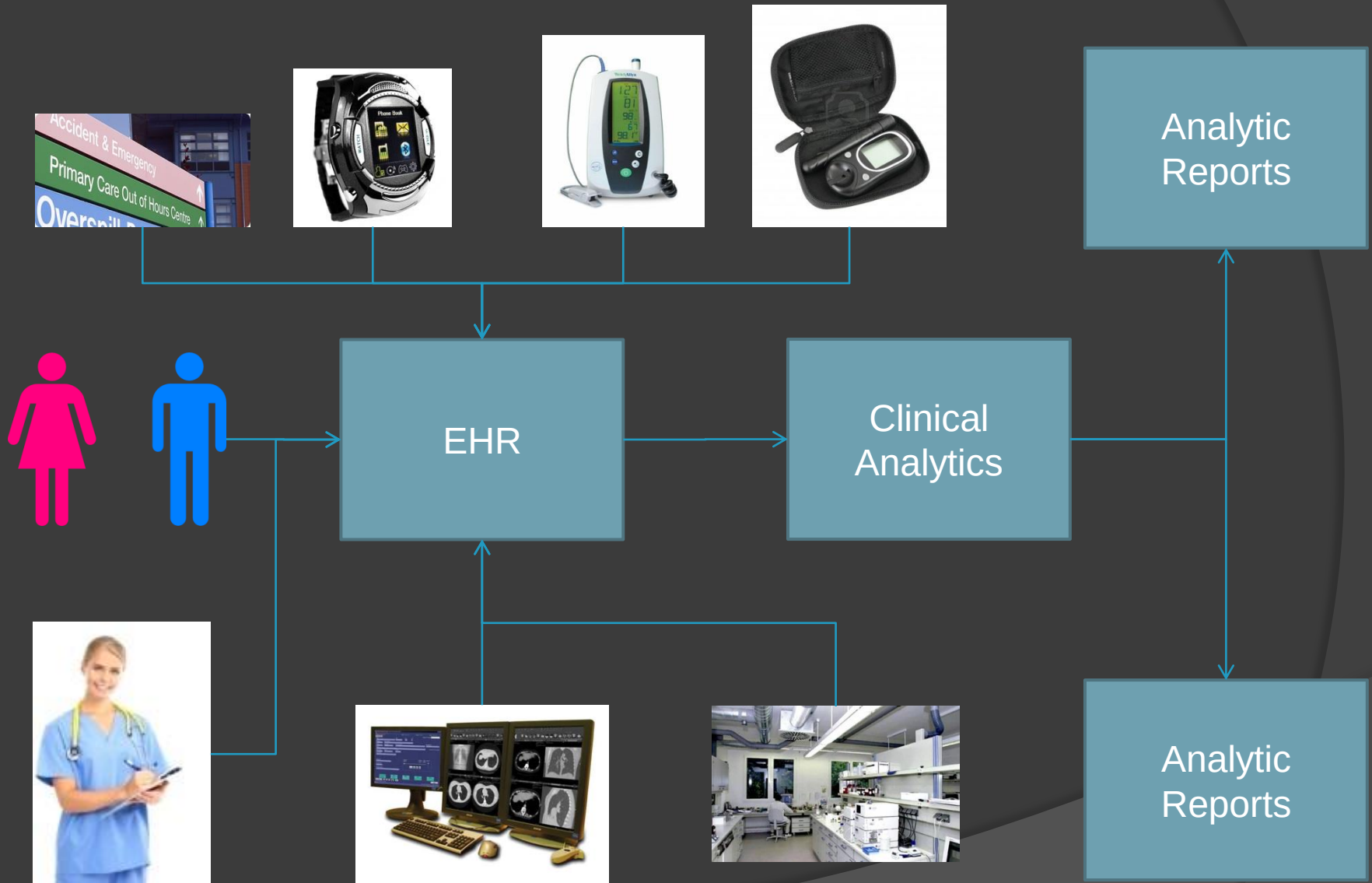
Few Indices

- ◉ Sensitivity
- ◉ Specificity
- ◉ Likelihood Ratio (+/-)
- ◉ Predictive Value (+/-)
- ◉ Prevalence
- ◉ Pre-test/Post-test Odds
- ◉ Post-test Probability (+/-)
- ◉ Kaplan-Meier Survival Curves
- ◉ Total Cost of Treatment
- ◉ Relative Risk
- ◉ Relative Risk Reduction
- ◉ Absolute Risk Reduction
- ◉ Odds Ratio
- ◉ Numbers Needed to Treat (Harm)
- ◉ Quality of Life Year Adjusted
- ◉ Receiver Operator Characteristic (ROC) Curve

Outcomes

- ⦿ Patient better, same or worse
- ⦿ Cost less, same or more
- ⦿ Needs lesser time, same time, longer time to recover/for relief
- ⦿ Needs lesser time, same time, longer time to cure
- ⦿ Cure vs. Recover/Relief

Schematic Process



Data → Analytic Reports

⦿ Data Management: ETL

- Acquire Data
- Clean Data
- Prepare Data (incl. anonymisation)

⦿ Query Preparation

- Formulate Null Hypothesis
- Determine Data Requirements

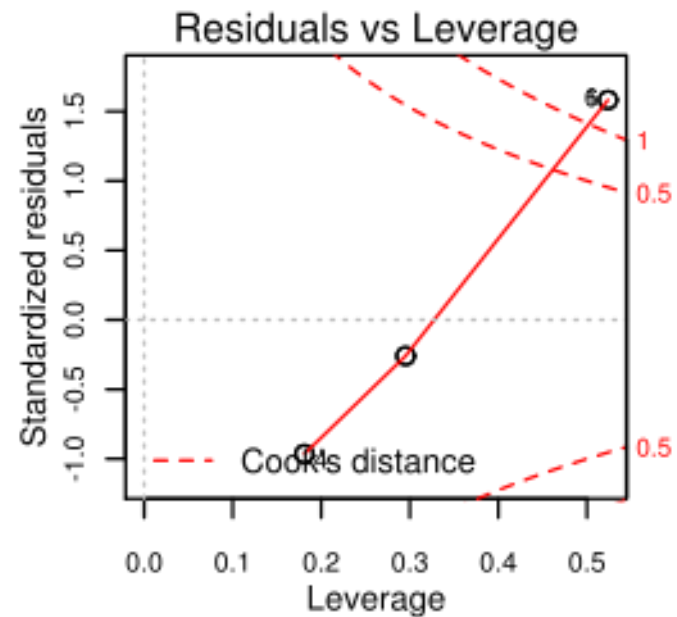
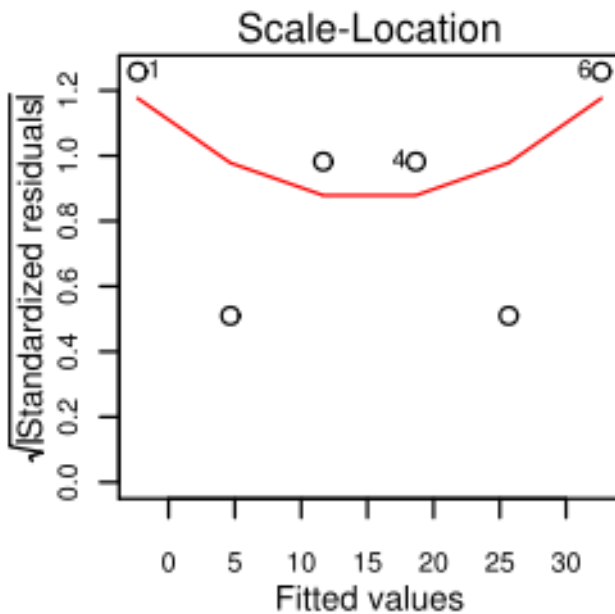
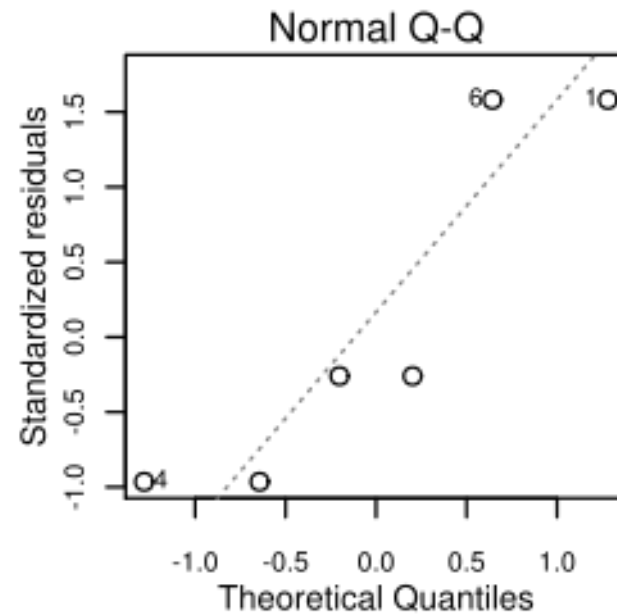
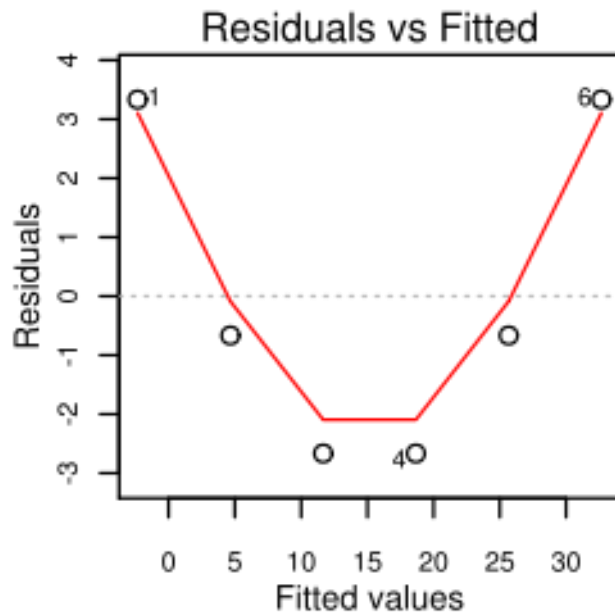
⦿ Analytics Management

- Prepare Analysis
- Run Analysis
- Review Results
- Review Analytical Steps

⦿ Repeat Cycle

- Analytics Management Onwards
- Query Management Onwards
- Data Management Onwards





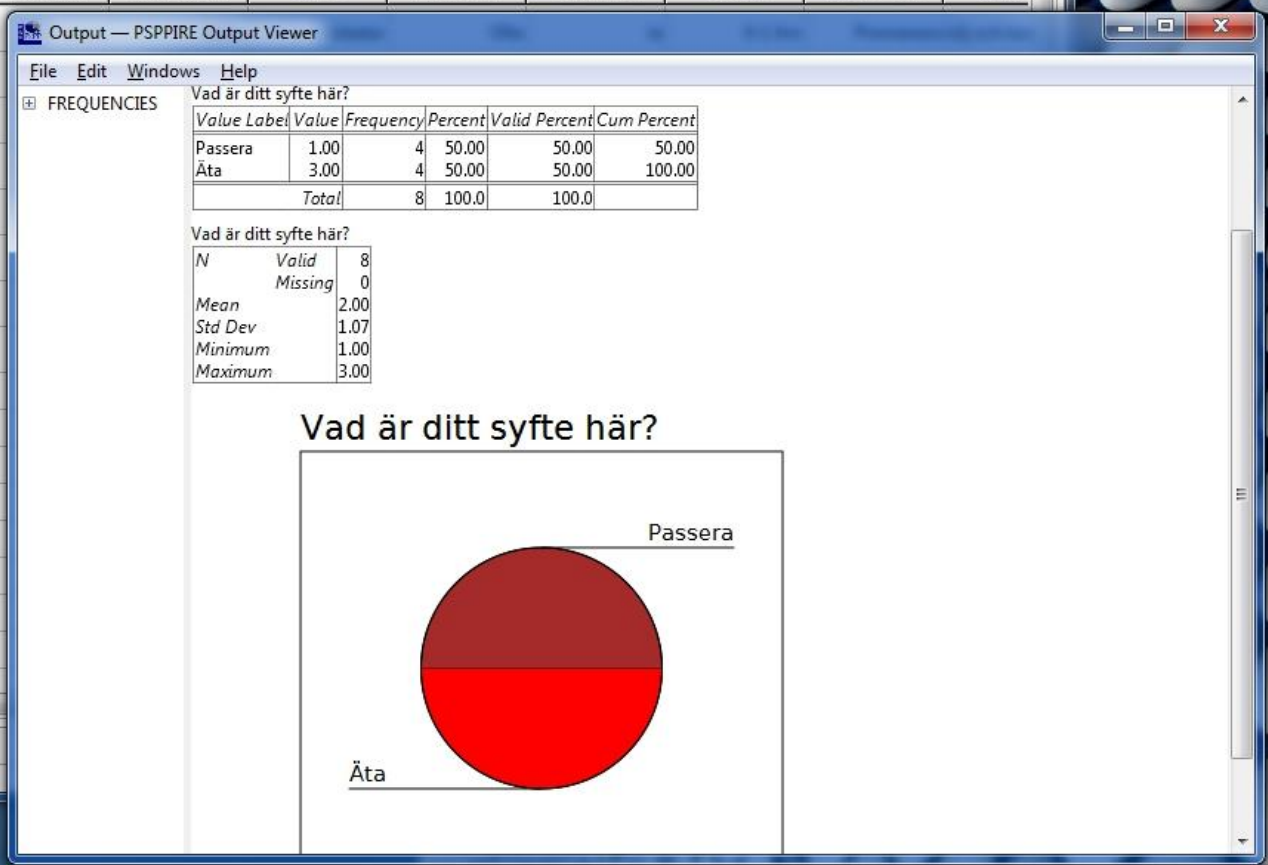
*savfil.sav [DataSet1] — PSPPIRE Data Editor

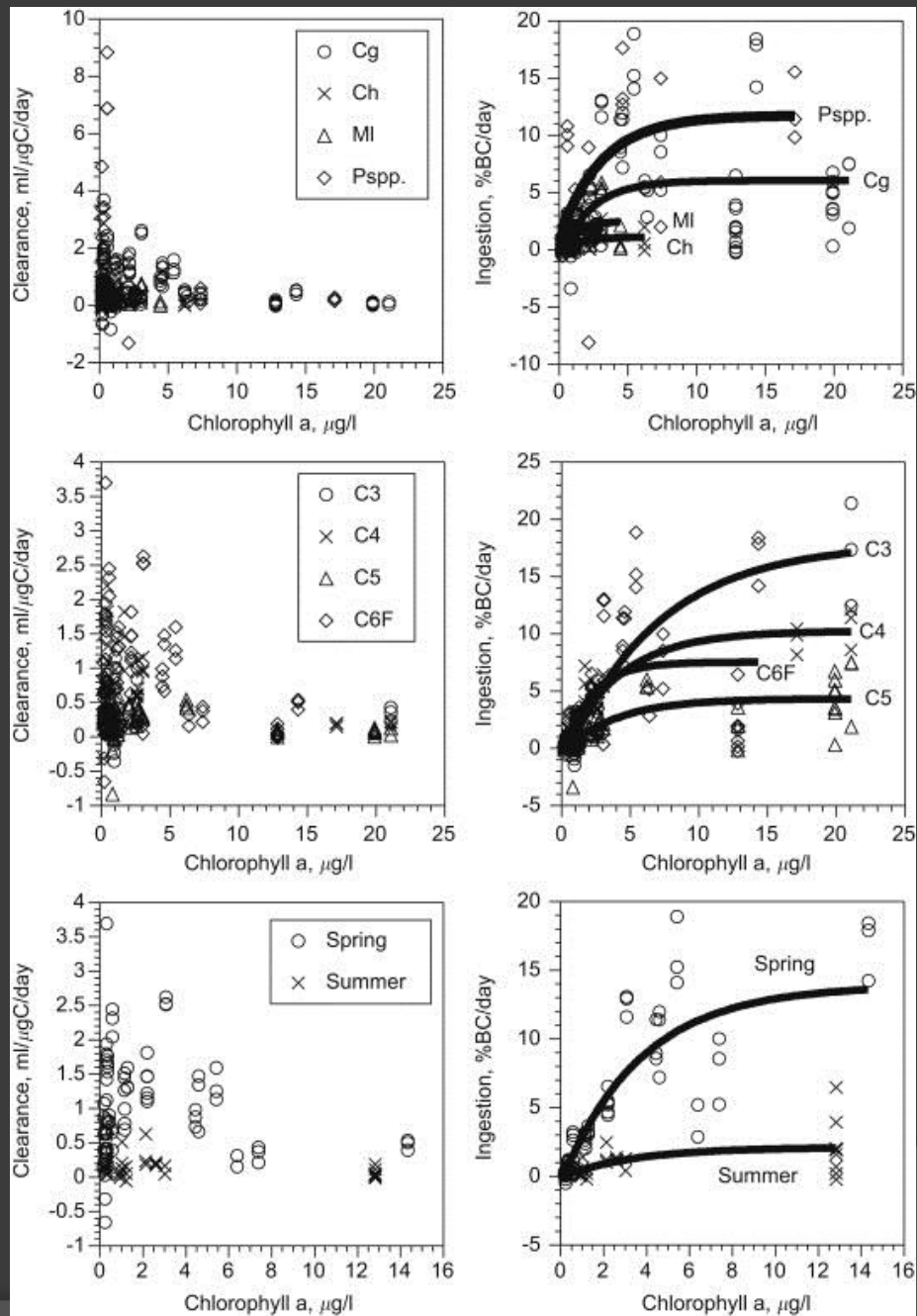
File Edit View Data Transform Analyze Utilities Windows Help

Open... Save Go To Case... Variables... Find... Insert Cases Insert Variable Split File... Weight Cases... Value Labels

	ID	KÖN	ÅLDER	SYFTEHÄR	ANNATHÄR	SYSSEL	HUROFTA	BORДУHÄR	OMINTE	TRANSPORT	KÄNN
1	1	Man	21-30								
2	2	Kvinna	21-30								
3	3	Kvinna	31-45								
4	4	Kvinna	21-30								
5	5	Man	21-30								
6	6	Man	15-20								
7	7	Kvinna	15-20								
8	8	Man	15-20								
9											
10											
11											
12											
13											
14											
15											
16											
17											

Data View Variable View





Chi-Squared [χ^2] Test

- For most clinical data/observation purposes the degree of freedom (*df*) is always 1 and *a* (probability) is always 0.05 (5%), the *a* being the pre-determined (assumed) level of significance.
- The corresponding critical value of Chi square (χ^2) is **3.841** (from Chi square critical values distribution tables). Any c^2 value above this is **significant**.

Likelihood Ratio

- The **Likelihood Ratio** (LR) is a *ratio of likelihoods* (or probabilities) for a condition.
- The LR basically **measures** the **power** of a test to **change** the *pre-test* into the *post-test probability* of a *particular outcome happening*.

Observations on Measures

- Likelihood Ratio is the best indicator.
- Chi square (χ^2) test is the moderate indicator.
- p-Value is the poorest indicator.



Thanks!

A stylized graphic of a folder or document with a central tab. The tab is a light gray rectangle with rounded corners, outlined in white, and contains the word "Thanks!" in a bold, green, sans-serif font. The tab is positioned in the center, with two larger, light gray shapes on either side that resemble the flaps of a folder, also outlined in white. The background is a dark gray gradient with a subtle curved line on the right side.