Statistics and Experimental Design in Toxicology

Georgia Giotopoulou

Comprehensive Pneumology Center

Lung Carcinogenesis Group

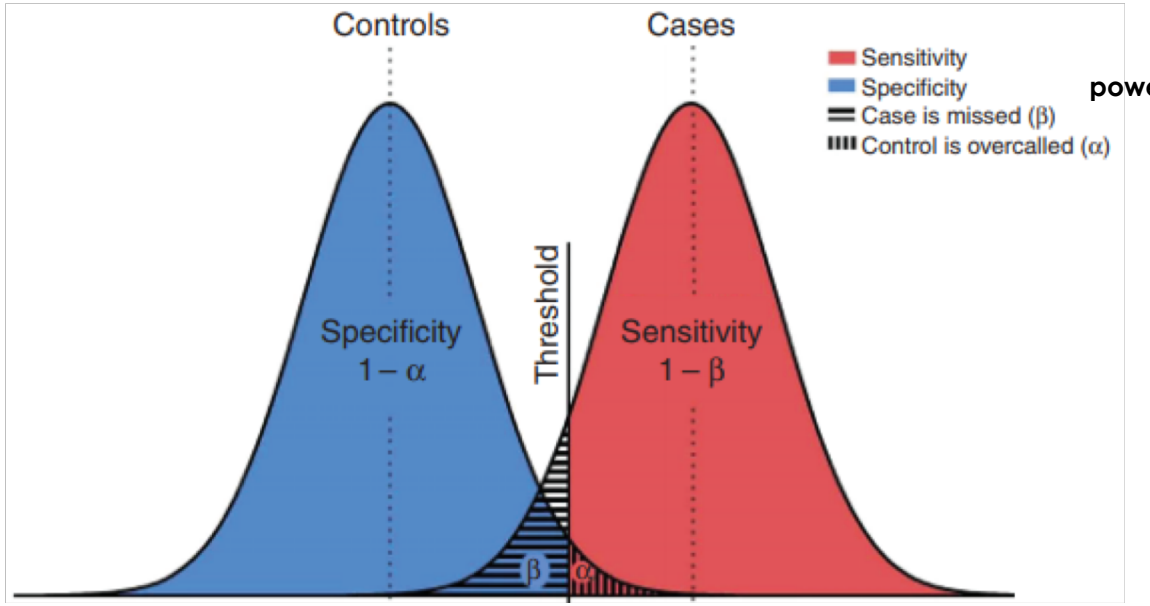
Comprehensive Pneumology Center, Ludwig-Maximilians-University and Helmholtz Zentrum München, Munich, Germany

[U] https://www.helmholtz-muenchen.de/ilbd/research/cpc-experimental-pneumology/stathopoulos-lab/scientific-focus/index.html

Basic terms in Statistics

Toxicology: descriptive science —**statistics**→ mechanistic studies

		Statistical Significance	
		No	Yes
Biological	No	Case I	Case II (false positive)
Significance	Yes	Case III (false negative)	Case IV



power

Causality ≠ correlation

Basic terms in Statistics

datum: each individual piece of experimental information

data: collection of pieces of experimental information

variables: ***independent*****:** treatment variables directly controlled (predictors or explanatory variables)
dependent**:** effect variables dependent on the treatment

population: all the possible measures of a given set of variables

sample: represantative group

Basic terms in Statistics

AIM of Toxicological study: treatment -> effect in a biological system?

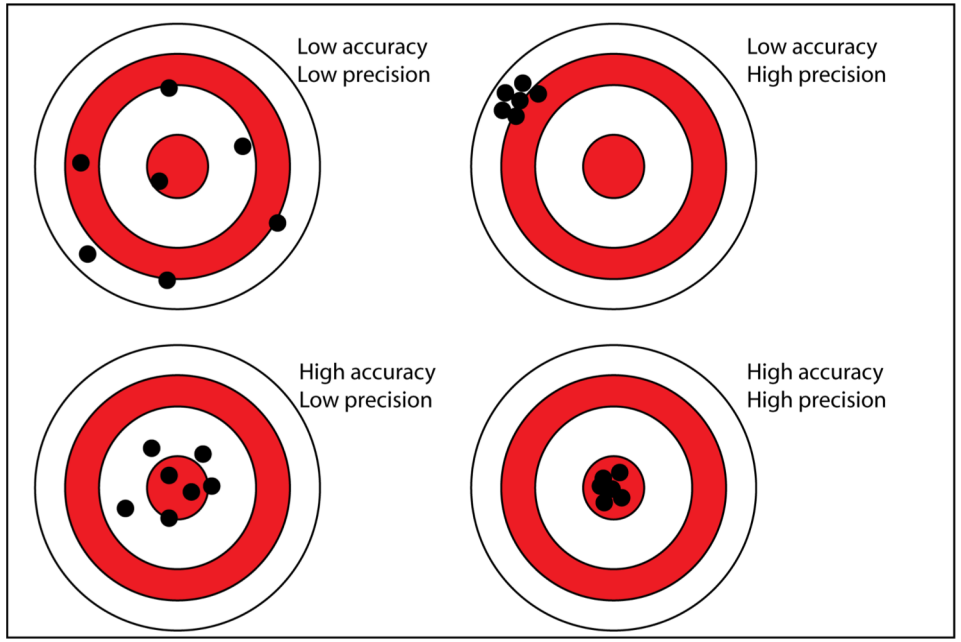
bias: systematic differences other than treatment between the groups (e.g. selection bias)

chance: random differences

signal/noise ratio: ↑ chance of picking up a true effect with confidence

accuracy: quality of measurements of variables/expression of the closeness of a measured value to its true value

precision: reproducibility of a series or repeated measurements



Hypothesis testing and probability (p) values

significance level/*p* value: the probability that a difference has been erroneously declared to be significant, typically 0.05 and 0.01, corresponding to 5% and 1% chance of error

one-tailed *p* value: the probability of getting by chance a treatment effect in a specified direction as great as or greater than that observed

two-tailed *p* value: the probability of getting by chance a treatment difference in either direction that is as great as or greater than that observed

p value does **not** give direct information about the size of any effect

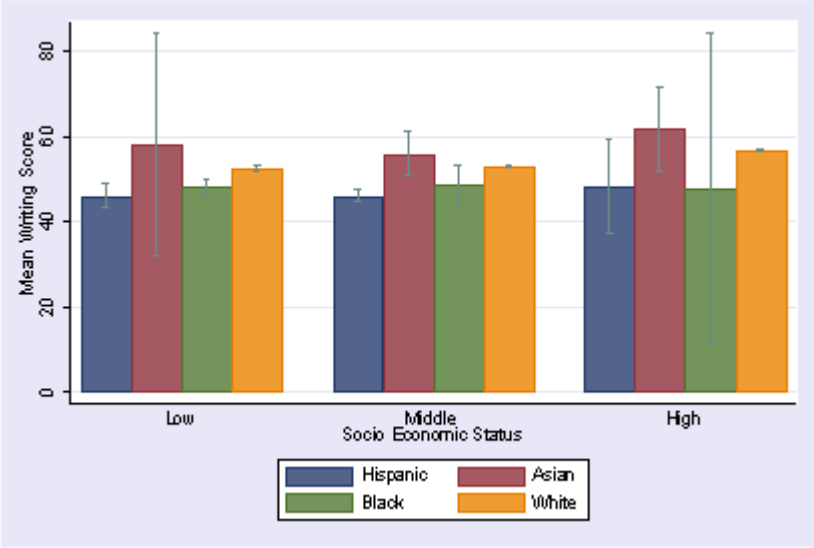
Estimating the size of the effect

Confidence interval (CI) 95%: A range of values (above, below, or above and below) the sample (mean, median, mode, etc.) that has a 95% chance of containing the true value of the population (mean, median, mode); also called the fiducial limit equivalent to $p < 0.05$

Cohen 's D: estimating the size of effect taking into account the standard deviation

$$d = \frac{M_{group1} - M_{group2}}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{(SD^2_{group1} + SD^2_{group2})/2}$$



Effect size	*d*	**Reference**
Very small	0.01	Sawilowsky, 2009
Small	0.20	Cohen, 1988
Medium	0.50	Cohen, 1988
Large	0.80	Cohen, 1988
Very large	1.20	Sawilowsky, 2009
Huge	2.0	Sawilowsky, 2009

Principles for experimental design in toxicologic studies

SAMPLING

- bias / chance
- independence
- replacement
- sample size – power, significance level

EXPERIMENTAL DESIGN

- replication
- randomization
- concurrrent control
- balance

Statistical efficiency
Saving of resources

Minimizing the role of chance

Choice of species and strain: responses of interest rare in untreated control group, evoked by appropriate treatment

Dose level: range of doses → dose-response curve

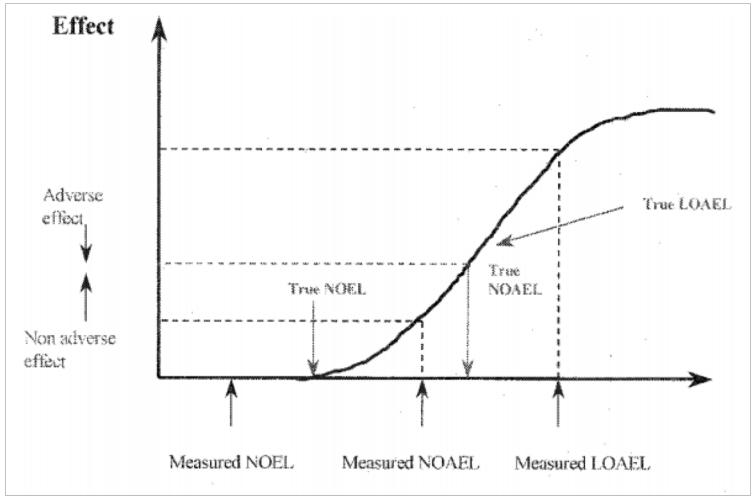
Duration of the experiment: For nonfatal conditions → ideal killing when average prevalence ≈ 50%

Accuracy of determinations

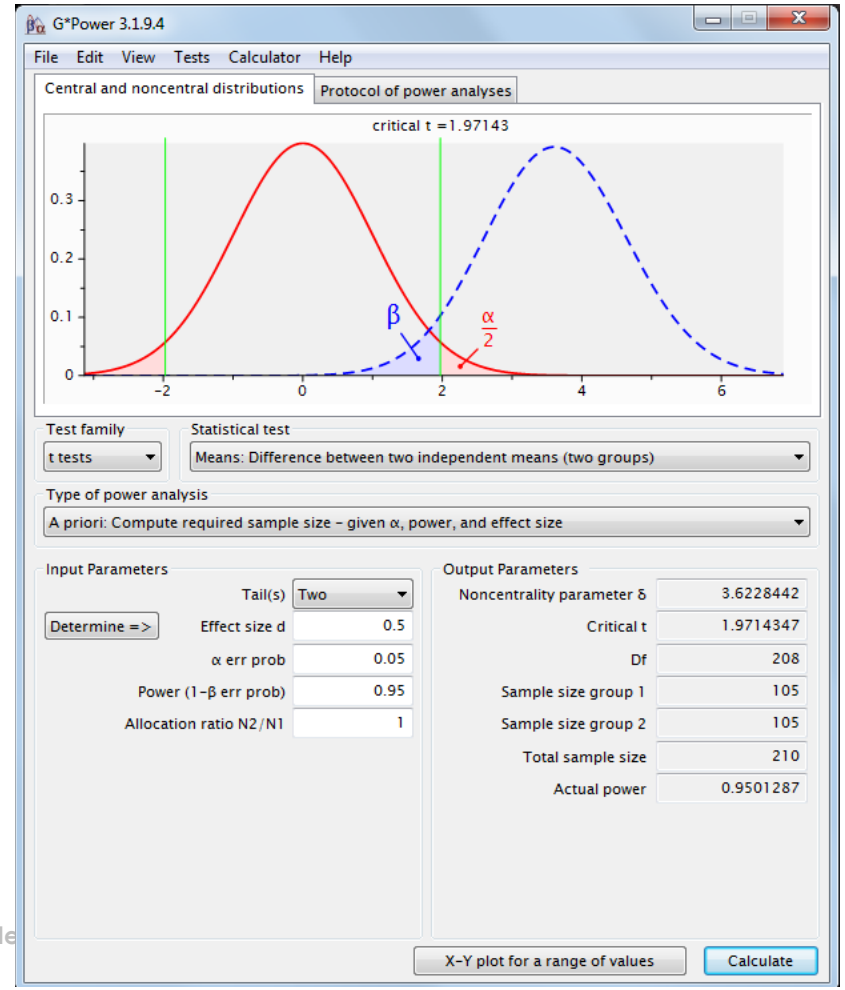
Sampling: without bias, independently, with replacement

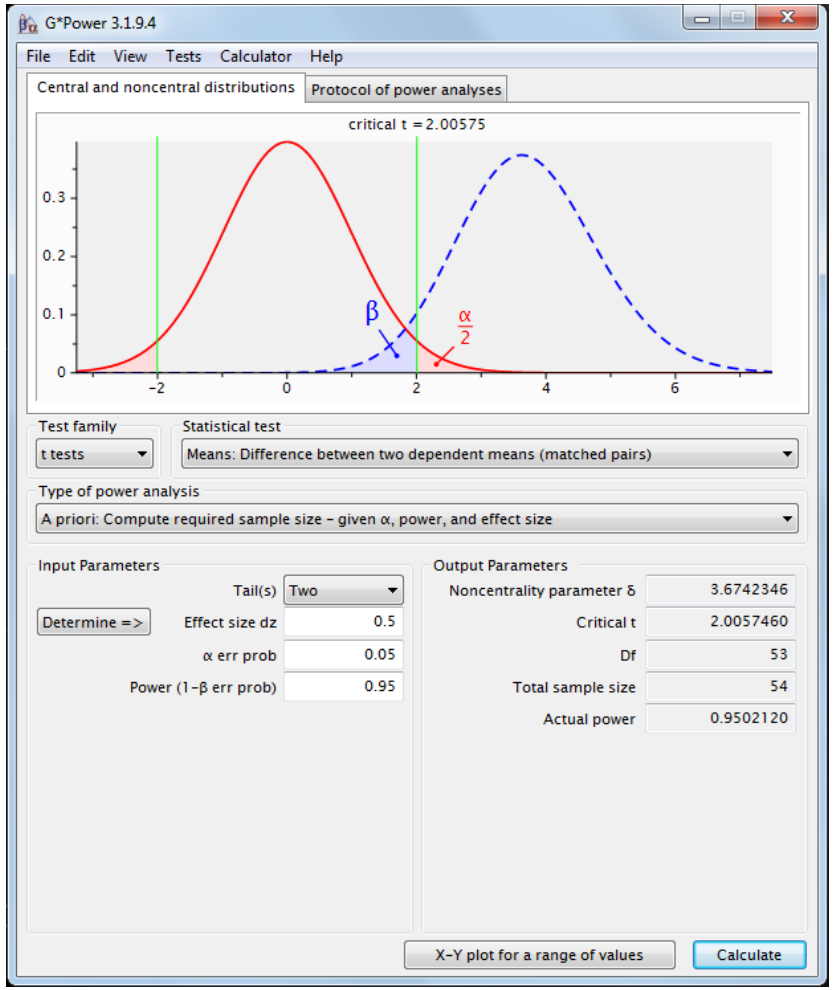
Number of animals: depends on

- The critical difference
- The false-positive rate
- The false-negative rate
- The variability in the material



Experimental design in toxicologic studies- Power analysis





Avoidance of bias

Stratification: homogenous groups, control of confounding variables

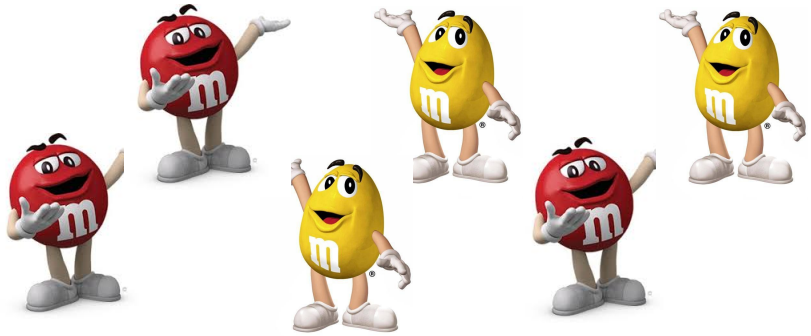
Balance: simultaneous evaluation of the effect of several different factors → *requirement* → the contributions of the different factors can be separately distinguished and estimated

Randomization

Adequacy of control group *(eg same route of administration)*

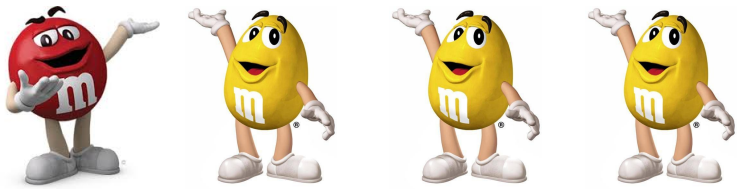
Experimental design – sampling methods

Population

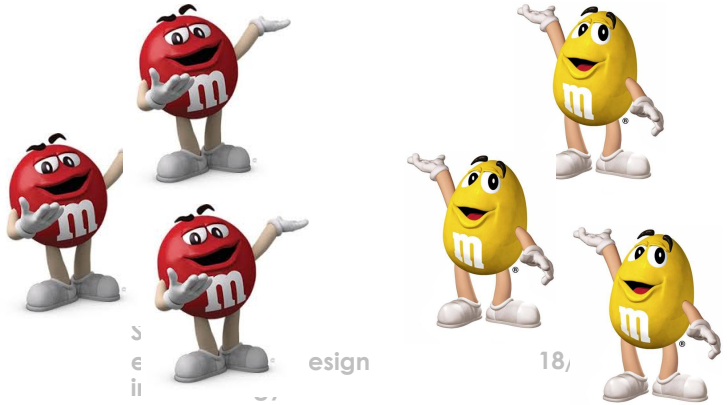


A. Random sampling

Sample size: 4

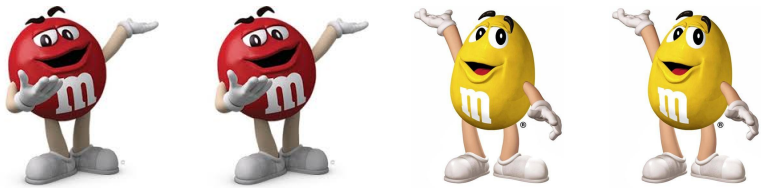


Population



B.Stratified sampling

Sample size: 2+2



C. Systematic sampling
employed in quality control

D. Cluster sampling
pool already divided into separate groups → selection of small ses of groups → selection of a few members from each set

Experimental design types in toxicology

Randomized block:

Blocking: arrangement or sorting of the members of a population into groups based on characteristics that may alter an experimental outcome (genetic background, sex age)

Randomization:each treatment group → at least one member of each blocking groups
Members of each block → assigned in unbiased-random fashion

Latin Square:

	Age (Weeks)			
Source Litter	6–8	8–10	10–12	12–14
1	A	B	C	D
2	B	C	D	A
3	C	D	A	B
4	D	A	B	C

A: control
B: low
C: intermediate
D: high

Experimental design types in toxicology

Factorial design: all levels of a given factor combined with all levels of every other factor in the experiment → ***interaction effect, synergism, antagonism***

1. *No treatment*
2. *Treatment A*
3. *Treatment B*
4. *Treatment A and B*

Nested design (dependent): each subfactor evaluated only within the limits of a single larger factor

Types of variables and frequency distributions

Types of Variables (Data) and Examples of Each Type

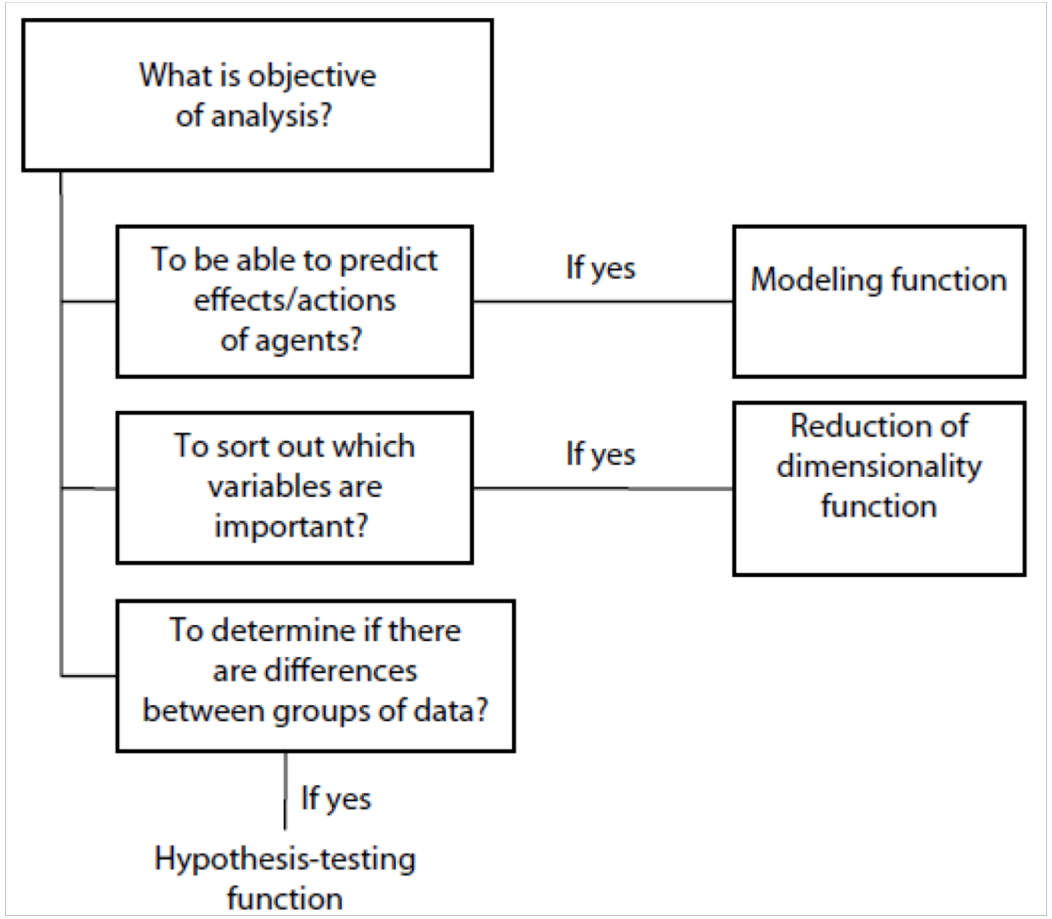
Classified by	Type	Example
Scale	Scalar	Body weight
Continuous	Ranked	Severity of a lesion
Discontinuous	Scalar	Weeks until the first observation of a tumor in a carcinogenicity study
	Ranked	Clinical observations in animals
	Attribute	Eye colors in fruit flies
	Quantal	Dead/alive or present/absent
Frequency distribution	Normal	Body weights
	Bimodal	Some clinical chemistry parameters
	Others	Measures of time to incapacitation

Types of variables and frequency distributions

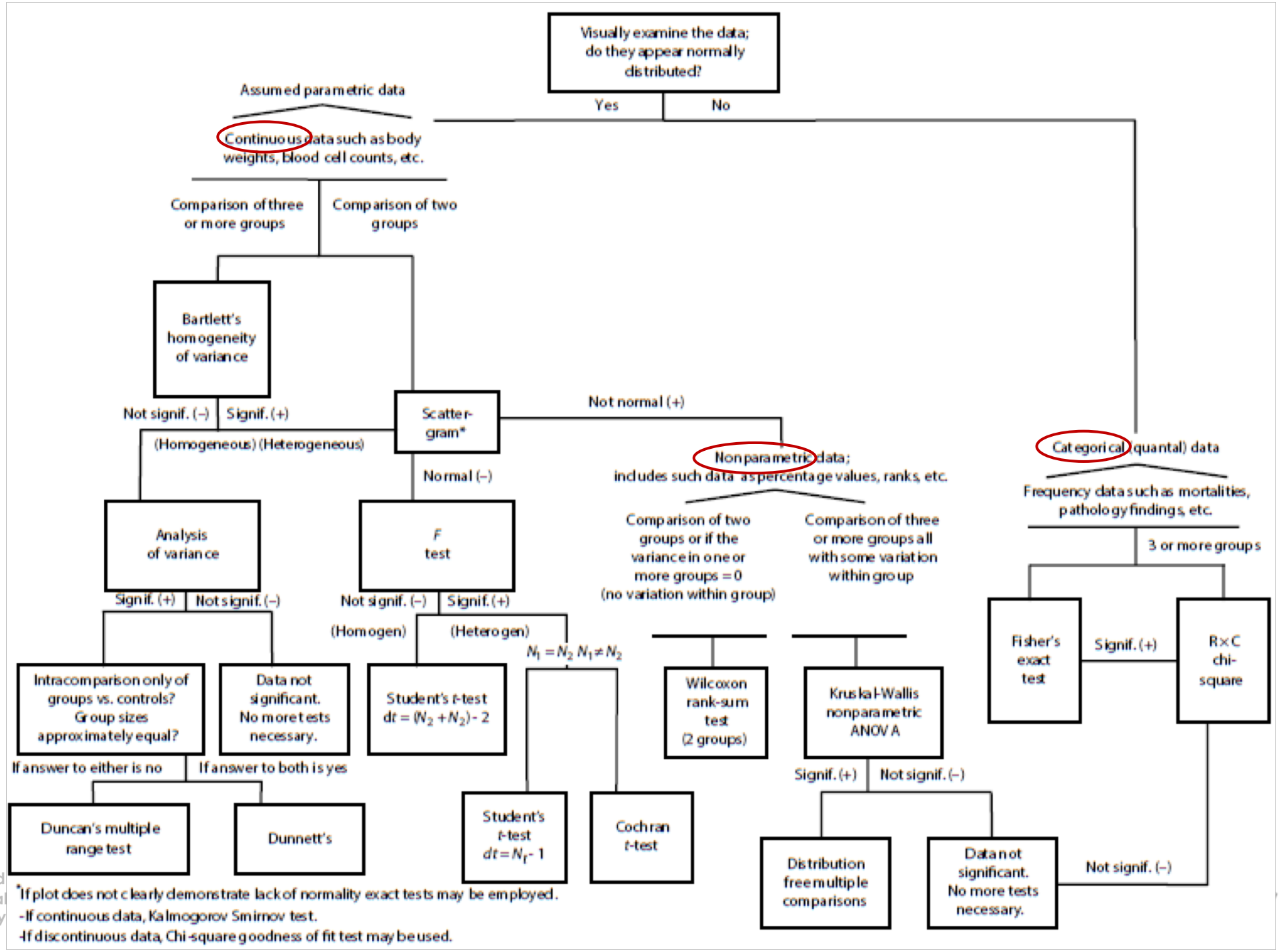
Classification of Data Commonly Encountered in Toxicology

Type of Data	Examples
Continuous normal	Body weights
	Food consumption
	Organ weights: absolute and relative
	Mouse ear swelling test (MEST) measurements
	Pregnancy rates
	Survival rates
	Crown–rump lengths
	Hematology (some)
	Clinical chemistry (some)
Continuous but not normal	Hematology (some; WBC)
	Clinical chemistry (some)
	Urinalysis
Scalar data	Neurobehavioral signs (some)
	PDI scores
	Histopathology (some)
Count data	Resorption sites
	Implantation sites
	Stillborns
	Hematology (some; reticulocyte counts, Howell-Jolly, WBC differentials)
Categorical data	Clinical signs
	Neurobehavioral signs (some)
	Ocular scores
	GP sensitization scores
	Mouse ear swelling test (MEST) sensitization
	Counts
	Fetal abnormalities
	Dose/mortality data
	Sex ratios
	Histopathology data (most)

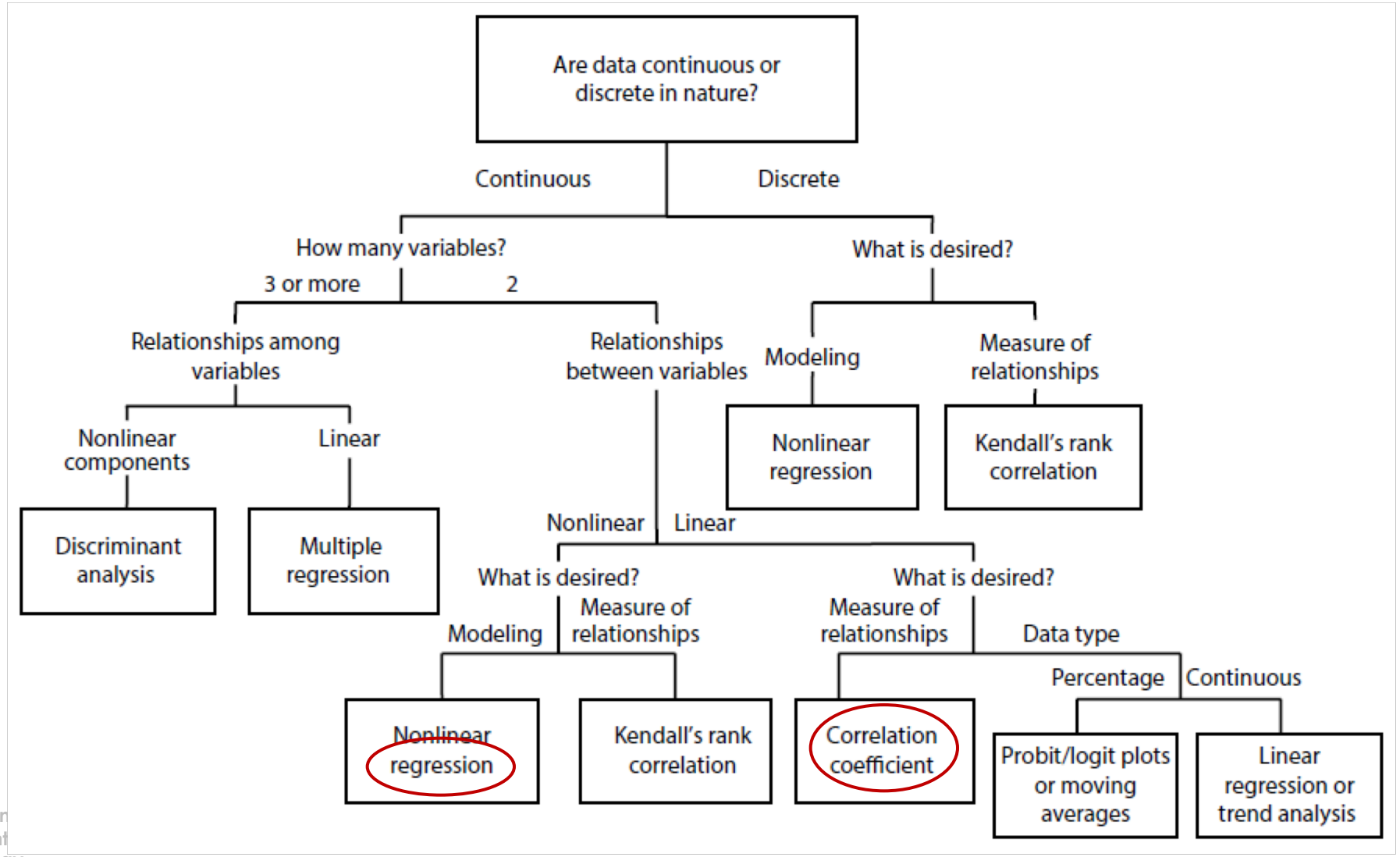
Methodology selection



Methodology selection: Hypothesis-testing



Methodology selection: Modeling



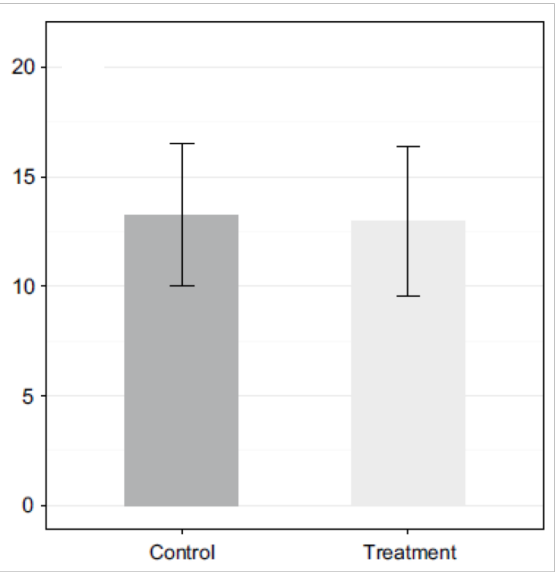
Statistical graphics

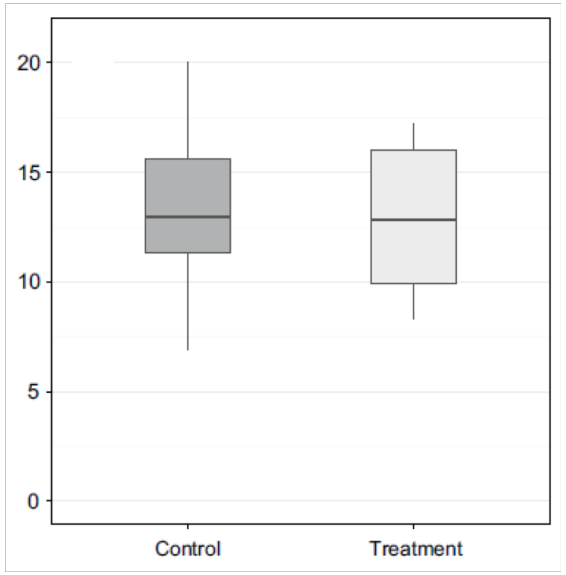
Exploration: summarizing data, deciding on appropriate forms for further analysis

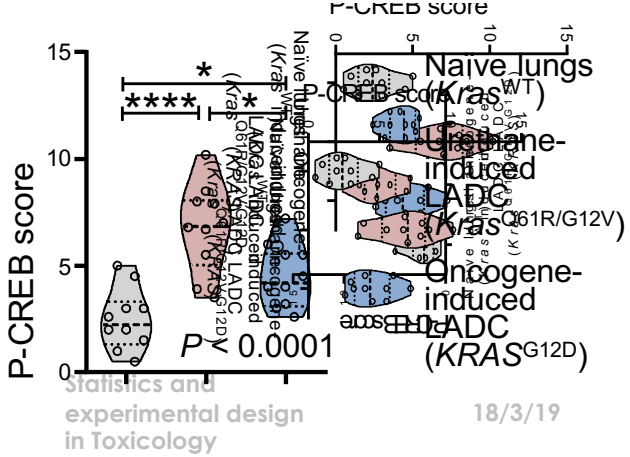
Analysis: use of graphs to evaluate aspects of data, determining outliers

Communication and display of data: showing important trends or relationships in the data

Construction of graph







Forms of Statistical Graphics (by Function)

Exploration

Data Summary	Two Variables	Three or More Variables
Box and whisker plot	Autocorrelation plot	Biplot
Histogram	Cross-correlation plot	Cluster trees
Dot-array diagram	Scatterplot	Labeled scatterplot
Frequency polygon	Sequence plot	Glyphs and metroglyphs
Ogive		Face plots
Stem and leaf diagram		Fourier plots
		Similarity and preference maps
		Multidimensional scaling displays
		Weathervane plot

Analysis

Distribution Assessment	Model Evaluation and Assumption Verification	Decision Making
Probability plot	Average vs. standard deviation	Control chart
Q–Q plot	Component-plus-residual plot	Cusum chart
P–P plot	Partial residual plot	Half-normal plot
Hanging histogram	Residual plots	Ridge trace
Rootagram		Youden plot
Poissonness plot		

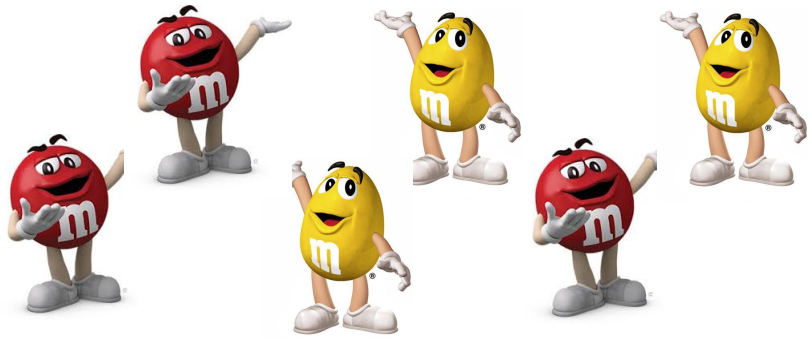
Communication and Display of Data

Quantitative Graphics	Summary of Statistical Analyses	Graphical Aids
Line chart	Means plot	Confidence limits
Pictogram	Sliding reference distribution	Graph paper
Pie chart	Notched box plot	Power curves
Contour plot	Factor space/response	Nomographs
Stereogram	Interaction plot	Sample–size curves
Color map	Contour plot	Trilinear coordinates
Histogram	Predicted response plot	
	Confidence region plot	

Thank you !

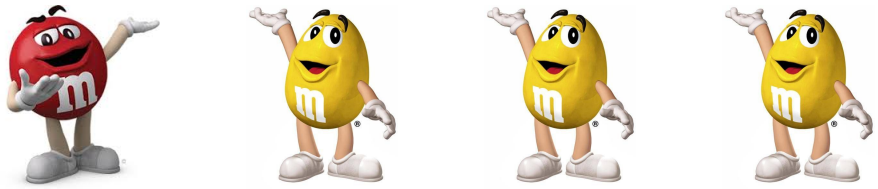
Experimental design – sampling methods

Population

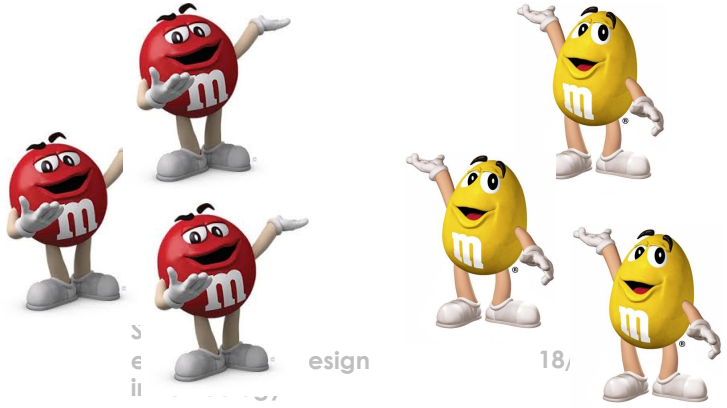


Random sampling

Sample size: 4



Population



Stratified sampling

Sample size: 4

Low-dose extrapolation / NOEL estimation

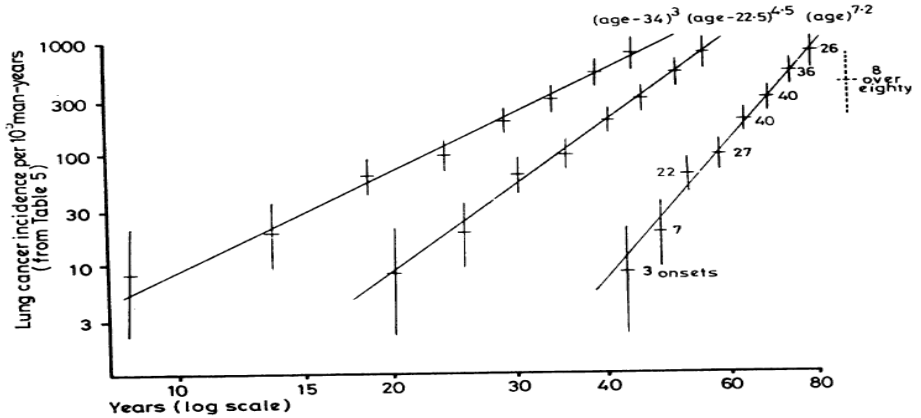


Fig. 2 *Age-specific incidence rates, standardised for dose. The numbers of onsets in each group are given, and 90% confidence intervals are given as vertical lines.*

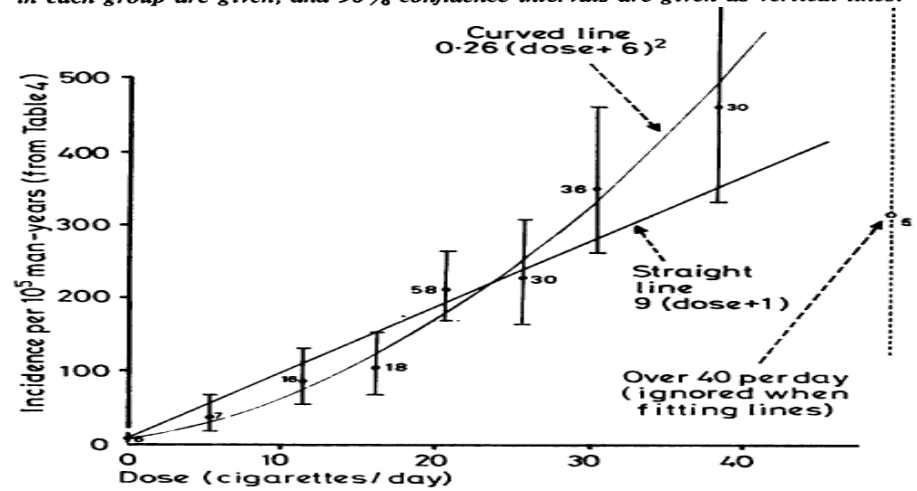


Fig. 1 *Dose-response relationship, standardised for age. The numbers of onsets in each group are given, and 90% confidence intervals are plotted.*

Methodology selection: Hypothesis-testing

Table 1. Parametric and Nonparametric Statistical Methods for Quantitative Data

Content of analysis	Parametric	Nonparametric
Unpaired two-group comparison	Student's *t*-test	Wilcoxon rank-sum test
Paired two-group comparison	Paired *t*-test	Wilcoxon signed rank-sum test
Analysis of homogeneity of population means among multiple groups	One-way ANOVA	Kruskal-Wallis test
Analysis of homogeneity of population means among multiple groups with block	Randomized Block method	Friedman test
Multiple comparisons		
Comparisons with a control group	Dunnett test	Steel test
All pairwise comparisons	Tukey test	Steel-Dwass test
Comparisons with a control group (assuming dose dependency)	Williams test	Shirley-Williams test Shirley test
Dose dependency analysis	Regression analysis	Jonckheere test

Methodology selection: Reduction of dimensionality