Guidelines for medical research

Yossi Levy
May 19, 2020
Rabin Medical Center Beilinson/Hasharon

To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of.

Sir Ronald Fisher

Outline

- Design considerations
  - Sample size and power
- Type of variables and statistical methods
- What to avoid

Design considerations

- Common designs
- Selection of study population
- Selection of outcomes and statistical methods
- Sample size and power

Common designs

- Randomized vs. Observational
- Superiority vs. Equivalence vs. Non-Inferiority
- Controlled vs. Single arm
- Blinded vs. Unblinded
- Pre-Post design
- Adaptive designs

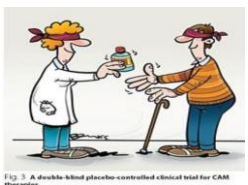


Fig. 3 A double-blind placebo-controlled clinical trial for CAM therapies.

Selection of study population

- Study population should represent the general population
- In medical studies, representativeness is usually achieved by appropriate specification of inclusion and exclusion criteria
- Representative population is essential to ensure external validity

Sample size and power considerations

- Avoid loss of information:
  - Always choose the variable which is measured in the highest measurement scale
  - Avoid dichotomization
- Sample size should be calculated according to:
  - Primary outcome
  - Minimal clinically meaningful effect
  - Reasonable estimate of variation
  - reasonable power, usually 80-90%

- Underpowered study
  - Sample size is too small
  - Higher probability of false negative result
  - Clinically meaningful results may not be statistically significant
  - Unethical
- Overpowered study
  - Sample size is too large
  - Statistically significant results may not be clinically meaningful
  - Waste of resources
  - Unethical

- Qualitative
  - Nominal
  - Ordinal
- Quantitative
  - Continuous
  - Count
  - Survival time

Nominal variables

- The nominal type differentiates between items or subjects based only on their names or (meta-)categories and other qualitative classifications.
- Numbers serve as “tags” or “labels” only, to identify or classify an object.

- Examples
  - Diagnosis
  - Smoking status
  - Treatment: active or placebo
  - Clinical outcome such as survived or died

Analysis of nominal variables

- Descriptives – Frequency tables
- Measures of association: Crammer's V, Phi coefficient, contingency coefficient C, Odds ratio
- Tests of independence/association: Chi-square test, Fisher's exact test
- Comparison of proportions: z-test
- Stratified data: Cochran–Mantel–Haenszel test (CMH)
- Nominal response: logistic regression, multinomial regression, probit regression (results mostly align with logistic regression)
- Advanced: log-linear models

Ordinal variables

- The ordinal type allows for rank order (1st, 2nd, 3rd, etc.) by which data can be sorted, but still does not allow for relative degree of difference between them

- Examples
  - Cancer stage
  - BMI class: Underweight/Normal/Overweight/Obese
  - Socio-Economical Status
  - Chili pepper heat level
  - Class on the Titanic

Analysis of ordinal variables

- Descriptive statistics
  - Location: median, quartiles, percentiles etc.
  - Dispersion: IQR , range, etc.
- Measures of association
  - Spearman correlation coefficient
  - Kendall's Tau
  - Goodman and Kruskal's gamma

Analysis of ordinal response

- Analysis by ranks
  - Kruskal-Wallis test
  - Friedman test
  - Wilcoxon's signed ranks test
  - Jonckheere test
  - Mann-Whitney test
  - Cochran–Armitage test for trend
- Ordinal regression
- Scoring system: replace ordinal categories by numerical scores that will portray the distances between the levels

Quantitative variables

- Quantitative variables are variables in which the numerical values have meaning.
- Two types of quantitative variables: interval and ratio scales
- Special types of quantitative data: count data, time to event

- Examples
  - Body temperature – interval scale
  - HbA1c level – ratio scale
  - Number of relapses within a year– count data
  - Time to heart attack – time to event




Analysis of quantitative variables

- Descriptives: mean, variance, standard errors
- Measures of association:
  - Association between two quantitative variables: Pearson's correlation coefficient
  - Association between a quantitative variable and a qualitative variable: Intraclass correlation
- Quantitative response – linear models
  - Continuous data: linear regression, ANOVA, ANCOVA, t-test
  - Count data: Multinomial regression, Poisson regression, Negative Binomial regression
  - Time to event: Kaplan-Meyer curves and log rank test, Proportional Hazard regression (Cox model). Parametric survival regression (e.g. Weibull model)

What to avoid

- Study design related deficiencies
- Study conduct related deficiencies
- Data analysis related deficiencies
- Data presentation related deficiencies
- Result interpretation related deficiencies

Study design related deficiencies

- Inappropriate power
- Inappropriate choice of endpoints
  - "change from baseline" endpoint is less efficient/powerful than baseline adjustment
- Inappropriate choice of statistical tests
- Non-representative sample (inappropriate inclusion/exclusion criteria)
- Ignoring multiple comparisons
- Unspecified or not detailed enough analyses
- Lack of "fallback positions" in protocol

Study conduct related deficiencies

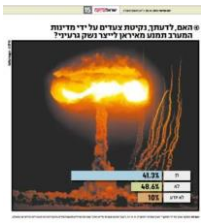
- Unblinding / leak of information
- Failure in randomization
- Missing values/information

Data analysis related deficiencies

- Failure to validate statistical assumptions
- Lack of baseline comparisons and adjustments
- Inappropriate handling of multiplicity issues
- Inappropriate handling of missing values / dropouts
- Post-hoc/data driven analyses
- HARKing / p-hacking

Data presentation related deficiencies

- Failure to present measures of uncertainty/variability
- Use of inappropriate descriptive statistics
- Presentation of confidence interval for groups but not for group differences
- Non-meaningful precision
- Misleading graphs



Result interpretation related deficiencies

- Statistical significance vs. clinically meaningful results
- Misinterpretation of association as causality
- Misinterpretation of p-values
- Misinterpretation of confidence intervals