TransMedRi Workshop in Biostatistics:

Critical evaluation of statistical analysis in scientific paper

Power analysis in research

Mladen Petrovečki

Statistics in health research

contribution. **It is important to note that funding limits will be applied as eligibility criteria. Proposals that do not respect this limit will be considered ineligible.**

- ***Statistics in health research:*** Appropriate study design, data processing and statistical analysis of results are important for the quality and efficiency of the science and reliability of conclusions, and hence also ethically. So, whenever applicable, the proposal should explain them. This may for example include description of experimental plan and data gathering, method for uncertainty or measurement error estimation, statistical analysis of data and methods of inference (e.g. statistical tests and p-values to be used, accounting for multiple comparisons or small sample size, dealing with missing or noisy data), statistical power analysis and estimate (justification) of the number of needed animals or human subjects. If these are not applicable or not justified, the proposal should briefly explain it.
- ***Innovative clinical trials[39] to verify safety and efficacy:*** Specific actions under clinical trials will have a major European added value in translating research into clinical practice, increasing therapeutic options for patients, stimulating the implementation of best

- Proposed priorities for innovative health research 2012

(Working document – not legally binding; Indicative publication date for all documents including the final work programme is 20 July 2011. All related documents will then be accessible via http://cordis.europa.eu/fp7/dc/index.cfm)

Journal article

during the 6 months of follow-up. Data entry was done by administrative assistants, blinded to the assigned treatment.

Statistical analysis

Based on an α of 5% and 80% power, we calculated that 250 patients were needed in each treatment arm to detect a clinically relevant reduction in the incidence of zoster-associated pain 1 month after the onset of the zoster rash from 12% to 5%. Although earlier studies showed an incidence of postherpetic neuralgia of 9–34% when using standard treatment,[1] we used a more conservative incidence to be expected in our control group based on data from a pilot study.[15] Allowing for 10% loss of follow-up, the required number of study patients was 550.

Analyses were carried out with an intention-to-treat approach. The presence of pain (at the different time points) in the two groups was compared by calculating

- Lancet 2006; 367: 219–24

(Albert J M van Wijck, Wim Opstelten, Karel G M Moons, Gerrit A van Essen, Robert J Stolker, Cornelis J Kalkman, Theo J M Verheij. The PINE study of epidural steroids and local anesthetics to prevent postherpetic neuralgia: a randomized controlled trial)

Statistical power

- *...the ability of a test statistic to detect a specified alternative hypothesis or difference of a specified size when alternative hypothesis is true*
  - B. Dawson & R.G. Trapp
  - Basic & clinical biostatistics, 4th ed., Lange Medical Books/McGraw-Hill, New York-Toronto, 2004
- *...is a measure of the likelihood that a researcher will find statistical significance in a sample if the effect exists in the full population*
  - M.L. McHugh
  - Power analysis in research, Biochemia Medica 2008;18(3):263–74
  - http://www.biochemia-medica.com/content/18-3-259-402

Statistical power

- *...ability of a study to detect an actual effect or difference*
  - B. Dawson & R.G. Trapp
  - Basic & clinical biostatistics, 4th ed., Lange Medical Books/McGraw-Hill, New York-Toronto, 2004

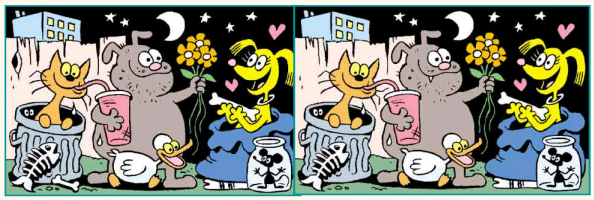
Hypothesis

- Null hypothesis
  - H_0 – no difference
- Alternate hypothesis
  - H_1 – difference exists
- Only one can be truthful
  - Only one can be accepted
  - Other will be rejected

Hypothesis testing

Reality		Conclusion from hypothesis test
No difference (H_0)	Difference (H_1)	
type I error (α error)	correct **power**	**difference exists (reject H_0)**
correct	type II error (β error)	**no difference (accept H_0)**

Hypothesis testing: No difference



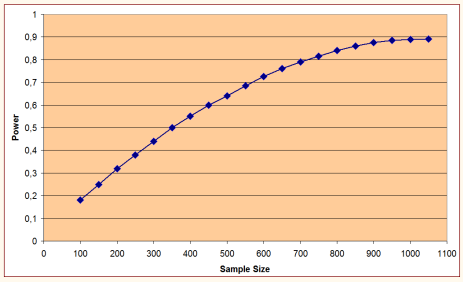
Hypothesis testing vs. diagnostic testing

- Type I error → FP rate
  - significance found with no significance in real
- Type II error → FN rate
  - significance not found for significance in real
- Test power → Sensitivity
  - detecting significance when result is significant

Statistical power

- Value
  - 0 – 1; **≥ 0.8** mostly suggested
- Factors
  - primary
    - size of the effect ← clinical significance
    - variability of observations
    - level of significance ← statistical significance
    - sample size
  - secondary
    - statistical test

Power curve



Factors of statistical power

1. Size of the effect (clinical significance)
2. Variability of observations
3. Level of significance (statistical significance)
4. Sample size
5. Statistical test

Representative sample!

1. Effect size

- Magnitude of the effect, smallest effect of interest
- Smallest size of the difference between groups
  - e.g. I, humans, proportion with *Helicobacter pylori*
    - sample: 65/100 = 65%
    - control: 24/30 = 80%
    - effect size: 15%
  - e.g. II, glucose, human subject
    - sample: 5±1 mmol/L
    - control: 4.5±1 mmol/L
    - effect size: 0.5 mmol/L
- ↑ effect = ↑ power

Effect size

- ↑ effect = ↑ power

Determining the effect

- Prior studies published
- Estimation of minimal effect
  - from pilot study
  - set by definition
  - peer review before starting the study
    - e.g., study protocol evaluation

Effect size of association

- What is the size of effect?
- Measuring: correlation
  - numeric variables → Pearson's r
  - ordinal variables → Spearman's r
  - t-test → point biserial r
  - ANOVA → η^2 (eta-squared)
  - etc.
- Importance: for studies with huge N
- r^2 → variance explained

2. Variability (numerical data)

- ↑ variability = ↓ power

3. Significance level

- Notation, according to testing:
  - before – α
  - after – P
- ↓ α = ↑ power
- $\alpha < 0.05$
  $\alpha < 0.01$
  $\alpha < 0.001$

4. Sample size

- **Must be** planned for the study
- ↑ sample size = ↑ power
  - sample **must** be representative
  - population study → power = 1
- What is appropriate sample size?
  - answer: power analysis
  - variables:
    - level of significance (α)
    - power (←β)
    - effect size (*E*)
    - variability (*v*)

Power analysis

- Sample size estimation
  - N_{sample} ~ α, β & *E*, *v*
- Always **before** beginning of the study
- Sample size
  - the smallest possible (ethics)
  - methodology
    - occasionally
    - general & quick formulae
    - tables
    - nomogram
    - software

Lehr's equation for unpaired t-test with α=0.05 and 80% power

$N = 16/SD^2$

SD – standardized difference

SD = the smallest difference in means still clinically important / assumed equal standard deviation

Power analysis

- Sample size estimation
  - N_{sample} ~ α, β & *E*,
- Always **before** begi
- Sample size
  - the smallest possi
  - methodology
    - occasionally
    - general & quick
    - tables
    - nomogram
    - software

Sample size table for correlation (quick table)

This is the truncated version of the full table (next section), for quick references to sample size required to establish a correlation coefficient (ρ) of a particular value, in the most common situation where power=0.8 and alpha (α)=0.05

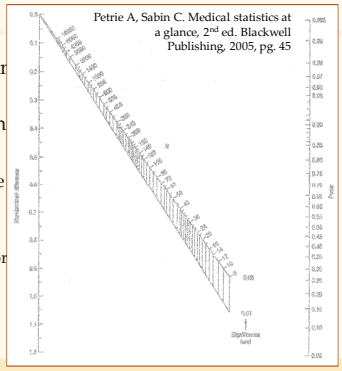
- r = correlation coefficient
- ss = sample size

http://www.stattools.net/

r(ρ)	ss	r(ρ)	ss	r(ρ)	ss	r(ρ)	ss	r(ρ)	ss
0.02	15455	0.22	126	0.42	33	0.62	15	0.82	8
0.04	3862	0.24	106	0.44	30	0.64	14	0.84	7
0.06	1716	0.26	90	0.46	28	0.66	13	0.86	7
0.08	964	0.28	77	0.48	25	0.68	12	0.88	6
0.10	617	0.30	67	0.50	23	0.70	11	0.90	6
0.12	428	0.32	59	0.52	21	0.72	10	0.92	6
0.14	314	0.34	52	0.54	20	0.74	10	0.94	5
0.16	240	0.36	46	0.56	18	0.76	9	0.96	5
0.18	189	0.38	41	0.58	17	0.78	9	0.98	5
0.20	153	0.40	37	0.60	16	0.80	8		

Power analysis

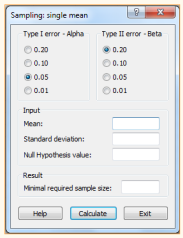
- Sample size estimation
  - N_{sample} ~ α, β & *E*, *v*
- Always **before** beginn
- Sample size
  - the smallest possible
  - methodology
    - occasionally
    - general & quick for
    - tables
    - nomogram
    - software



Petrie A, Sabin C. Medical statistics at a glance, 2nd ed. Blackwell Publishing, 2005, pg. 45

Power analysis

- Sample size estimation
  - N_{sample} ~ α, β & *E*, *v*
- Always **before** beginning of the study
- Sample size
  - the smallest possible (ethics)
  - methodology
    - occasionally
    - general & quick formulae
    - tables
    - nomogram
    - software

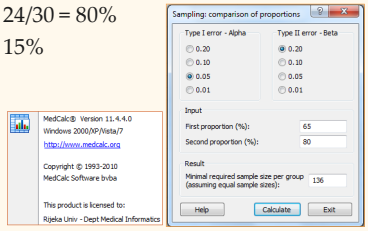


Sample size estimation, examples

1. humans, proportion with *Helicobacter pylori*
   - sample: 65/100 = 65%
   - control: 24/30 = 80%
   - effect size: 15%
2. glucose, human subject
   - sample: 5±1 mmol/L
   - control: 4.5±1 mmol/L
   - effect size: 0.5 mmol/L

Sample size estimation, examples

1. humans, proportion with *Helicobacter pylori*
   - sample: 65/100 = 65%
   - control: 24/30 = 80%
   - effect size: 15%



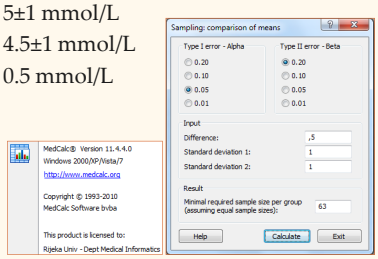
Sample size estimation, examples

1. humans, proportion with *Helicobacter pylori*
   - sample: 65/100 = 65%
   - control: 24/30 = 80%
   - effect size: 15%

- http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html

Sample size estimation, examples

2. glucose, human subject
   - sample: 5±1 mmol/L
   - control: 4.5±1 mmol/L
   - effect size: 0.5 mmol/L



Sample size estimation, examples

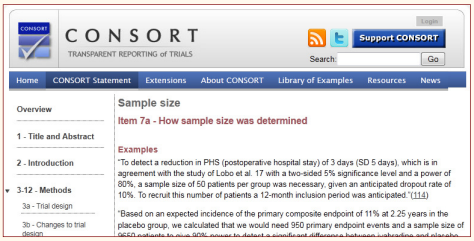
2. glucose, human subject
   - sample: 5±1 mmol/L
   - control: 4.5±1 mmol/L
   - effect size: 0.5 mmol/L

- http://www.stat.ubc.ca/~rollin/stats/ssize/n2.html
- http://www.stat.uiowa.edu/~rlenth/Power/

Sample size adjustment

- losses to follow up
  - drop-out rate, *d*
  - $N_{adjusted} = N \cdot 100/(100-d)$
- independent groups of different sizes
- final-stage critical value to protect the overall type-I error rate

Important!



- http://www.consort-statement.org/home/

Consolidated Standards of Reporting Trials, encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials (RCTs)

5. Statistical test

- Secondary factor
- Each statistical tests ← power level
  - parametric > non-parametric
  - chi-squared test > Fisher exact probability test
  - etc.
- Most important: **appropriate statistical test**
  - nominal/ordinal/numeric variables
  - data distribution
  - sample size
  - etc.

Conclusion

- Power of the study
  1. Size of the effect (clinical significance)
  2. Variability of observations
  3. Level of significance (statistical significance)
  4. Sample size
  5. Statistical test