Statistical Planning and Evaluation of Translational Trials

Frank Konietschke

Institut für Biometrie und Klinische Epidemiologie

Charité - Universitätsmedizin Berlin, Berlin

frank.konietschke@charite.de

Clinical vs. Pre-Clinical

- Both areas come from the same origin
- But...

www.pinterest.com

Clinical vs. Pre-Clinical

- Both areas come from the same origin
- But...

www.pinterest.com

Biometry in Animal Experiment Projects

- **Clinical trials**
  - Ethical approval is legally regulated (§40 + §42 Drug Law)
  - Statisticians are usually members of the committee (state law)
  - Statistical planning and analysis are strictly regulated (e.g., ICH E9)
- **Animal testing**
  - Approval from the local animal welfare committee is required (§8 para. 1 sentence 1 Animal Welfare Act)
  - Committee may provide feedback (§2 para. 4 Animal Welfare - Animal Experimentation Ordinance)
  - **Statisticians** are not required (§42 Animal Welfare - Animal Experimentation Ordinance)
- **Only 20% of all committees in Germany involve biometricians**

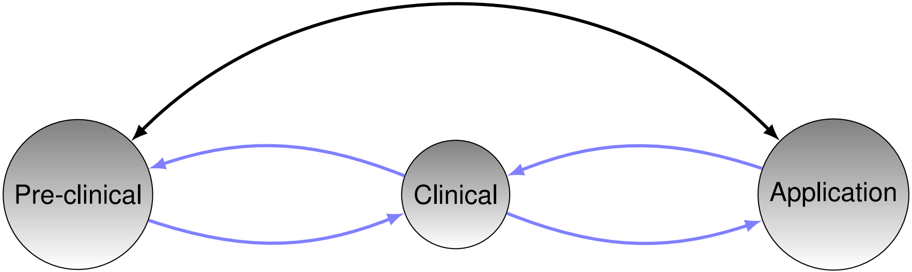
https://www.berlin.de/lageso/gesundheit/veterinaerwesen/tierversuche/tierversuchskommission/

Consequences

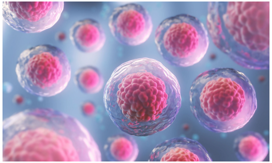
- Experiences show that...
  - Endpoints are not defined
  - Statistical planning is inadequate
  - Statistical analysis is flawed
  - Sample sizes are arbitrarily chosen
  - Text blocks from previous applications are merely copy-pasted
  - ...
- **Uniform standards are important and lacking**

Translational Studies

- Translational research
  - Implementation of results
  - Transition between clinical phases



Example I: Proof DESFERAL



- Osteoinduction (bone formation)
- **6 measurements** per subject

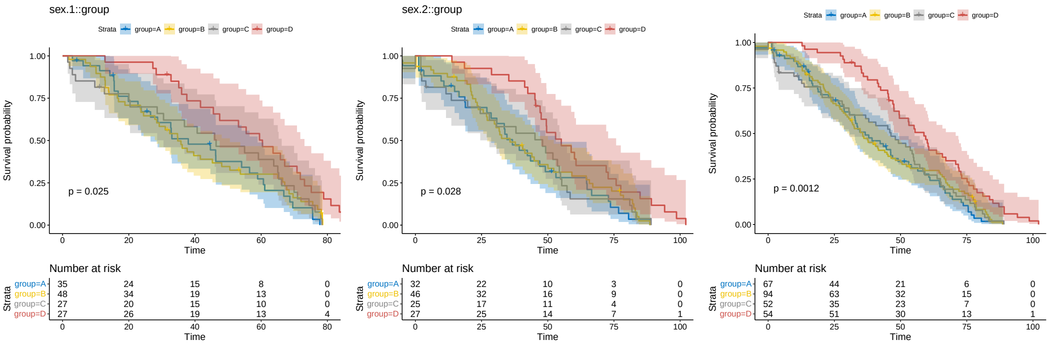
- Bone volume
- **6 groups**
- Show a 20% difference in at least one dose

- Bone volume
- **2 groups**

- How many donors, mice, and sheep are required?
- Statistical analysis

Example II: SOXALS

- Endpoint: Survival
- Oxprenolol, Placebo, Riluzole, and Edaravone + Gender
- Simulated data:



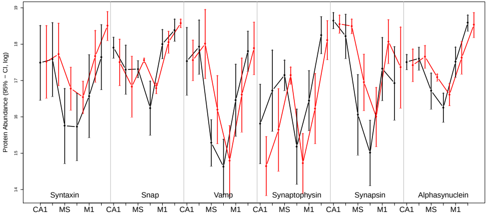
- Objective: $\theta_1 = 0.45; \theta_2 = 0.48; \theta_3 = 0.56$
- How many mice are needed? Statistical analysis?

Example III: PINPRICS

- **PINPRICS**
- Nerve damage as AE in cancer
- Endpoint: SNAP
- sensory nerve action potential amplitude
- Nilotinib, Lithium carbonate, IL-6 neutralizing antibodies, negative control
- **4 groups**
- Objective: Increase of at least 4 μV in at least one comparison
- How many animals are needed? Statistical analysis?

Example IV: Alzheimer

- Alzheimer research
- Wild-type ($n_1 = 10$) and L1 Tau transgenic mice ($n_2 = 9$)
- Response: protein abundance of **six different proteins**
- Measured in **six different brain regions** (each mouse)



Statistical Inference

- Statistical model $X_i \sim F(\theta_i), ; i = 1, \ldots, a$
- θ_i: Parameter of group i (or time point)
  - Mean of group i: $\mu_i = E(X_i)$
  - Mann-Whitney effect $p_i = \int G dF_i$
  - Survival (probability, hazard)
  - ...
- **Classical** statistical procedures

$$H_0 : \theta_1 = \ldots = \theta_a \qquad \text{vs} \qquad H_1 : \theta_i \neq \theta_j$$

- ANOVA: Equality of means
- Kruskal-Wallis: Equality of distributions
- Is this the main question of the study?
- With a significant result: **Some group** is different
- How to plan the sample size?

Local Inference

- The global hypothesis is **not** the main question of the study
- Of interest are **local hypotheses**

$$H_0^{(\ell)} : \mathbf{c}_\ell' \boldsymbol{\theta} = 0, ; \ell = 1, \ldots, q$$

  - Multiple comparisons with control: $H_0^{(\ell)} : \theta_1 = \theta_j$
  - Pairwise comparisons: $H_0^{(\ell)} : \theta_i = \theta_j$
  - Reject H_0 globally if any local comparison is rejected

Local Inference / Maximum Tests

- Test for each local null hypothesis (t-test)
- Reject the global null hypothesis if any local hypothesis is rejected
- **Account for the correlations** of the test statistics

- Test statistics

$$
\begin{aligned}
T_\ell &= \frac{\mathbf{c}'_\ell(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\widehat{\sigma}_\ell} \\
T_0 &= \max(|T_\ell|)
\end{aligned}
$$

- Required: Distribution of T_0: $\Psi(t_0)$
- Test decisions
  - Local p-values: $1 - \Psi(T_\ell)$
  - Global p-value: $1 - \Psi(T_0)$
- Calculation of $\Psi(t_0)$: **not trivial**

Sample Size Planning

- Sample size planning is possible for individual comparisons

$$
\begin{aligned}
Power(\boldsymbol{\delta}|N) &= P_{H_1}(T_0 \geq z_{1-\alpha}(\boldsymbol{R})) \\
&= 1 - P_{H_1}(z_{1-\alpha}(\boldsymbol{R}) \leq T_1 \leq z_{1-\alpha}(\boldsymbol{R}), \ldots, -z_{1-\alpha}(\boldsymbol{R}) \leq T_q \leq z_{1-\alpha}(\boldsymbol{R})) \\
&= 1 - \Psi(-z_{1-\alpha}(\boldsymbol{R})\mathbf{1}_q, z1 - \alpha(\boldsymbol{R})\mathbf{1}_q),
\end{aligned}
$$

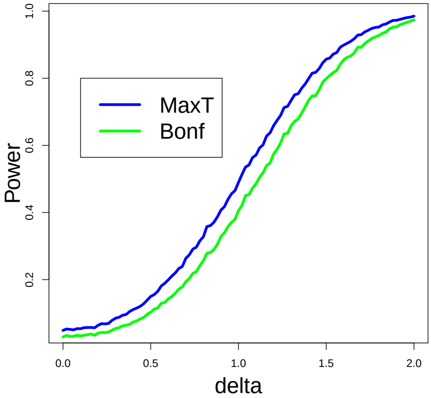
- Sample size N is the solution to

$$f(N|\alpha, \beta, \boldsymbol{\delta}, \mathbf{t}) = Power(\boldsymbol{\delta}_N) - (1 - \beta)$$

- $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)'$ represents the effect size for each individual comparison
- For example, when comparing to a control group: $\boldsymbol{\delta} = \left(\sqrt{N}\frac{(\theta_1-\theta_2)}{\sigma_1}, \ldots, \sqrt{N}\frac{(\theta_1-\theta_a)}{\sigma_q}\right)'$
- $https://s-quest.bihealth.org/SampleSize4MCTP/$

Power Comparisons

- Power comparison: ANOVA
  - Depending on the alternative
  - In many cases, the power is roughly the same
- Power comparison: Bonferroni
  - Higher power
  - Depending on the effect: T_0 up to 10% more power
  - Requires significantly fewer animals



Examples

- **Example I**
- 6 doses
- Improvement of 20% in at least one dose
- $\mu_1 = \ldots = \mu_5 = 1; \mu_6 = 1.2$
- Conservative assumption: $\sigma = 0.3$
- Cohen's $d = 0.67$ in at least one comparison
- $\Rightarrow N = 120$ ($\alpha = 5\%$, Power=80%)
- Bonferroni: $N = 156$ (23% saved)

- **Example II**
- Comparison of Oxprenolol to Placebo, Riluzole, and Edaravone
- $\theta_1 = 0.45$, $\theta_2 = 0.48$, $\theta_3 = 0.56$
- $\Rightarrow N = 440$
- **Example III**
- Pairwise comparisons: Difference of 4 μV
- Standard deviation: 4
- Cohen's $d = 1$ in at least one component
- $\Rightarrow N = 60$ ($\alpha = 5\%$, Power=80%)
- Bonferroni: $N = 72$

Example: Analysis

- **Alzheimer Experiment**
- RM MANOVA: $p < 0.0001$ (there is a difference somewhere)
- Maximum Test:

$\ell =$	$\widehat{\delta\ell}$	T_ℓ	$95\% L$	$95\% U$	p-Value
1	1.35	2.50	-0.68	3.38	0.384
6	1.62	3.23	-0.26	3.51	0.289
11	1.10	6.58	0.47	1.73	$<$ .0001
15	-6.67	-8.45	-9.63	-3.71	$<$ .0001
16	-3.98	-17.27	-4.84	-3.11	$<$ .0001
17	1.46	3.34	-0.18	3.10	0.091
18	2.83	9.47	1.71	3.95	$<$ .0001
32	1.07	4.59	0.20	1.94	$<$ .0001
33	-2.18	-6.77	-3.40	-0.97	$<$ .0001
34	-3.10	-11.54	-4.11	-2.09	$<$ .0001
35	0.85	3.59	-0.04	1.75	0.053
36	2.91	21.80	2.41	3.41	$<$ .0001
		$T_0 = 21.80$			

Software

- R packages
  - nparcomp
  - rankFD
  - nparLD
  - Over 200,000 installations

References

- Brunner, E., Bathke, A. C., Konietschke, F. (2018). *Rank and pseudo-rank procedures for independent observations in factorial designs*. Springer International Publishing.
- Konietschke, F., Hothorn, L. A., Brunner, E. (2012). Rank-based multiple test procedures and simultaneous confidence intervals. *Electronic Journal of Statistics*, **6**, 738–759.
- Konietschke, F., Schwab, K., Pauly, M. (2021). Small sample sizes: A big data problem in high-dimensional data analysis. *Statistical Methods in Medical Research*, **30(3)**, 687–701.
- Rubarth, K., Sattler, P., Zimmermann, H. G., Konietschke, F. (2021). Estimation and testing of Wilcoxon-Mann-Whitney effects in factorial clustered data designs. *Symmetry*, **14(2)**, 244–264.
- Umlauft, M., Placzek, M., Konietschke, F., Pauly, M. (2019). Wild bootstrapping rank-based procedures: Multiple testing in nonparametric factorial repeated measures designs. *Journal of Multivariate Analysis*, **171**, 176–192.