Statistik III

Advanced Statistics

Florian Wickelmaier

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Summer 2026

Clustered Data

Hierarchical structure

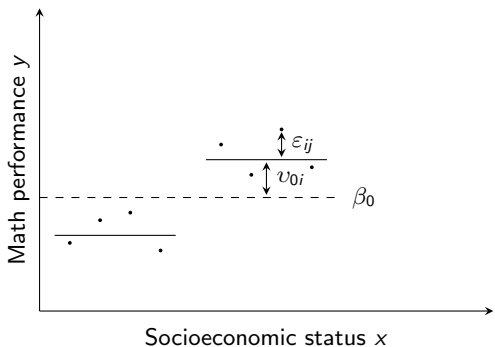
- Observations often do not come from a simple random sample, but result from a hierarchical structure
  - Individuals are organized in groups (e.g., students nested in classes, or schools)
  - Persons are observed multiple times (observations nested in persons, longitudinal data)
- Statistical models for clustered data are called multilevel models, mixed-effects models, random-effects models, covariance components models, or hierarchical models

Example: Mathematics Achievement Study

- Data from the National Center for Education Statistics' (NCES) "High School & Beyond" national survey of U. S. public and Catholic high schools (cf. Bryk & Raudenbush, 2002; Singer, 1998, 2003)
- The data contain information on student performance in a mathematics test and on their socioeconomic status for 7185 students from 160 schools
- Hierarchical data structure
  - Students are organized in schools
    y_{ij} Mathematics achievement of student j in school i
    x_{ij} (relative) socioeconomic status of student j in school i (overall mean 0, centered)
  - Two levels
    Level 1: Student attributes
    Level 2: School attributes

Models with Random School Effects

- Research questions
  - What is the mean math achievement of the students?
  - How much do schools vary in their mean math achievement?



Models with Random School Effects

- Model equations

$$\begin{aligned} \text{(Level 1)} \quad & y_{ij} = b_{0i} + \varepsilon_{ij} \\ \text{(Level 2)} \quad & b_{0i} = \beta_0 + \upsilon_{0i} \\ \text{(2) in (1)} \quad & y_{ij} = \beta_0 + \upsilon_{0i} + \varepsilon_{ij} \end{aligned}$$

  where $\upsilon_{0i} \sim N(0, \sigma_\upsilon^2)$ independent, $\varepsilon_{ij} \sim N(0, \sigma^2)$ independent, and υ_{0i} and ε_{ij} independent

- Interpretation

β_0	mean math achievement
υ_{0i}	random school effect, mean deviation of school i from mean math achievement
ε_{ij}	residual of student j in school i

Models with Random School Effects

Random effects

- The grouping of observations in schools induces dependencies
- These dependencies are modeled by introducing school effects
- For many schools this calls for (too) many parameters
- School effects are therefore modeled as random effects (random variables) υ_{0i}
- Only their variance σ_υ^2 has to be estimated in the model
- The total variance of y_{ij} then is decomposed into the variance between schools σ_υ^2 and within schools σ^2

Models with Random School Effects

Results

- The above-posed research questions can be answered based on the parameter estimates $\hat{\beta}_0$, $\hat{\sigma}_v^2$, and $\hat{\sigma}^2$
  - The estimated mean math achievement of the students is $\hat{\beta}_0$
  - The estimated standard deviation of schools in mean math achievement is $\hat{\sigma}_v$
  - The proportion of the total variance accounted for by the variance between schools is

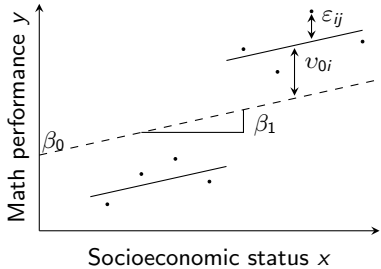
$$\text{ICC} = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

(intra-class correlation)

Models with Random School Effects

Effects of covariates

- Research questions
  - How strong is the relationship between students' socioeconomic status and their math achievement on average?
  - How much do schools vary in mean math achievement for students with average socioeconomic status?



Models with Random School Effects

- Regression with random school effects

$$\begin{aligned} \text{(Level 1)} \quad & y_{ij} = b_{0i} + b_{1i}\, x_{ij} + \varepsilon_{ij} \\ \text{(Level 2)} \quad & b_{0i} = \beta_0 + \upsilon_{0i} \\ & b_{1i} = \beta_1 \\ \text{(2) in (1)} \quad & y_{ij} = \beta_0 + \beta_1\, x_{ij} + \upsilon_{0i} + \varepsilon_{ij} \end{aligned}$$

where $\upsilon_{0i} \sim N(0, \sigma_\upsilon^2)$ independent, $\varepsilon_{ij} \sim N(0, \sigma^2)$ independent, υ_{0i} and ε_{ij} independent

- Interpretation

β_0	mean math achievement for $x_{ij} = 0$ (intercept)
β_1	mean change in math achievement for unit change in socio-economic status
υ_{0i}	school-specific deviation from the mean intercept
ε_{ij}	residual of student j in school i

Mixed-Effects Models

Possible extensions of the models used so far

- Random effects
  - School specific slopes
  - Correlation between intercept and slope
- Fixed effects
  - at school level: type of school (public vs. catholic)
  - at the student level: gender, intelligence
- Grouping, levels
  - Classes, school districts or states (hierarchical)
  - Teacher effect within schools (crossed)

References

Bryk, A.S., & Raudenbush, S.W. (2002). Hierarchical linear models. Thousand Oaks, CA: Sage Publications.

Faraway, J.J. (2006). Extending linear models with R: Generalized linear, mixed effects and nonparametric regression models. Boca Raton, FL: Chapman & Hall/CRC.

Hedeker, D.R., & Gibbons, R.D. (2006). Longitudinal data analysis. Hoboken, NJ: Wiley.

Pinheiro, J.C., & Bates, D.M. (2000). Mixed-effects models in S and S-Plus. New York: Springer.

Singer, J.D., & Willett, J.B. (2003). Applied longitudinal data analysis: Modeling change and event occurrence. New York: Oxford University Press.

Snijders, T.A.B., & Bosker, R. (1999). Multilevel analysis: An introduction to basic and advanced modeling. London: Sage Publications.

Verbeke, G., & Molenberghs, G. (2000). Linear mixed models for longitudinal data. New York: Springer.

Analysis of Longitudinal Data

Advantages

- Increased power compared to cross-sectional studies
- Subjects serve as controls for themselves
- Separation of maturation effects and cohort effects
- Information about individual change

Challenges

- Some statistical test results only available via simulations
- Missing values (completer analysis, last observation carried forward, observed cases)
- Predictors may change over time
- In cross-over designs: carry-over effects

Notation

- $i = 1, \ldots, N$ persons
- $j = 1, \ldots, n_i$ time points for person i
- $\sum_{i}^{N} n_i$ total number of observations
- vector of observations from person i

$$(\mathbf{y}_i)_{n_i \times 1}$$

- vector of covariates of person i at time j

$$(\mathbf{x}_{ij})_{p \times 1}$$

- all covariates of person i

$$(\mathbf{X}_i)_{n_i \times p}$$

Data Structure

person	time point	observation	covariates		
1	1	y_{11}	x_{111}	$\dots$	x_{11p}
1	2	y_{12}	x_{121}	$\dots$	x_{12p}
.	.	.	.	$\dots$	.
1	n_1	y_{1n_1}	x_{1n_11}	$\dots$	x_{1n_1p}
.	.	.	.	$\dots$	.
.	.	.	.	$\dots$	.
N	1	y_{N1}	x_{N11}	$\dots$	x_{N1p}
N	2	y_{N2}	x_{N21}	$\dots$	x_{N2p}
.	.	.	.	$\dots$	.
N	n_N	y_{Nn_N}	x_{Nn_N1}	$\dots$	x_{Nn_Np}

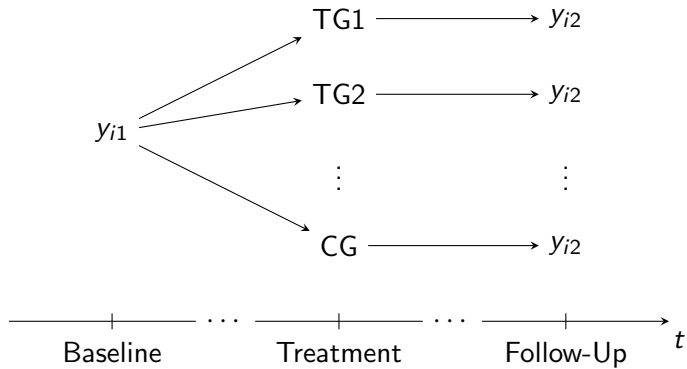
Statistical Methods

- Introductory statistics courses already discussed some procedures that allow for answering various research questions related to measuring change, such as
  - t-test for paired samples
  - Analysis of variance with repeated measurement
  - Analysis of covariance
- For a comparative evaluation of these methods, as well as the presentation of more general approaches to the analysis of hierarchical data, all of these procedures are recast as linear models
- In addition, the relevant effects are modeled in different ways as
  - Fixed effects
  - Random effects

Two Time Points

Randomized controlled trial

- At least two groups are observed before (y_{i1}) and after (y_{i2}) a treatment



- Question
  - Do the groups differ in their magnitude of change?

t-Test for Comparison of Two Time Points

- For each person consider the change score $d_i = y_{i2} - y_{i1}$ and the regression model

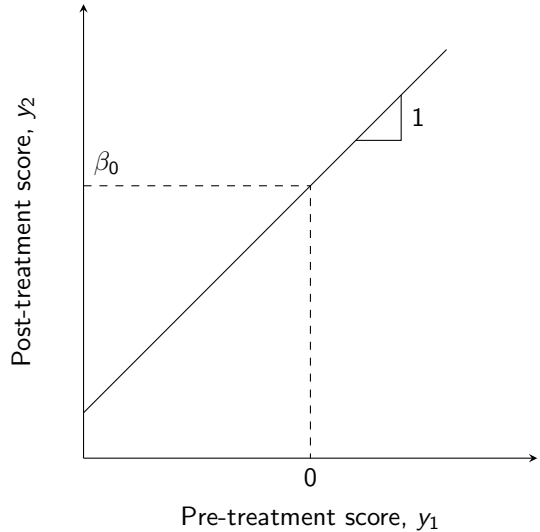
$$d_i = \beta_0 + \varepsilon_i$$
$$y_{i2} = \beta_0 + y_{i1} + \varepsilon_i$$

  with $\varepsilon_i \sim N(0, \sigma^2)$ independent
- Interpretation

$$\beta_0 \quad \text{mean change score}$$

- Hypothesis
  - Do values change from first to second time point ($H_0 \colon \beta_0 = 0$)?
- It remains unclear whether the change is due to the treatment or the result of a maturation process

t-Test for Comparison of Two Time Points



t-Test of the Follow-up Values in Two Groups

- Regression representation with $x_i = 1$ if i-th person belongs to the treatment group, and 0 otherwise:

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

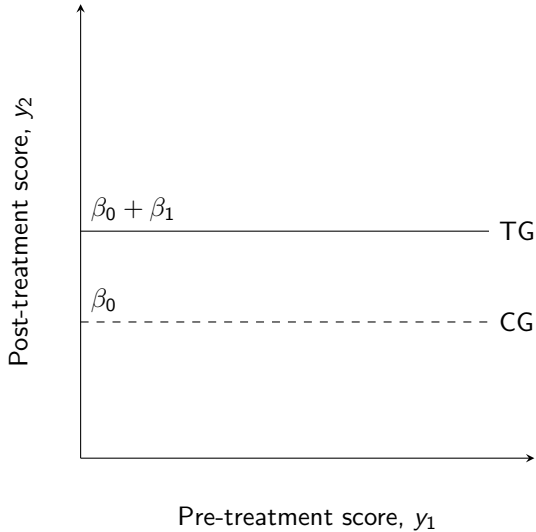
  with $\varepsilon_i \sim N(0, \sigma^2)$ independent

- Interpretation

β_0	mean follow-up value in reference group
β_1	effect of the treatment group

- Hypothesis
  - Do the groups differ at the second time point ($H_0 \colon \beta_1 = 0$)?
- Since the baseline value is not taken into account, the estimate of the change is biased

t-Test of the Follow-up Values in Two Groups



Change Score Analysis

- Regression representation

$$
\begin{aligned}
d_i &= \beta_0 + \beta_1 x_i + \varepsilon_i \\
y_{i2} - y_{i1} &= \beta_0 + \beta_1 x_i + \varepsilon_i \\
y_{i2} &= \beta_0 + y_{i1} + \beta_1 x_i + \varepsilon_i
\end{aligned}
$$

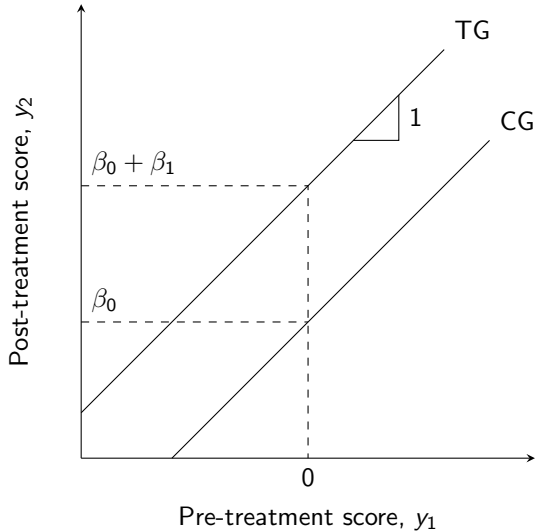
  with $\varepsilon_i \sim N(0, \sigma^2)$ independent

- Interpretation

  β_0 mean change score in the reference group
  β_1 difference from β_0 in the treatment group

- Hypotheses
  - Is there a change in the reference group ($H_0\colon \beta_0 = 0$)?
  - Does the change differ across groups ($H_0\colon \beta_1 = 0$)?

Change Score Analysis



Regression to the Mean

- The change score analysis is based on the often too restrictive assumption that the follow-up value depends on the baseline value with slope 1
- Baseline values are often negatively correlated with change scores
  - Persons with low (poor) scores improve more than persons with high scores
- This phenomenon is known as regression to the mean
- It suggests a slope < 1, which should therefore be estimated from the data

Analysis of Covariance (ANCOVA)

- Regression representation

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

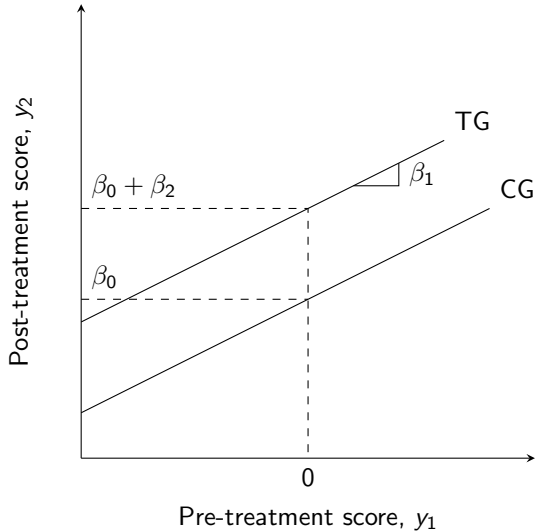
  with $\varepsilon_i \sim N(0, \sigma^2)$ independent

- Interpretation

β_0	mean follow-up value in the reference group for $y_{i1} = 0$
β_1	effect of the baseline value
β_2	difference from reference follow-up value in the treatment group

- Hypotheses
  - Are baseline and follow-up values related ($H_0 \colon \beta_1 = 0$)?
  - Do the follow-up values differ across groups for individuals with the same baseline value ($H_0 \colon \beta_2 = 0$)?

Analysis of Covariance (ANCOVA)



Adjusted Means

- Using the ANCOVA model, mean follow-up values can be predicted for the groups based on individuals sharing the same baseline value
  - For example, one obtains
    $$\hat{y}_{i2} = \hat{\beta}_0 + \hat{\beta}_1 \bar{y}_1 + \hat{\beta}_2 x_i$$
    for the average baseline value $\bar{y}_1$
- These conditional means are sometimes referred to as (baseline-)adjusted means

Different Slopes in the Groups

- Regression model

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \beta_3 (y_{i1} \cdot x_i) + \varepsilon_i$$

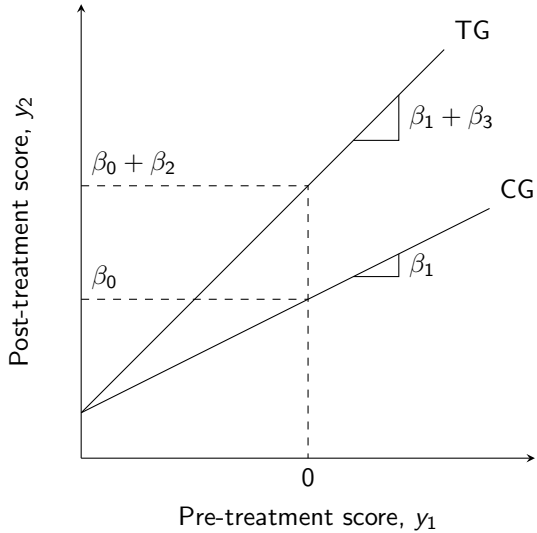
  with $\varepsilon_i \sim N(0, \sigma^2)$ independent

- Interpretation

β_0	mean follow-up value in the reference group for $y_{i1} = 0$
β_1	effect of the baseline value
β_2	deviation of follow-up value in the treatment group for $y_{i1} = 0$
β_3	difference of slopes between reference and treatment group

- Hypothesis
  - Does the treatment effect depend on the baseline value ($H_0 \colon \beta_3 = 0$)?
- Any interpretation of the adjusted means independent of the baseline value assumes $\beta_3 = 0$

Interaction between Baseline Value and Group



Analysis of Covariance with Change Score

- Regression model

$$
\begin{aligned}
d_i &= \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i \\
y_{i2} - y_{i1} &= \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i \\
y_{i2} &= \beta_0 + (1 + \beta_1) y_{i1} + \beta_2 x_i + \varepsilon_i
\end{aligned}
$$

- For testing whether changes differ across groups (average treatment effect, β_2), it does not matter whether the target variable of the ANCOVA is the follow-up score or the change score

Example: Acupuncture and Shoulder Pain

- Kleinhenz et al. (1999) studied the effectiveness of acupuncture in improving mobility in 52 patients with shoulder pain
- Patients were randomly assigned to two groups (placebo and acupuncture)
- A mobility score (Constant Murley Score) was assessed before and at the end of the treatment
- Vickers and Altman (2001) use these data to illustrate the advantages of analysis of covariance over other methods

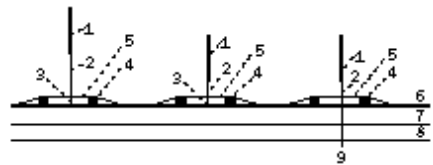
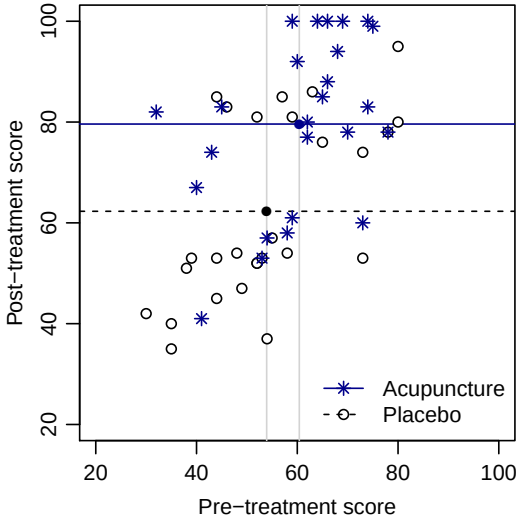


Figure 1: **Placebo needle**

Example: Acupuncture and Shoulder Pain

Follow-up analysis



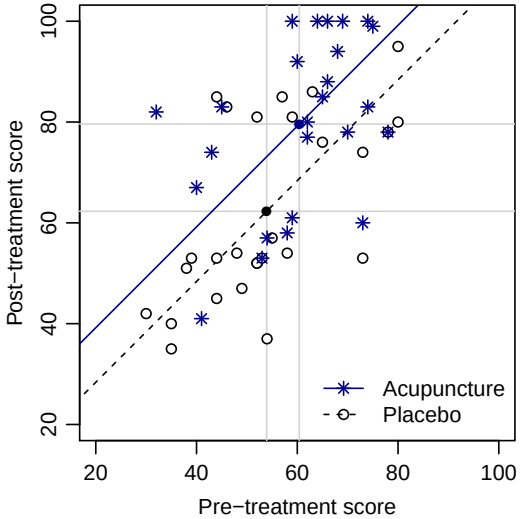
	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$\hat{\beta}_1 = 17.3, 0.95\text{-CI: } (7.5, 27.1)$$

Example: Acupuncture and Shoulder Pain

Change score analysis



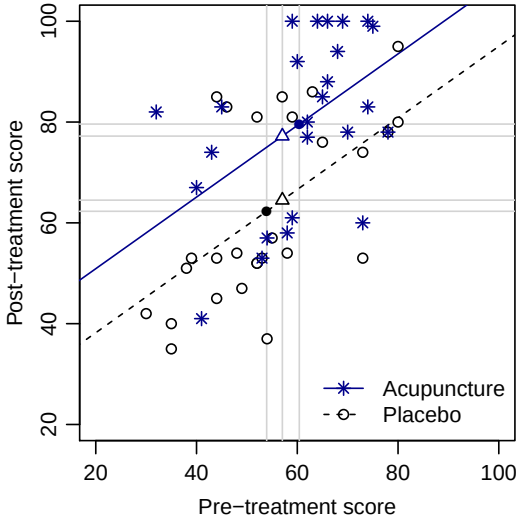
	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + y_{i1} + \beta_1 x_i + \varepsilon_i$$

$$\hat{\beta}_1 = 10.8\ (2.3, 19.4)$$

Example: Acupuncture and Shoulder Pain

ANCOVA



	Pla	Acu	Diff
Baseline	53.9	60.4	6.5
Follow up	62.3	79.6	17.3
Change sc.	8.4	19.2	10.8
ANCOVA			12.7

$$y_{i2} = \beta_0 + \beta_1 y_{i1} + \beta_2 x_i + \varepsilon_i$$

$$\hat{\beta}_2 = 12.7\ (4.1, 21.3)$$

Example: Acupuncture and Shoulder Pain

Data preparation

```
dat <- read.table("kleinhenz.txt", header = TRUE)
dat$grp <- factor(dat$grp, levels = c("plac", "acu"))
```

Follow-up analysis

```
m1 <- lm(post ~ grp, dat)
summary(m1)
confint(m1)
```

Change-score analysis

```
m2 <- lm(post ~ offset(pre) + grp, dat)
```

ANCOVA

```
m3 <- lm(post ~ pre + grp, dat)
```

Example: Acupuncture and Shoulder Pain

Testing equal slopes

```
m4 <- lm(post ~ pre*grp, dat)
anova(m3, m4)
```

Baseline-adjusted means

```
predict(m3,
        data.frame(pre = mean(dat$pre),
                   grp = c("plac", "acu")))
```

Summary

- The simplest type of longitudinal data is when multiple groups are observed at two time points (baseline and follow-up)
- The analysis of this kind of data refers to the differences

$$d_i = y_{i2} - y_{i1}$$

  or the adjusted follow-up values, which are considered to be independent
- ANCOVA
  - provides the most powerful test among all methods for detecting differences in mean change
  - however, must be interpreted with caution in case of non-randomized groups

References

Hedeker, D.R., & Gibbons, R. D. (2006). Longitudinal data analysis. Hoboken, NJ: Wiley. Chap. 1

Kleinhenz, J., Streitberger, K., Windeler, J., Güssbacher, A., Mavridis, G., & Martin, E. (1999). Randomised clinical trial comparing the effects of acupuncture and a newly designed placebo needle in rotator cuff tendinitis. *Pain*, 83, 235–241.

Vickers, A.J., & Altman, D.G. (2001). Analysing controlled trials with baseline and follow up measurements. *British Medical Journal*, 323, 1123–1124. https://doi.org/10.1136/bmj.323.7321.1123