Statistical properties of hypothesis tests using Goal Attainment Scaling

Susanne Urach
Section of Medical Statistics, Medical University of Vienna

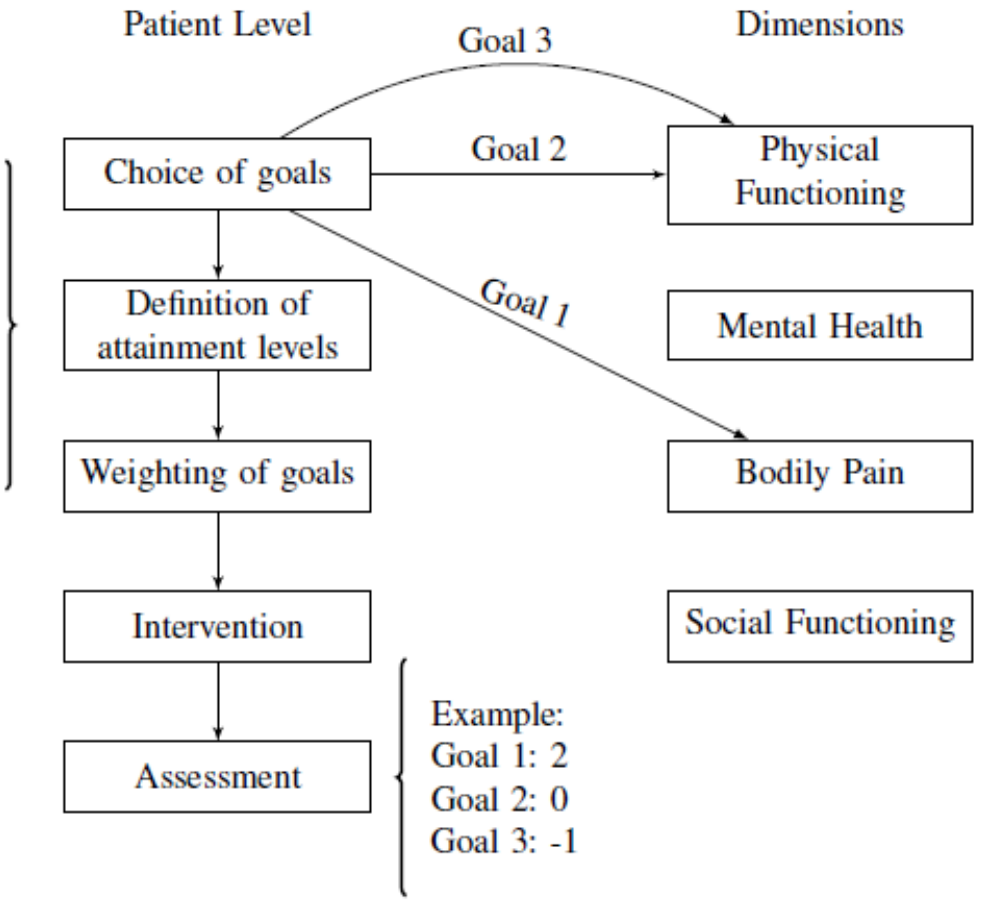
4th International Clinical Trials Methodology Conference (ICTMC), Liverpool, 10th May 2017

joint work with Gaasterland C.M.W., Rosenkranz G., Jilma B., Roes K., Van der Lee J.H., Posch M., Ristl R.

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number FP HEALTH 2013-603160. ASTERIX Project - http://www.asterix-fp7.eu/

Process of goal setting and measurement

Example:
-2 unable to walk
-1 can walk for 3 steps
0 can walk for 5 minutes
+1 can walk for 15 minutes
+2 can walk for a longer period



Use of GAS for finding a treatment effect?

Advantages and disadvantages

- Advantages:
  - Relevance of the endpoint to the patient
  - Increasing the possible sample size for the clinical trial because not all endpoints are measurable in each patient affected by a certain disease

- Disadvantages and open questions:
  - Process of goal setting very time consuming
  - Not a validated measurement instrument
  - What concept does GAS measure? Treatment effect?
  - What kind of test should one perform with GAS data?
  - How can a significant hypothesis test be interpreted?
  - Clinical interpretation of estimated change?

Research questions

- Analyzing trials:
  - How to test for a treatment effect in an optimal way?
  - What kind of weights should be applied to the individual goals?
  - Interpretation of significant hypothesis test?

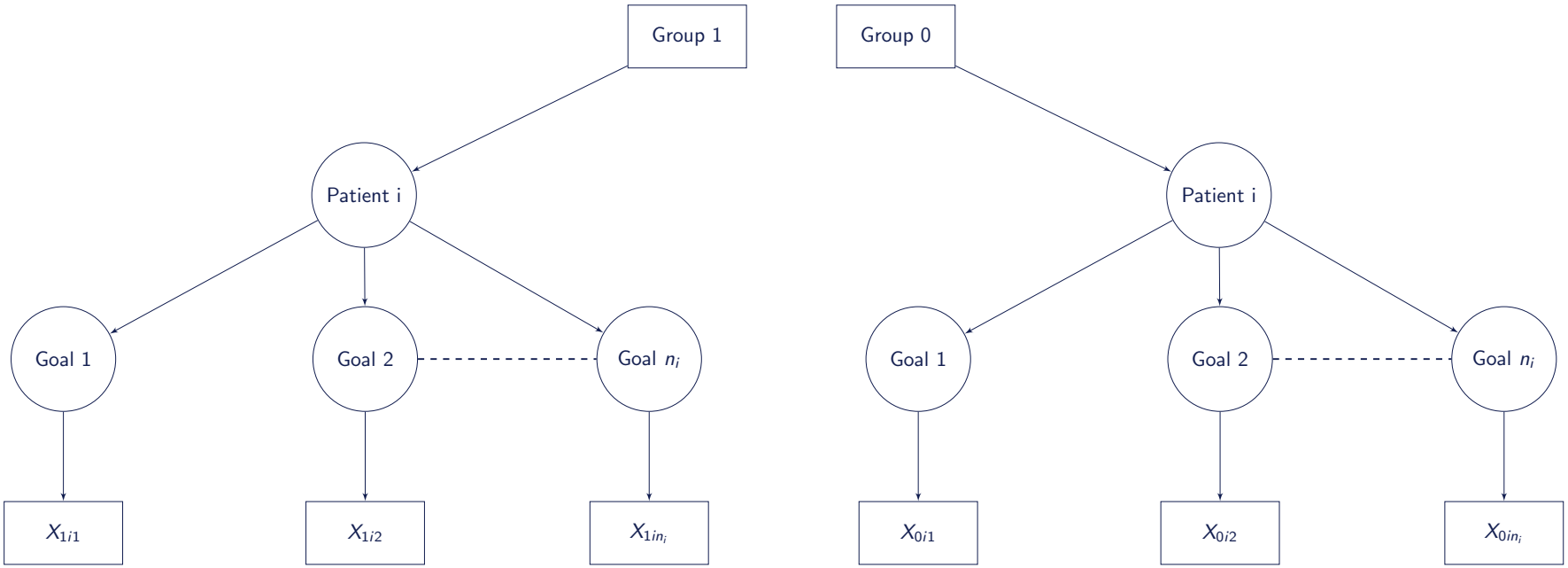
- Designing trials:

  How is a hypothesis test using a GAS endpoint affected by
  - Maximum number of goals
  - Correlation between the goals
  - Proportion of goals affected by the treatment
  - Number of attainment levels

Trial design assumptions

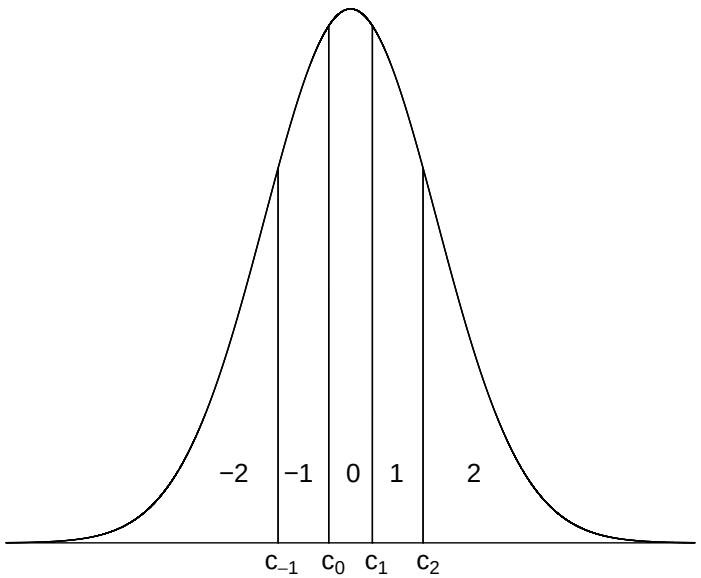
- The treatment affects the underlying mechanism of the disease and thereby several symptoms/goals.
- Randomized parallel group comparison between two arms.
- Goal outcomes are correlated within patients.
- Patients individually choose number and kind of goals.
- Same set of attainment levels for all goals, e.g. $\{-2, -1, 0, 1, 2\}$.

Multilevel hierarchical model



Discretization of continuous goal scores

- The observed ordinal goal attainment level X_{gik} for goal k of patient i in group g is the result of a discretization of a latent continuous normal variable Z_{gik}.
- The continuous variables are discretized based on thresholds c_j.



Discretization

$$
\begin{aligned}
-\infty < Z_{gik} < c_{-1} &\rightarrow X_{gik} = -2 \\
c_{-1} \leq Z_{gik} < c_0 &\rightarrow X_{gik} = -1 \\
c_0 \leq Z_{gik} < c_1 &\rightarrow X_{gik} = 0 \\
c_1 \leq Z_{gik} < c_2 &\rightarrow X_{gik} = 1 \\
c_2 \leq Z_{gik} < \infty &\rightarrow X_{gik} = 2
\end{aligned}
$$

Generating clustered ordinal outcomes

Random effect model for latent continuous goal outcome

$$Z_{gik} = u_{gi} + \mu_{gik} + \epsilon_{gik}$$

Z_{gik} ... continuous outcome for goal k of patient i in group g=0,1
u_{gi} ... random patient effect in group g $\sim N(0, \sigma_u^2)$
μ_{gik} ... random treatment effect on goal k of patient i
with $E(\mu_{gik}) = \mu_g$ and $Var(\mu_{gik}) = \sigma^2_{\mu_g}$
ϵ_{gik} ... noise $\sim N(0, 1)$

- The difference in expected goal attainment across goals and patients between treatment and control group $\delta = \mu_1 - \mu_0$ can be interpreted as the average treatment effect.

Estimating $E(X_g)$ and testing $E(X_1) = E(X_0)$

Null hypothesis $H_0 : E(X_1) = E(X_0)$

The average goal attainment level $E(X_1)$ across patients and goals of the experimental group and $E(X_0)$ of the control group are the same.

- **Challenges:**
  - Clustered observations:
    Since goal attainment levels from within patients tend to be more alike than observations from different patients, those observations provide less information about a group.
  - Different number of goals per patient:
    Less correlated or more goals of a patient provide more information about the overall treatment effect.
- **Methods:**
  - Kiresuk and Sherman formula
  - Generalised estimation equation (GEE) approach

Kiresuk and Sherman formula and GEE approach

- Composite goal score ("T score") for patient i in group g:

$$T_{gi} = 50 + \frac{10 \sum_k (W_{gik} X_{gik})}{\sqrt{(1-\rho_{gi}) \sum_k W_{gik}^2 + \rho_{gi} (\sum_k W_{gik})^2}}$$

  $X_{gik} \ldots$ ordinal goal attainment levels
  $W_{gik} \ldots$ weigths for the individual goal attainment levels
  $\rho_{gi} = \rho = 0.3 \ldots$ weighted average correlation

- The T score is a standardized weighted average of the goal attainment levels. For testing $E(T_1) = E(T_0)$ a t test can be applied.

- Generalized estimating equation (GEE) approach is used to estimate the average goal attainment $E(X_g)$ in group g

Comparison of GEE and Kiresuk method

If we assume equal correlations ρ for all pairs $(X_{gik}, X_{gik'})$:

Kiresuk method

$$\frac{\bar{T} - 50}{10} = \frac{1}{m} \sum_{i=1}^{m} \sqrt{\frac{n_{gi}}{1 + (n_{gi} - 1)\rho}} \bar{X}_{gi}$$

Sum of the standardised mean goal attainment levels.

GEE method

$$J' \Sigma^{-1} X = \sum_{i=1}^{m} \frac{n_{gi}}{1 + (n_{gi} - 1)\rho} \bar{X}_{gi}$$

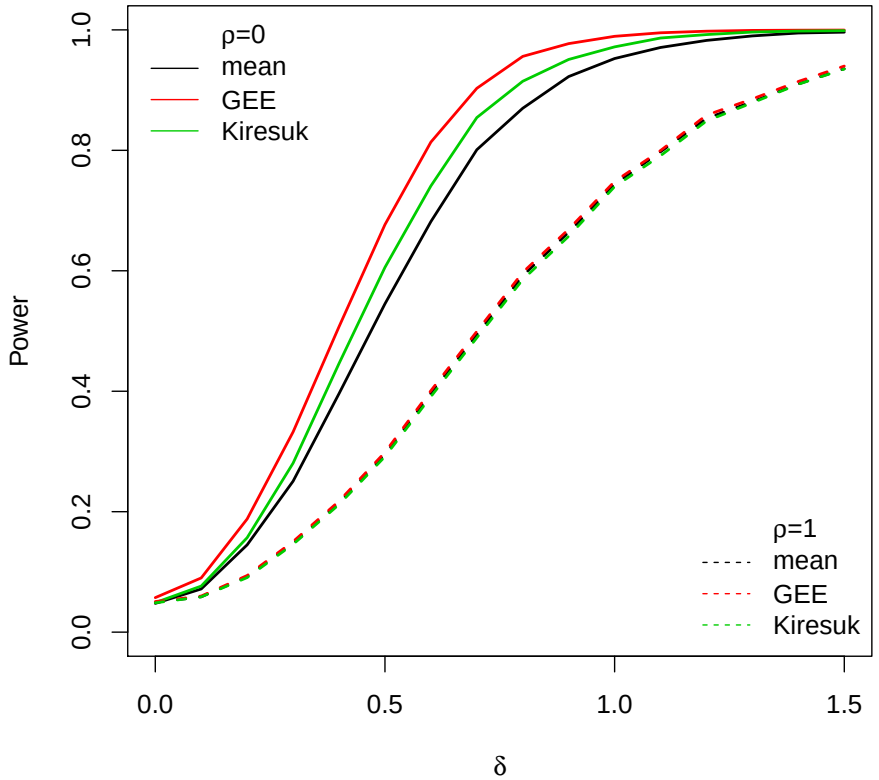
Weighting the goal attainment levels with the inverse of the covariance matrix Σ^{-1}.

Similarities between the GEE and Kiresuk method

- The means are weighted accounting for the different numbers of goals and the correlation between them.
- If the number of goals n_{gi} are independent of the goal attainment levels, it holds that $E(T_1) = E(T_0) \Leftrightarrow E(X_1) = E(X_0)$.

Power of the hypothesis test: GEE vs Kiresuk

The GEE approach has better power for testing $E(X_1) = E(X_0)$:



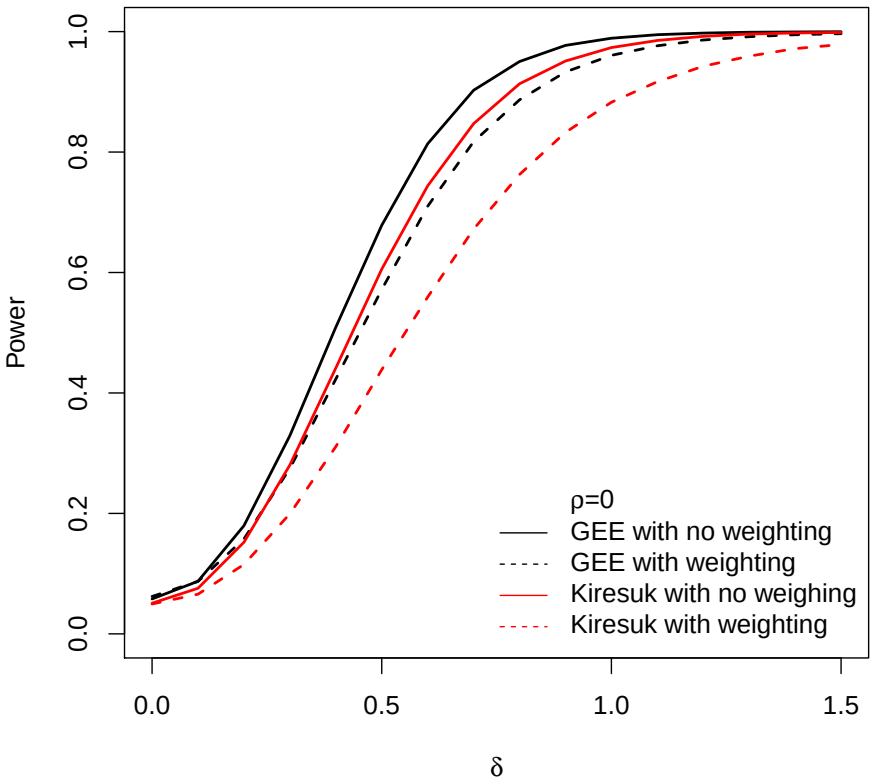
Power, $\delta = 0.5$

GEE: 68%
Kiresuk: 61%
mean: 55.4%

m=20, $n_{max} = 5$, $c_j = \Phi^{-1}(p_j)$, $p = (0.2, 0.4, 0.6, 0.8)$
$n_{gi} \sim U\{1, \ldots, n_{max}\}$, $\mu_{gik} \sim U(0, 2\mu_g)$, $\delta = \mu_1 - \mu_0$

Weighting of goal attainment outcomes

If the weights are not correlated with the treatment effect on the goals, weighting leads to a substantial loss in power.



Power, $\delta = 0.5$, $\rho = 0$

GEE no weighting: 68%
GEE with weighting: 57%
Kiresuk no weighting: 61%
Kiresuk with weighting: 43%

Impact of design aspects on power

- The power increases with the number of goals affected by the treatment, but the increase levels off: For weak correlation between goals, there can be substantial power increase up to about 5 goals.
- If goals chosen by a patient are very similar, the gain in power by adding goals is small.
- Including goals that are not affected by the treatment can lead to a substantial loss in power.
- A scale with 5 levels appears to be sufficient. Further increasing the number of level has little influence on the power.

Conclusions

- The optimal way to test for a change in average goal attainment levels between groups would be to use the GEE approach ($m \geq 20$).
- Using weights for the goal attainment levels which are not correlated with the treatment effect reduces power.
- The statistical implications of design choices (as, e.g., the maximum number of goals) should be considered.
- Clinical interpretation of a significant hypothesis test: There is a difference in the average attainment of goals.
- When presenting the results, the individual goals chosen should be investigated as well, maybe for certain domain clusters.

References


Agresti, A. and M. Kateri (2011).
Categorical data analysis.
Springer.


Hedeker, D. and R. D. Gibbons (1994).
A random-effects ordinal regression model for multilevel analysis.
Biometrics, 933–944.


Kiresuk, T. J. and M. R. E. Sherman (1968).
Goal attainment scaling: A general method for evaluating comprehensive community mental health programs.
Community mental health journal 4(6), 443–453.


Liang, K.-Y. and S. L. Zeger (1986).
Longitudinal data analysis using generalized linear models.
Biometrika, 13–22.