

## E-Appendix

### Glossary of Statistical Terms As Used in the Present Study

*Effect size* reflects the ratio between the mean change in a group over time and the (preoperative) standard deviation. An effect size of 0.2 is considered to be small, whereas a ratio of 0.8 is considered to be large. This measure gives an indication of the sensitivity of an outcome measure to change.

*Minimal detectable change* estimates the extent to which a difference between the mean score at two time points in the same population indicates a true change in outcome rather than measurement error. The minimal detectable change<sub>90</sub> refers to the threshold above which there is 90% confidence that the change in the score on the outcome measure score is due to a real change.

*Reliability* represents the proportion of the total variance in a score that is due to actual differences rather than measurement error. Reliability measures the extent to which scores obtained at different time points would provide similar results if there had been no change over that time.

*Responsiveness* is an indication of the ability of an outcome measure to detect change over time. This is distinct from discrimination, which is the ability of the measure to detect differences between groups at a single point in time.

*Standard error of measurement* is a measure of the variability in the scores of a single subject if he or she had completed the outcome measure multiple times at the same time point.

*Standardized response mean* represents a ratio between the mean change in a group over time and the standard deviation of the change. This ratio gives an indication of the sensitivity of a measure to change over time.

*Validity* relates to the degree of confidence in the inferences about the respondent that can be made on the basis of the questionnaire results. There are three predominant types of validity. Construct validity measures the extent to which the hypothesis explaining the relationship between the determinants of the attribute and the functions being measured is true. Criterion validity measures the extent to which the measure is related to the results of another measure that assesses the same attribute. Content validity determines the extent to which the measure contains a representative sample of the symptoms and functions of interest.