Clinical decisions in temporomandibular joint treatment planning and evaluation

Chapter 1. Introduction
‘Temporomandibular disorders’ is a generally accepted term used to designate the disorders of the mandibular locomotor apparatus. For convenience, disorders of the temporomandibular joint proper are commonly distinguished from non-articular disorders, which most frequently affect the masticatory musculature. The most common disorder affecting the temporomandibular joint is osteoarthritis.

Symptomatically, temporomandibular joint osteoarthritis is characterized by joint pain. Furthermore the pathological process may give rise to mechanical derangements, which may involve a displacement of the articular disk generally distinguished in disk displacement with reduction and disk displacement without reduction. Joint pain and mechanical phenomena, which are related to restriction of mandibular movements, are likely to be associated with an impairment of mandibular functions. Therefore, the treatment of temporomandibular joint osteoarthritis is usually aimed at improving mandibular function by reducing pain and enhancing joint mobility. The most favorable therapeutic interventions are non-invasive (physical therapy) and minimally invasive treatment modalities (arthrocentesis and arthroscopic surgery).

Temporomandibular joint osteoarthritis and disk displacement without reduction is characterized by restricted movement, pain and function impairment. The most representative variables for the disease of the patient are maximal mouth opening, pain and mandibular function. The clinician, however, has to decide to what extent each of these variables depart from ‘normal’ values in case of complaints. Furthermore he or she should have an idea to what extent an intervention can be expected to lead to pain reduction and increase of mobility. After re-evaluating the patient following treatment, it must be decided to what extent the relevant outcome variables must change for the intervention to be considered effective. No consistent measure of improvement, nor an operational definition of success which informs the clinician about the ‘statistically significant change’ in the individual patient is available in the temporomandibular joint field. The aims of this study were to review the available literature and identify evidence for non-invasive and minimally
invasive treatment modalities in patients with painfully restricted temporomandibular joint function (Chapter 2). Furthermore to analyse the smallest detectable difference in previously used outcome variables related to a painful restriction of the temporomandibular joint (Chapter 3) and to analyse the reliability of instruments to measure maximal mouth opening, pain and function impairment analysed by the generalizability theory and to determine their smallest detectable difference (Chapter 4, 5 and 6). Finally, to demonstrate the relevance of the smallest detectable difference, the effects of non invasive and minimally invasive treatment modalities in patients with temporomandibular joint osteoarthritis associated with non-reducing disk displacement were determined in a randomized clinical trial (Chapter 7).

**Chapter 2. Therapeutic outcome assessment**

In permanent temporomandibular disk displacement outcome studies many authors claim positive effects of non-invasive and minimally invasive treatment modalities. This literature review was undertaken to analyze whether the claimed effects are based on acceptable methodology. The recorded papers were analyzed by two independent observers according to 1. method of investigation, 2. therapeutic intervention studied, 3. therapeutic outcome variables used, and 4. claimed effectiveness of the intervention. Agreement between observers was calculated. Twenty-four papers were found in which therapeutic outcome of interventions on temporomandibular disorders were studied. Six studies applied a true experimental design. Each of these six studies compared a different set of interventions. Twenty two papers used maximal mouth opening (MMO) as an outcome variable, 9 studied pain intensity on a visual analogue scale, one paper assessed the mandibular function impairment questionnaire. Kappa for overall agreement concerning the reviewing criteria was 0.82 (p < 0.001). No distinguishing effects on MMO, pain or function impairment were reported between arthroscopic surgery, arthrocentesis and physical therapy. More results of methodological sound outcome studies evaluating the effects of non-invasive and minimally invasive modalities are needed.

**Chapter 3. Smallest Detectable Difference (SDD) in outcome variables**

The SDD is the smallest statistically significant change in measurement results. In the field of temporomandibular disorders, the SDD is not a commonly used concept. Most outcome studies are based on comparison
of group means, although this does not provide information about individual changes or about the clinical relevance thereof. The SDD for maximal mouth opening was calculated from previously published reliability coefficients and the standard deviations of different samples of healthy subjects and patients with complaints of the temporomandibular joint. In this chapter the SDD of pain intensity measured with different visual analogue scales was calculated from the reliability coefficients and standard deviations of a heterogeneous group of pain patients. The SDD of function impairment was calculated for a group of patients with complaints of the temporomandibular joint. For maximal mouth opening in healthy subjects the SDD was 5 mm. Repeated measurements improved it to 3 mm. The SDD on a visual analogue scale was 28 mm for actual pain intensity and 22 mm for minimal pain as well as for maximal pain intensity. For total function impairment of patients with complaints of the temporomandibular joint, the SDD was 8 units on a 0-68 scale. Further explicit research to quantify the SDD for different outcome variables and for different groups of patients is needed.

Chapter 4. SDD of maximal mouth opening
Changes in maximal mouth opening reflect the impact of temporomandibular disorders and the effect of a therapeutic intervention. No information about the amount of change in maximal mouth opening with regard to reasoned decision making, is available. The SDD, as a measure of reliability assessment, provides this information and is expressed in the unity of the measurement instrument. Twenty-five consecutive patients (5 males, 20 females) with a painfully restricted temporomandibular joint, participated in this study. Measurements of maximal mouth opening were performed by two well-trained observers on two separate measurement days, one week apart. The maximal mouth opening measurements were repeated thrice. Inter-, intra-observer, and test-retest reliability varied between 0.90 and 0.96. Inconsistency in measurement results analyzed in terms of absolute error variance i.e. the measurement facets plus all the interactions, represented 11% of total variance. The SDD of maximal mouth opening varied from 9 to 6 mm. For being successful in painfully restricted temporomandibular joint patients, statistically as well as clinically, the clinicians has to measure at least 9 mm of improvement in maximal mouth opening. To reduce the SDD from 9 to 6 mm, repeated measurement is necessary.
Chapter 5. Repeated pain assessment
Pain due to osteoarthritic and/or muscular pain is influenced by movement, pressure, social surroundings and rest and it can be estimated on uni- and multidimensional pain levels. A unidimensional method to measure pain is on a visual analogue scale of 100 mm and a multidimensional method to measure pain is by use of the McGill pain questionnaire. In this chapter the SDD of both uni- and multidimensional pain measurement instruments is estimated by means of the generalizability theory. Twenty-five consecutive patients (5 males, 20 females) with a painfully restricted temporomandibular joint, participated in this study. The visual analogue scales as well as the McGill pain questionnaire were completed by the patients within half an hour, followed by an assessment of the patient by a physician. Within an hour after the first session a second physician assessed the patient after the patient had completed the questionnaires. No treatment was provided within the assessment period of one week. Between-patient variance for all unidimensional visual analogue scales accounted for at least 64% of total variance, and absolute error variance accounted for at least 31% of total variance. For the multidimensional pain rating index, the between-patient variance accounted for 76% of total variance while error variance was 24%, respectively. The SDD for each single observer assessing a patient once, is 35 mm for average pain, 25 mm for minimal pain and 43 mm for maximal pain. Carrying out the measurements on two separate days and repeating it twice on each measurement day, the SDD decreased to 24 mm for average pain, 15 mm for minimal pain and 29 mm for maximal pain. The SDD for the pain rating index varied from 22.6 for a single measurement design to 14.4 on a 0 to 141.8 scale for a repeated measurement design with 2 measurement sessions and two repetitions. The generalizability coefficient for each separate pain intensity measured in separate random measurement designs varied from 0.64 for single measurement of minimal pain to 0.81 for repeated measurement of average pain and 0.89 for repeated measurements of the pain rating index.
In conclusion, an improvement of at least 43 mm to 15 mm on a VAS and 22.7 to 14.4 units of the pain rating index, depending on the dimension of experienced pain recorded and the number of repetitions on different days, is needed to be statistically significant in patients with temporomandibular disorders. The SDD of temporomandibular joint pain measurement is relatively large. Experienced pain recorded on visual
analogue scales is influenced by various external and internal factors to a larger extent than pain assessed on the pain rating index.

**Chapter 6. SDD of mandibular function impairment**

Mandibular functions such as speech, laughing, yawning, mastication and taking a large bite may be impaired in case of temporomandibular disorders. The Mandibular Function Impairment Questionnaire (MFIQ) is a validated instrument that scores perceived difficulty of representative mandibular functions in relation to jaw complaints. However, the reliability of the MFIQ has never been adequately tested. Generalizability and decision studies are currently proposed to assess the reliability of a measurement device. The purpose of this study was to assess the reliability of the MFIQ in terms of the SDD. The MFIQ was completed by 25 consecutive patients with painfully restricted temporomandibular joints on two separate measurement days, one week apart, using two consecutive sessions per day. Spearman’s $r$ was calculated for test-retest reliability. Variance components such as patients, measurement days, repetitions and all their interactions were analyzed in the generalizability study. In the decision study the SDD was calculated for different days and repetitions. Spearman’s $r$ varied from 0.69 to 0.96. The between patient variance and the error variance contributed 81% and 19% to total variance, respectively. The MFIQ is a reliable instrument to assess mandibular function impairment. The minimal amount of change to detect is 14 units on a 0-68 scale. Reliability in terms of the SDD increases by repeating the measurement twice on two separate days the SDD improves to 10 units.

**Chapter 7. Deciding on success with regard to function impairment**

Temporomandibular joint osteoarthritis is commonly accompanied by a disk derangement giving rise to impaired mandibular function due to pain and a restricted mouth opening. Probably because of the lack of randomized clinical trials, many authors in the temporomandibular joint field use ‘success rates’ to assess the therapeutic outcome of their interventions. However, published success rates are commonly based on convention rather than on proper research designs and statistical analyses. In the context of temporomandibular disorders, a logical measure for ‘success’ as primary outcome measure would be ‘sufficiently improved mandibular function impairment’ as a result of restored mobility and pain reduction in the temporomandibular joint. In this chapter the SDDs of mandibular function impairment, maximal mouth
opening and pain were used as a measure of statistically significant change in individual patients in a randomized clinical trial evaluating the effects of physical therapy, arthrocentesis, and arthroscopic surgery. Because of disappointing inclusion results (sample size was 90 patients) the analyses were based on the results of 20 patients (7 physical therapy, 7 arthrocentesis and 6 arthroscopic surgery). In this chapter we have summarized the results in a best-case and worst-case analyses. No statistically significant differences between the three treatment groups were found with regard to ‘therapeutic success’ as ‘success based on the change achieved by one of the therapeutic interventions’, neither in the best nor in the worst case analysis. The power of the analyses was 0.33 and 0.45 in case of the best and worst case analyses, respectively. Analyzing the results for two groups, i.e. a non-invasive group (n=8) and minimally invasive group (n = 13), the Chi-square p value was 0.19 in the best case analyses and 0.01 in the worst case analyses, with a power of 0.71 and 0.47, respectively. Prior to the start of this randomized clinical trial, the influences of the natural course of the disease, biological variations and measurement errors were analysed using a so-called generalizability study and a decision study in a comparable group of patients. Most of the outcome studies evaluated in chapter 2 used pre-experimental designs in which the variance components mentioned could be responsible for supposed treatment effects. To avoid this negative side effect and to be able to make the optimal clinical decision as to whether an individual improved on mandibular function, mouth opening or uni- and multidimensional pain, the SDD has proved to be an objective, statistically significant, and individually applicable cut-off point for change. Mandibular function is likely impaired due to pain and/or a restricted mobility of the temporomandibular joint. It is striking that the improvement noted for mandibular function and the reduction of pain was more impressive than the improvement of mouth opening for these three interventions. In the worst cases analyses only 0 to 13% of the cases improved on mouth opening whereas 56 to 13% improved on function impairment after arthroscopic surgery, arthrocentesis or physical therapy. Until demonstrated otherwise, we expect that clinical differences between non-invasive and minimally invasive interventions with regard to the actual therapeutic effect are marginal. Considerations such as general versus local anesthesia and medical costs could be more decisive criteria than expected statistical significant change in mandibular function impairment.
Chapter 8. General discussion
From our review described in chapter 2, it has become apparent that the evidence for the beneficial effects of physical therapy, arthrocentesis and arthroscopic surgery is incomplete. No properly conducted clinical trials are available to perform a meta-analysis. In addition, it became apparent that a wide variety of outcome measures have been used for evaluating treatment modalities for temporomandibular disorders. This variety makes comparison between studies difficult, if not impossible. Moreover, the validity and reliability of a considerable number of these measures has not been sufficiently established. Thus, there is not only a need for properly designed and conducted trials, but also for outcome measures that can be generally applied as described in chapter 4, 5 and 6 to enable future comparison between studies and meta-analysis of different trials. A major advantage of the SDD is that it exceeds the natural course of the disease, the biological variance of the outcome variable and measurement error and is generalizable to all patients that comprise the sample. In temporomandibular joint osteoarthritis accompanied by a disk derangement impaired mandibular function is largely due to pain and restricted mouth opening. Because of the interrelationship between pain, range of opening and mandibular function impairment, the variation of each of these variables is influenced by the other variables. It has been made plausible that the natural course of the disease contributes to changes within treatment as well as in no-treatment groups and this aspect should, therefore, be incorporated in the outcome measure. The SDD is a measure for statistically significant change that can be used in the individual patient as a measure of change. Responsiveness is defined as the ability of an outcome instrument to detect clinically important changes in a specific condition. However, there is no consensus on the appropriate strategy to quantify responsiveness. Most of these approaches are based on the average change in scores since baseline in self rated clinically stable and improved patients. These responsiveness ratios are unitless ratios of the variances of only one sources of variation (measurement days) in a changed and a non-changed group of patients. The SDD, however, is based on the analyses of error variance of different sources of variation (observer, days and repetitions and their interactions) causing variance around the observed (change) scores. The higher the variance (large SDD's) the less responsive the outcome variable is expected to be. Increasing the number of repetitions the SDD is reduced and could therefore improve responsiveness of the outcome variable. Because of an insufficient number of patients the Minimally
Clinically Important Difference could not be estimated. Further research is needed to investigate the SDD for its responsiveness and clinical significance in individual patients. The more reliable the measurement procedure the smaller the SDD will be and we do not know the borders of the SDD with respect to a notified change by the patient. We recommend to include analyses of patients expectations of pain reduction and patients perceived pain history because we belief in the interaction between disappointment (high expectations versus bad remembering) and therapeutic success are interrelated. However, a more structural data collection could verify this hypothesis.

**Recommendations**
The following recommendations are suggested based on this thesis:

- Implementation of the smallest detectable difference in clinical practice for the benefits of patients and clinicians information about real expectations and percentages of success.
- To continue research on personalized smallest detectable differences (i.e. each clinician his/her own smallest detectable difference).
- To investigate responsiveness and the minimally clinically important difference in relation to the smallest detectable difference.
- To develop a Relational Database Management System based on Internet technology for the purpose of a structural standardized data collection. Large multi-center randomized clinical trials may be supported by such a system and the smallest detectable difference could be integrated as a reliable criterion of success.