Deciding on success with regard to mandibular function of noninvasive and minimally invasive treatment of temporomandibular joint osteoarthritis in a randomized clinical trial

Kropmans ThJB, Dijkstra PU, Stegenga B, Spijkervet FKL, de Bont LGM

Introduction

Temporomandibular joint osteoarthritis is commonly accompanied by a disk derangement giving rise to impaired mandibular function due to pain and a restricted mouth opening. Common treatment options include physical therapy, arthrocentesis, and arthroscopic surgery (Stegenga et al., 1989; Stegenga et al., 1993; Gray et al., 1994; Linde et al., 1995; Fridrich et al., 1996). However, few studies have been published reporting on differences in treatment effects based on a randomized clinical trial design (Kropmans et al., 1999a). Both arthrocentesis and arthroscopic surgery are frequently followed by a physical therapy protocol, thus resulting in a combined treatment effect. In these cases, it is not clear to what extent each of these modalities contribute to the treatment outcome.

In the field of temporomandibular joint disorders, therapeutic success is frequently based on designs using a one group pre- and post treatment comparison (Kropmans et al., 1999a). However, it is now generally accepted that cause-effect relationships can be evaluated only by including a control group and at random treatment assignment (Altman, 1997). Clinicians are supposed to base their treatment of choice on the best available evidence, according to the principles of ‘evidence based medicine’ (Fletcher and Sackett, 2000). With regard to treatment outcome, the best possible evidence is that from randomized clinical trials of sufficient size and/or the combined outcome of smaller randomized clinical trials resulting from a meta-analysis in a systematic review (Sackett and Rosenberg, 1995; McNeill, 1997; Raphael and Marbach, 1997; Greene et al., 1998). From the systematic review of outcome assessment by Kropmans et al. (1999a) it appeared that well-performed large clinical trials evaluating physical therapy, arthrocentesis and arthroscopic surgery in patients with permanent temporomandibular
joint disk displacement are not available (Kropmans et al. 1999a). Differences in the outcome variables used in smaller trials preclude the proper performance of a meta-analysis. 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 (Kropmans et al., 1999a). However, published success rates are commonly based on convention rather than on proper research design and statistical analyses. In the context of temporomandibular joint disorders, a logical measure for ‘success’ as primary outcome measure would be ‘sufficiently changed mandibular function impairment’ as a result of restored mobility and pain reduction in the temporomandibular joint.

In previous chapters we described and evaluated the smallest detectable difference as an individual measure of change in mandibular function impairment, maximal mouth opening and unidimensional pain and multidimensional pain of the temporomandibular joint. The smallest detectable difference is the smallest statistically significant difference that can be detected by a measurement device. This measure is increasingly used to enhance decision making based on statistically significant change in the individual patient (Kahn, et al. 1996; Balakrishnan 1998). Thus, success may well be defined as a positive change of at least the smallest detectable difference of mandibular function impairment. Moreover, by using the smallest detectable difference in a randomized clinical trial as the threshold for the definition of successful improvement, the patient and/or clinician are better informed about the a priori chances of success in individual cases as compared with using classical outcome analyses of a randomized clinical trial, which is based on between-group comparison.

The purpose of this study was to evaluate the short and long-term ‘therapeutic success’ (i.e. improvement of mandibular function due to a specific intervention) of non-invasive (physical therapy) and minimally invasive (arthrocentesis and arthroscopic surgery) treatment of patients with temporomandibular joint osteoarthritis involving a non-reducing disk displacement using the smallest detectable difference as threshold for ‘success’.
Patients and Methods

Patients
Potential participants of the study were patients referred to the department of Oral and Maxillofacial Surgery of the Groningen University Hospital with clinically suspected temporomandibular joint osteoarthritis associated with non-reducing disk derangement. These patients were initially treated with load reduction instructions, a diet restricted to soft food, and standard medication with non-steroidal anti-inflammatory drugs. After four weeks, they were re-assessed and only those who obeyed pre-defined inclusion criteria and gave their written informed consent were enrolled in a randomized clinical trial for the evaluation of physical therapy, arthrocentesis or arthroscopic surgery. The inclusion criteria included restriction of maximal mouth opening to less than 33 mm, mild to severe pain larger than 20 mm assessed on an uni-dimensional visual analogue scale for average pain, and mild to severe function impairment at least equal to 10 units on the Mandibular Function Impairment Questionnaire. In addition, the non-reducing disk displacement had to be confirmed with magnetic resonance imaging. The inclusion criteria were compatible with the research diagnostic criteria previously defined by Dworkin and LeResche (1992).

Interventions
Eligible patients were randomly assigned to one of three treatment groups, i.e., physical therapy, arthrocentesis, arthroscopic surgery. A computerized minimization method was used to balance the treatment totals for three patient factors, i.e., mouth opening restriction, age and gender. All treatments were initiated between 4 to 6 weeks after randomization and were performed according to standardized written procedures (outlined below). Each of the treatment procedures was performed by one clinician.

Non-invasive treatment: physical therapy
Physical therapy consisted of techniques focusing on increasing range of motion (i.e., grade II joint distraction and anterior glide techniques combined with active protrusive and laterotrusive movements) and techniques to decrease pain, such as ice-application (Dijkstra, 1996). All physical therapy techniques were performed without the use of
medication. Three to six treatment sessions were applied within a time span of three weeks.

**Minimally invasive treatment: arthrocentesis and arthroscopic surgery**

Arthrocentesis was performed under local anesthesia in an outpatient setting. The upper joint space was punctured with a 19-gauge needle inserted in the region of the articular fossa and with a second needle inserted in the area of the articular eminence to enable a free flow of fluid through the upper compartment. Saline solution, warmed to body temperature, from an infusion bag placed about one meter above the level of the joint was connected to one of the needles and allowed to freely flow through the upper joint space. After a lavage of about 200-300 ml the procedure, which took about 15 minutes, was terminated and was not followed by physical therapy (Nitzan et al., 1991)

In case of arthroscopic surgery, the patient was administered to the hospital one day prior to, and released from the hospital one or two days after the procedure. The procedure was performed under general anesthesia and complete neuromuscular relaxation. Under constant irrigation with saline solution warmed to body temperature, a double puncture technique was performed and the upper joint space was subsequently inspected. Depending on the findings during inspection, surgical procedures were carried out with the aim to reduce the synovitis and to mobilize the disk. The arthroscopic procedure was not followed by physical therapy post-operatively (Davis et al., 1991).

**Outcome measures**

A clinical evaluation was performed 6 weeks, and 3, 6 and 12 months after treatment. Outcome measures included mandibular function impairment, maximal mouth opening and uni- and multidimensional pain levels. All assessments were performed by an independent observer.

**Mandibular function impairment**

A statistically and clinically significant improvement of mandibular function was considered the primary outcome measure for therapeutic success. This was operationalized by an improvement of the score on the Mandibular Function Impairment Questionnaire at least equal to the smallest detectable difference. The Mandibular Function Impairment Questionnaire is a validated and reliable instrument that scores perceived difficulty of 17 representative mandibular functions in relation to jaw
complaints. It consists of two scales, a ‘mastic’ scale and a ‘non-mastic’ scale. The mastic scale assesses difficulty the patient has with a bite and chewing hard, soft and resistant foods, while the non-mastic scale consists of items assessing the extent of difficulty the patient has with non-masticatory jaw activities such as social activities, speaking and yawning. The possible answers to each item are scored on a Likert scale, ranging from no difficulty (0) to very much difficulty or impossible without help (4). The sum score for function impairment ranges from 0 - 68 (Stegenga, et al. 1993; Kropmans, et al. 1999).

**Maximal mouth opening**
Maximal mouth opening was defined as the maximal distance between the incisal edges of the maxillary and mandibular central incisors. Maximal mouth opening was measured interincisally with a millimeter ruler after applying gentle pressure reaching maximal mouth opening. The patient was sitting in an upright position in a dental chair, the head supported by the headrest. For each subject, the measurements were carried out by one independent observer in two measurement sessions, each consisting of three measurements (Kropmans, et al. 2000). The first measurement session was performed at the beginning and the second session at the end of the clinical evaluation after filling out the questionnaires.

**Pain**
The average, minimal and maximal experienced pain during the week prior to assessment was repeatedly recorded within the questionnaires on non-verbal horizontal VAS’s of 100 mm, the two extremes representing the absence of pain and the worst imaginable pain, respectively. Multidimensional pain experiences were assessed with the McGill Pain Questionnaire (Dutch Language Version, MPQ-DLV), scoring the ‘number of words chosen’ and the pain rating index (van der Kloot, et al. 1995; Kropmans, et al. 2000a).

**Statistical analysis**
Successful and unsuccessful treatments were defined as at least equal to the smallest detectable differences respectively smaller than the smallest detectable difference of mandibular function, mouth opening and uni- and multidimensional pain. Between group differences were calculated using a chi-square analyses (Kruskal Wallis) of the percentage successful change of the outcome measures. Based on the assumption
that an effect size of 5 mm of mouth opening would have been sufficient to improve mandibular function, a power of at least 0.80 could be obtained on a significance level of 0.05, when including at least 30 patients. Lost cases were analysed in a best and worst case analyses. In the best case analyses, lost cases were considered as being successfully treated versus in the worst case analyses they were considered as unsuccessfully treated. Worst and best case analyses and the summation of total medical costs were performed twelve month after the interventions.

Results

Patients
Initially, during a three year period, 69 patients had a clinical diagnosis of temporomandibular joint osteoarthritis accompanied with non reducing disk displacement. The characteristics of these patients are listed in table 7.1. After four weeks of initial therapy, the inclusion criteria were fulfilled and the diagnosis could be confirmed with magnetic resonance imaging in 24 patients. These patients were randomized to the therapeutic groups.

Table 7.1 Patients characteristics (n = 69)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age years</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>Duration of complains months</td>
<td>25</td>
<td>41</td>
</tr>
<tr>
<td>Maximal mouth opening mm</td>
<td>33</td>
<td>10</td>
</tr>
<tr>
<td>Average pain (visual analogue scale) mm</td>
<td>38</td>
<td>22</td>
</tr>
<tr>
<td>Minimal pain (visual analogue scale) mm</td>
<td>22</td>
<td>17</td>
</tr>
<tr>
<td>Maximal pain (visual analogue scale) mm</td>
<td>55</td>
<td>23</td>
</tr>
<tr>
<td>Pain Rating Index (McGill painq. DLV) units</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>Mandibular Function Impairment (MFIQ) units</td>
<td>33</td>
<td>10</td>
</tr>
</tbody>
</table>

Interventions
Three patients withdrew after being assigned to the arthroscopic surgery group, one patient (assigned to the arthrocentesis group) was lost to follow-up due to moving to another part of the country, and one patient (assigned to the physical therapy group) refused further treatment. Based
on missing data of these patients no intention to treat analyses is available. So, 20 patients have actively been treated, seven non-invasively and 13 with invasive modalities (6 patients with arthroscopic surgery, 7 with arthrocentesis).

**Outcome measures**

Success (i.e., improved function impairment of at least equal 10 units), as well as improvement of maximal mouth opening and pain release after 6 weeks, and 3, 6 and 12 months are listed in table 7.2.

<table>
<thead>
<tr>
<th>Table 7.2 Within group success rates of the primary outcome variable ‘success’ on mandibular function and the secondary outcome variables maximal mouth opening, average, minimal and maximal pain (VAS) in the week prior to the investigation and Pain Rating Index of the McGill pain questionnaire (DLV).</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDD</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td><strong>Arthroscopic Surgery</strong></td>
</tr>
<tr>
<td>after 6 weeks</td>
</tr>
<tr>
<td>after 3 months</td>
</tr>
<tr>
<td>after 6 months</td>
</tr>
<tr>
<td>after 12 months</td>
</tr>
<tr>
<td><strong>Arthrocentesis</strong></td>
</tr>
<tr>
<td>after 6 weeks</td>
</tr>
<tr>
<td>after 3 months</td>
</tr>
<tr>
<td>after 6 months</td>
</tr>
<tr>
<td>after 12 months</td>
</tr>
<tr>
<td><strong>Physical Therapy</strong></td>
</tr>
<tr>
<td>after 6 weeks</td>
</tr>
<tr>
<td>after 3 months</td>
</tr>
<tr>
<td>after 6 months</td>
</tr>
<tr>
<td>after 12 months</td>
</tr>
</tbody>
</table>

No statistically significant differences were found between arthroscopic surgery, arthrocentesis and physical therapy (Kruskall Wallis p varied from 0.11 to 0.99)

In table 7.3, we have summarized the therapeutic success that would have been obtained in case the randomized patients who were not available for further analysis all would have been successful (best-case) and unsuccessful (worst-case), respectively (Table 7.3). No statistically significant differences between the three treatment groups were found with regard to ‘therapeutic success’ as ‘success based on the change achieved by the therapeutic intervention’, neither in the best nor in the worst case analysis (Kruskal-Wallis p values varied between 0.15 and 0.91). However, the power of the analysis 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.

Table 7.3 Best and worst case analyses (percentages success rates) one year after arthroscopic surgery, arthrocentesis and physical therapy

<table>
<thead>
<tr>
<th>SDD</th>
<th>Therapeutic success on Impairment 10 units</th>
<th>Maximal Mouth Opening 6 mm</th>
<th>Average Pain (VAS) 24 mm</th>
<th>Minimal Pain (VAS) 15 mm</th>
<th>Maximal Pain (VAS) 29 mm</th>
<th>Pain Rating Index 14 units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthroscopic Surgery</td>
<td>89 / 56</td>
<td>33 / 0</td>
<td>67 / 33</td>
<td>33 / 0</td>
<td>100 / 67</td>
<td>67 / 33</td>
</tr>
<tr>
<td>Arthrocentesis</td>
<td>63 / 38</td>
<td>38 / 13</td>
<td>88 / 63</td>
<td>50 / 25</td>
<td>75 / 50</td>
<td>63 / 38</td>
</tr>
<tr>
<td>Physical Therapy</td>
<td>50 / 13</td>
<td>50 / 13</td>
<td>75 / 37</td>
<td>50 / 13</td>
<td>75 / 38</td>
<td>50 / 13</td>
</tr>
</tbody>
</table>

No statistically significant differences were found for best nor for worst case analyses between arthroscopic surgery, arthrocentesis and physical therapy (Kruskall Wallis p varied from 0.15 to 0.91)

Discussion

Throughout the years, very few studies have been published reporting on differences in treatment effects based on a randomized clinical trial design (Kropmans et al, 1999a). At first glance this may be surprising, because of the wide agreement that the randomized clinical trial is the standard design for evaluating treatment outcome (Altman, 1997). It is important that a clinical trial is carried out according to a strict protocol, in which its main features are specified (Pocock, 1996):

- patient selection and randomization (inclusion criteria, registration and randomization procedure, study population size)
- treatment schedules
- methods of patient evaluation
- trial design and monitoring of trial progress

Prior to the start of this randomized clinical trial, the influences of the natural course of the disease, biological variation and measurement error were analysed using a so-called generalizability study and a decision study in a comparable group of patients. Most of the outcome studies

Chapter 7: deciding on success 86
Chapter 7: deciding on success

Evaluating outcomes measured by Kropmans et al. 1999 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 smallest detectable difference has proved to be an objective, statistically significant, and individually applicable cut-off point for improvement after the initial period of load reduction, a diet of soft food and medication. In case of an unsuccessful initial treatment inclusion in the trial was considered.

Several important factors, some of which are beyond the direct influence of the researcher, have to be taken into account to obtain a clinically and statistically useful study population. The size of the study population, strictly speaking, depends on the estimated difference in effect of the interventions, the predefined significance level, and the power of the analysis to be conducted. Although we had expected to include 90 patients with temporomandibular joint osteoarthritis accompanied with non-reducing disk displacement in a three years period (based on a power-analysis, and non-response and drop-out estimation), monitoring the trial progress we decided to terminate the inclusion period after two years because of the disappointing inclusion result. In the temporomandibular joint field, treatment outcome studies have yielded results that are more or less comparable, regardless of the treatment modality. This would increase the sample needed for an RCT that compares different treatment modalities to a large extent (Pocock, 1996; Altman, 1997). A priori, we expected non-invasive treatment to be inferior to minimally invasive treatment with respect to the improvement of mandibular function. According to our best case analyses (table 2 and 3), 89% of the patients achieved a successful result with arthroscopic surgery, while 63% and 50% were successful after being treated with arthrocentesis and physical therapy, respectively. However, to achieve these results with a power of at least 80%, at least 81 patients would have been included in each treatment group.

Other factors also influence the sample size that will eventually be obtained. Restriction of the base population by patient selection criteria, especially with regard to the stage of the disease, will result in a homogenous population. However, there is a risk that an insufficient number of these patients actually exists in ‘the real world’. We chose to apply inclusion criteria that are widely accepted in the field, i.e. criteria compatible with the research diagnostic criteria defined previously.
(Dworkin and LeResche, 1992). A draw-back of using these criteria is that they have been defined by consensus rather than derived empirically. Empirically derived criteria probably are more realistic and consistent with the actual patient population than are criteria defined by consensus, which have a largely theoretical basis.

Another problem related to patient selection criteria was that the majority of the patients (in our study up to 65%) improved by initial therapeutic measures (i.e., reassurance and explanation, load reduction instructions, a soft food diet, and non-steroidal anti-inflammatory drugs) to an extent that they did not obey the inclusion criteria anymore. In addition, inherent to a randomization study is an expected drop-out percentage due to randomization as such. However, also following randomization patients may decide not to participate (as did three patients who appeared to be assigned to the arthroscopic surgery group), presumably because they had preference for a certain intervention modality.

Eventually, only 20 patients could be included in our analysis. Similar disappointing recruitment results have been reported before (Stegenga, et al. 1993). The limited number of clinical trials available in this field is, therefore, not so surprising after all.

Mandibular function is likely impaired due to pain and/or a restricted mobility of the temporomandibular joint (Stegenga, et al. 1993). 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 after arthroscopic surgery, arthrocentesis or physical therapy. These results confirm the results of Stegenga et al. (1993) reporting a rapid decrease in pain and a ‘significant’ improvement in functional ability. However, both studies suffer from a lack of patients and, consequently from insufficient power, to draw meaningful conclusions concerning the relative efficacy of the interventions investigated. Our results do not support other, non-randomized, studies reporting pain relief to be as effective as improvement of mobility (Nitzan et al., 1997a; Nitzan et al., 1997b; Hori et al., 1999; Miyamoto et al., 1999). None of these studies have reported data with respect to mandibular function impairment. We recommend that future studies on treatment outcome in this field include this variable, for which a valid and reliable questionnaire is available (Kropmans et al., 1999).
As postulated before, the smallest detectable difference has proven to be a statistically significant measure for success in the individual patient. It is based on proper research criteria rather than on convention and beliefs.

The smallest detectable difference is a measure of reliability complementary to classical psychometric analyses of reliability and is expressed in the same unit as the measurement device involved, in contrast to correlation coefficients (Bland and Altman, 1986). Another major advantage is that the smallest detectable difference provides a measure for the actual treatment effect, beyond influences from the natural course and biological variations. Therefore, it can be applied both in ‘good medical practice’ and in scientific studies.

In most (non-randomized, non-controlled) treatment outcome studies concerning both minimally invasive and non-invasive treatments of temporomandibular joint osteoarthritis and related disorders, a success rate of about 70 to 90% is reported, irrespective of the modality evaluated. Lower percentages are not commonly reported in the literature, which possibly may be due to publication bias (Shuster 1998).

As stated earlier, to demonstrate small differences in success rate between two (or more) interventions to be significant, large samples are required (for instance to demonstrate a 5% difference between two treatment modalities with \( \alpha = 0.05 \) and \( \beta = 0.20 \) requires a sample size of more than 2000 per treatment group). Under these circumstances, one may question the clinical significance of such a difference. Applying the smallest detectable difference, which generally is relatively large, the success rate will probably decrease because of a higher cut-off point for success and probably less patients will be needed for sufficient power. A large randomized clinical trial could be initiated in a center generating sufficient patients or by a multicenter trial. Although it is a realistic alternative, the latter solution is relatively expensive and time consuming, while specific methodological issues, such as homogeneity of patient characteristics and diagnosis, must be addressed. An efficient database management information system based on research criteria and build on Internet technology may generate structural and scientifically reliable data in more than one center and could, therefore, be an appropriate solution to achieve efficiency in a multi-center approach (Scully et al., 1997; Scott and Lenert, 1998; Frenot and Laforest, 1999). 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 a more decisive criteria (table 7.4).

<table>
<thead>
<tr>
<th></th>
<th>Therapeutic success on Impairment 10 units</th>
<th>Total medical costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthroscopic Surgery (n=8)</td>
<td>89 / 56 %</td>
<td>$ 14,736,-</td>
</tr>
<tr>
<td>Arthrocentesis (n=8)</td>
<td>63 / 38 %</td>
<td>$ 2,694,-</td>
</tr>
<tr>
<td>Physical Therapy (n=8)</td>
<td>50 / 13 %</td>
<td>$ 560,-</td>
</tr>
</tbody>
</table>

Using the smallest detectable difference as a measure of success in a randomized clinical trial is a valuable contribution to ‘evidence based medicine’ because both the clinician and the patient is informed about the best possible evidence available and the success rates that can be achieved. The outcome of a systematic review with a meta-analysis provides evidence for a therapeutic choice based on group results, but does not directly inform the patient about the chances of success.
References


Fletcher S, Sackett DL, Levels of Evidence. http://cebm.jr2.ox.ac.uk/docs/levels.html


*Chapter 7: deciding on success* 92


