2 Therapeutic outcome assessment in permanent temporomandibular joint disc displacement; a review

Kropmans ThJB, Dijkstra PU, Stegenga B, de Bont LGM
Reprinted with permission

Introduction

Therapeutic outcome studies are carried out to evaluate the effects of a therapeutic intervention on signs and symptoms of a disease (Stegenga et al., 1993b). Arthroscopic surgery, arthrocentesis and physical therapy have commonly been used as therapeutic interventions for permanent temporomandibular joint disc displacement over the past 15 years (Austin and Shupe, 1993; McCain et al., 1992; Nitzan et al., 1991; Nitzan et al., 1990; Nitzan and Dolwick, 1991; Austin and Shupe, 1993; Humberger and Humberger, 1992; Bertolucci, 1992). For the assessment of the therapeutic efficacy in permanent temporomandibular joint disc displacement, maximal mouth opening, pain intensity, and mandibular function impairment are currently used as outcome variables in group comparison studies. Many authors claim effects of their interventions on these variables after treating permanent temporomandibular joint disc displacement (Dimitroulis et al., 1995; Gray et al., 1994; Hoffman and Cubillos, 1994; Kuwahara et al., 1994; Petersson et al., 1994; Bertolami et al., 1993; McCarty, Jr. and Darnell, 1993; Stegenga et al., 1993b; McCain et al., 1992; Lundh et al., 1992; Nitzan et al., 1991; Tarro, 1991; Linde et al., 1995).

The objective of group comparison studies is to provide evidence on which a future choice of a therapeutic intervention in a particular patient can be based. The choice of the intervention is based on the proven efficacy of the intervention, which is related to pre- and post-therapeutic differences in outcome variables of the group comparison study (Stegenga et al., 1993b). However, the impact of these differences depend on the study design and the sample used, i.e. on methodological issues.

A major priority of proper therapeutic outcome studies is to distinguish between specific and non-specific effects of the intervention. In order to do so, the outcome of the therapy group should be compared with a non-therapy group under otherwise comparable conditions. Further, to
prove a causal relationship between the effects found and the therapeutic intervention, randomized assignment from a homogeneous sample to different study groups is necessary. Moreover, to avoid bias, the observers, should not be aware of the study group the patients were assigned to. A true experimental study design fulfills these criteria and is generally regarded as the proper research method for therapeutic outcome assessment. This study was undertaken to analyze whether the claimed effects of arthroscopic surgery, arthrocentesis and physical therapy with regard to temporomandibular joint disorders could be substantiated by means of a proper methodology (Levine et al, 1986).

Methods

A Medline search of the international literature published between 1966 and 1997 was carried out with the following medical subject headings: temporomandibular joint diseases, dislocations, treatment outcome and time factors. Additional hand search followed, based on the available references of the papers found. The recorded papers subsequently were analyzed according to 1. method of investigation, 2. therapeutic intervention studied, 3. therapeutic outcome variables used, and 4. claimed effectiveness of the intervention.

Method of investigation

Three different types of study designs were distinguished i.e., pre-experimental, quasi-experimental and true-experimental designs. As pre-experimental designs, pre- and post intervention comparison of one group of patients and post intervention comparison of one group and a non-treated group were accepted. As quasi-experimental designs, group comparison studies and time series both without randomization were accepted. As true experimental designs group comparison studies and time series were accepted in which the subjects were randomly assigned to both groups. For power analyses, the total number of subjects and the number of subjects with a permanent temporomandibular joint disc displacement under study were separately recorded.

Therapeutic intervention

Four different interventions were distinguished i.e., arthroscopic surgery (AS), arthrocentesis (AC), physical therapy (PT), or combinations of these interventions and no treatment or placebo treatment (NT). For arthroscopic surgery, the use of an arthroscope was the only inclusion criterium. Criteria for arthrocentesis were: 1. joint lavage 2. hydraulic
pumping or 3. arthroscopic lavage. Criteria for physical therapy were: continuous passive motion, manipulation, exercise therapy, massage and physical interventions such as ultrasound therapy, short wave diathermy, or transcutaneous electrical nerve stimulation (TENS). Splint therapy, as a physical intervention, was recorded separately. Exclusion criteria were open joint surgery, eminectomy or any other surgical technique used to improve functioning of the temporomandibular joint.

**Outcome variables**
As outcome variables maximal mouth opening, pain intensity and mandibular function impairment were included. Maximal mouth opening had to be measured with a millimeter ruler as the interincisal distance on maximal mouth opening. Pain intensity had to be recorded on a visual analogue scale of 100 mm (VAS). Mandibular function impairment had to be assessed by means of a validated mandibular function impairment questionnaire (MFIQ) (Dworkin et al., 1990; Dworkin et al., 1988; Stegenga et al., 1993; Scott, 1976; McCormack et al., 1988).

**Effectiveness of the intervention**
Effectiveness of the interventions as claimed by the authors for within-group-effects and for between-group-effects was recorded. The papers were analyzed according to the above mentioned criteria by two observers independently. Overall agreement between the observers (TJBK, PUD) and agreement for each of the reviewing criteria was calculated as Cohens Kappa values.

**Results**
Twenty-four publications, fulfilling the search criteria, were found out of 53 dentistry or oral and maxillo-facial surgery journals between 1986 to 1997 (Table 1). Overall agreement concerning the reviewing criteria between the observers was 0.82 (p ≤ 0.001). Agreement varied between 0.41 for quasi-experimental designs and 1.0 for the use of a validated MFIQ and effectiveness in these papers, respectively (Table 2.1).

**Method of investigation**
Six studies with a true experimental design were found (Petersson et al., 1994; Gray et al., 1994; Lundh et al., 1992; Linde et al., 1995; Fridrich et al., 1996; Stegenga et al., 1993b). Six studies were judged as quasi-experimental design, however the observers did not agree whether the studies of Clark et al. (1991), Davis et al. (1991) and Nitzan et al. (1990) should be judged as quasi or pre-experimental. The other 12
studies were judged as pre-experimental designs. The number of subjects in the studies with a true experimental design ranged from 21 (Stegenga et al., 1993) to 139 (Gray et al., 1994).

**Therapeutic intervention**

Each of the six true experimental designs compared a different set of therapeutic interventions. Fridrich et al. (1996) compared the outcome of arthroscopy and arthrocentesis (Fridrich et al., 1996). Linde et al. (1995) evaluated the effects of a flat occlusal splint and TENS. Gray et al. (1994) evaluated the effectiveness of a treatment based on short-wave diathermy, a pulsed short-wave diathermy, ultrasound therapy and laser therapy. Petersson et al. (1994) evaluated the effectiveness of arthrography with or without immediate lavage. Stegenga et al. (1993) evaluated the effects of arthroscopic surgery followed by post-operative physical therapy and physical therapy alone. Lundh et al. (1992) evaluated the effects of a flat occlusal splint versus non treatment.

**Outcome variables**

Maximal mouth opening, was used as an outcome variable in all of the papers, except in Lundh et al. (1992) and Sanders et al. (1987). Nine out of 24 papers used a VAS of 0 - 100 mm for pain intensity assessment, a 2 to 32 points scale for the assessment of pain was used by Murakami et al. (1995), while Dimitroulis et al. (1995) used a VAS of 0 to 150 mm. The validated MFIQ was used only by Stegenga et al. (1993). Other instruments used for function impairment are the Helkimo index (Bertolami et al. 1983), subjective ability to chew (Hoffmann et al. 1994) and non-validated jaw function scores.

**Effectiveness of the intervention**

All 24 authors claimed effectiveness of the therapeutic intervention(s). Eleven papers compared different sets of therapeutic interventions but none of them found a statistical significant difference between the effects of different interventions. Linde et al. (1995) found that the occlusal splint resulted in more pain relief compared to transcutaneous electro nerve stimulation. Gray et al. (1994) found a shorter period of improvement after short-wave diathermy and megapulse than following ultrasound and laser. Using a pre-experimental design, Kuwahara et al. (1994) claimed better results referring to joint noises after open joint surgery than after arthrocentesis.
Table 2.1 Results of the analyses of 24 scientific papers concerning the therapeutic outcome of arthroscopic surgery, arthrocentesis and physical therapy. Agreement is expressed as Cohen's Kappa. Overall $K = 0.82$ $p < 0.001$

<table>
<thead>
<tr>
<th>Method of investigation</th>
<th>Therapeutic intervention</th>
<th>Therapeutic outcome</th>
<th>Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author</strong></td>
<td><strong>Year</strong></td>
<td>PED</td>
<td>QED</td>
</tr>
<tr>
<td>Fridrich K L et al.</td>
<td>1996</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Murakami K et al.</td>
<td>1996</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Murakami K et al.</td>
<td>1995</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Linde, C et al.</td>
<td>1995</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dimitroulis G et al.</td>
<td>1995</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Gray RJM et al.</td>
<td>1994</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hoffman DC etc al.</td>
<td>1994</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Kuwahara T et al.</td>
<td>1994</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Petersson A et al.</td>
<td>1994</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>McCarty WL et al.</td>
<td>1993</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Stegenga B et al.</td>
<td>1993</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>McCain JP et al.</td>
<td>1992</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Lund H et al.</td>
<td>1992</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tarro AW et al.</td>
<td>1991</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Nitzan DW et al.</td>
<td>1991</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Clark G T et al.</td>
<td>1991</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Davis C et al.</td>
<td>1991</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Nitzan DW et al.</td>
<td>1990</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Israel H A et al.</td>
<td>1989</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Moses J Y et al.</td>
<td>1989</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Montgomery et al.</td>
<td>1989</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Indresano A T et al.</td>
<td>1989</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sanders B et al.</td>
<td>1987</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Holmlund A et al.</td>
<td>1986</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Kappa</strong></td>
<td></td>
<td>0.67</td>
<td>0.41</td>
</tr>
</tbody>
</table>

PED=pre-experimental design; QED= quasi-experimental design; TED=true experimetal design; N=number of subjects; PT=physical therapy; AC=arthrocentesis; AS=arthroscopic surgery; NT=non-therapy/placebo therapy; splint=splint-therapy; MMO=maximal mouth opening; PAIN=pain on a 100 mm visual analogue scale; MFIQ=mandibular function impairment questionnaire; Effectiveness within/between groups.
Discussion

Only a limited number of scientific papers could be found in which the effects of arthroscopic surgery, arthrocentesis and/or physical therapy on maximal mouth opening, pain and mandibular function impairment were evaluated.

The overall between observer agreement of $K = 0.82$ is sufficient (Altman, 1991). Agreement was moderate ($K = 0.63$) to complete ($K = 1.0$) for all criteria except for the criterion quasi-experimental designs ($K = 0.41$).

Twelve studies used pre-experimental designs. Because of the lack of a control group in this design, it is not possible to distinguish between specific and non-specific effects of the intervention. Therefore, causality between the effects found and the therapeutic intervention can not be assessed.

Six studies used a quasi-experimental design. However random assignment of patients was not performed in these studies and therefore pre-existing group differences were not excluded.

Although 6 studies used a true experimental design, the effects of arthroscopic surgery, arthrocentesis and physical therapy did not differ with regard to maximal mouth opening, pain intensity or function impairment. None of the studies used blinded designs. To avoid surgeon-observer bias at least the use of an independent observer would have been an alternative. Only Gray et al. (1994) and Stegenga et al. (1993) used an independent observer. However Gray et al. (1994) compared non-invasive interventions, so no conclusions can be drawn regarding the effectiveness of a non-invasive intervention versus arthroscopic surgery or arthrocentesis. In the study of Stegenga et al. (1993) the number of subjects was insufficient to draw firm conclusions concerning efficacy (Stegenga et al., 1993). Comparing ‘blindly’ a non-invasive and an invasive intervention is difficult because of scars caused by invasive intervention. A simple solution would be to provide all patients with a pre-auricular bandage during the assessment regardless the intervention given.

Pooling the effects found in the papers using a true experimental design is not possible because different sets of interventions were compared. However, regardless of the nature of the intervention i.e. therapeutic or placebo, in all groups with the clinical diagnosis of permanent
temporomandibular joint disc displacement, maximum range of mouth opening had improved and pain intensity had decreased. Although Murakami et al. (1995) observed greater improvement on maximal mouth opening after arthroscopic surgery than other authors did, his clinical study did not randomly assign therapeutic interventions. Thus, selection bias could well be the reason for the reported cumulative success rate.

One study evaluated the effects of electro-physiotherapy and placebo treatment on pain free mouth opening and maximal mouth opening (Gray et al., 1994). We calculated the actual effect of electro-physiotherapy by subtracting maximal mouth opening achieved with electro-physiotherapy and maximal mouth opening achieved with placebo treatment. The placebo effects accounted for 53.9% the outcome of electro-physiotherapy. After a trial period of three months the placebo effects vanished, while the beneficial effects of electro-physiotherapy sustained (Gray et al., 1994). The actual effect on both the pain free mouth opening and the maximal mouth opening due to electro-physiotherapy was limited to 5.8 mm.

The assessment of pain intensity is usually performed with a 100 mm VAS, with ‘no pain’ and ‘worst pain imaginable’ as extremes. However, other types of VAS measures have been used, making the results of pain assessments incomparable.

The MFIQ, only used by Stegenga et al. (1993), is a valid and reliable instrument to measure mandibular function impairment. Mandibular function impairment was further measured as subjective ability to chew, jaw function scores, and the Helkimo index. However, reliability and validity of the subjective ability to chew and jaw function scores have never been published. The Helkimo index is primarily an epidemiologic instrument and was not designed to evaluate therapeutic effects. The latter application of this index has never been validated.

It is striking that none of the reviewed scientific papers have reported measurement error of the procedures used. To decide whether a change in an outcome variable is in fact real change, knowledge about the borders of precision of the measurement procedure is mandatory (Mitchell, 1979). Borders of precision can be expressed as the standard error of measurement, 95% confidence limits and the smallest detectable difference (Cronbach et al., 1972; Kropmans et al., 1995). The smallest detectable difference of a measurement procedure is the smallest statistical significant difference in outcome, which can be detected by a measurement procedure. It is possible that the differences reported in the
analyzed papers are due to influences related to the measurement procedures, rather than therapeutic effects. Based on empiricism, effect differences between arthroscopic surgery, arthrocentesis and physical therapy are expected to be small. This makes the borders of precision of the outcome variables even more important in order to detect the effects and the differences in effects between the interventions. However, the borders of precision of measuring the outcome variables for temporomandibular joint disorders are unknown. Thus, reliability studies in which these borders are analysed for maximal mouth opening, pain and function impairment measurement are necessary. The statistical analyses of a difference in group effects due to a therapeutic intervention is of importance, moreover the magnitude of change with reference to some standard - such as ‘normal’ functioning - should be an additional measure for success, since it relates to ‘clinical significance’ (Jacobson et al., 1984). For example, a statistically significant improvement in maximal mouth opening in permanent temporomandibular joint disc displacement of 2 mm may have very little clinical impact. So, an additional question should be how much change is necessary before it can be regarded as ‘clinically significant’? The smallest detectable difference is a measure that detects clinically relevant as well as statistically significant change in maximal mouth opening, pain intensity, and mandibular function impairment assessment in permanent temporomandibular joint disc displacement patients, analogous to applications in other fields (Kropmans et al., 1995; Evans et al., 1981; Mitchell, 1979; Barnett and Mathisen, 1997).

Instability and inconsistency in outcome measurement of each separate variable has been reported previously (van der Kloot et al., 1995; Price et al., 1983; Mezitis et al., 1989; Agerberg, 1974; Dijkstra, 1995). However, since pain intensity, maximal mouth opening and mandibular function impairment are interrelated, instability in measurement results of each separate outcome variable can be expected (Stegenga et al., 1993; Stegenga et al., 1989; de Bont and Stegenga, 1993; Stegenga et al., 1993a). The amount of variation caused by this relationship is unknown. It can be expected that this interrelationship may increase measurement error in pain, maximal mouth opening, and function impairment measurement.

We conclude that no differences in effects on maximal mouth opening and pain intensity or mandibular function impairment were found.
between arthroscopic surgery, arthrocentesis and physical therapy. Methodological sound outcome studies reporting the smallest detectable difference in outcome variables and evaluating the effects of arthroscopic surgery, arthrocentesis and physical therapy are needed. We recommend reliability studies that focus on determining the smallest detectable difference in maximal mouth opening, pain intensity and mandibular function impairment in patients with temporomandibular joint disc displacement.
References


