Fatal attraction: chemokines and rheumatoid arthritis
Haringman, J.J.

Citation for published version (APA):
Chapter 4

THE RELIABILITY OF COMPUTERIZED IMAGE ANALYSIS FOR THE EVALUATION OF SERIAL SYNOVIAL BIOPSIES IN RHEUMATOID ARTHRITIS RANDOMIZED CONTROLLED TRIALS

Jasper J. Haringman¹, Marjolein Vinkenoog¹, Danielle M. Gerlag¹, Tom J.M. Smeets¹, Aeilko H. Zwinderman², Paul P. Tak¹.

¹Division of Clinical Immunology and Rheumatology, Academic Medical Center/University of Amsterdam, The Netherlands. ²Department of Clinical Epidemiology and Biostatistics, Academic Medical Center/University of Amsterdam, The Netherlands.

Arthritis Research and Therapies 2005, 7:R862-R867
Analysis of biomarkers in synovial tissue is increasingly used in the evaluation of new targeted therapies in rheumatoid arthritis (RA) patients. This study determined the intra- and interrater reliability of digital image analysis (DIA) for the analysis of synovial biopsies of rheumatoid arthritis patients participating in clinical trials.

Arthroscopic synovial biopsies were obtained before and after treatment in 19 RA patients who participated in a randomized controlled trial with prednisolone. Immunohistochemistry was used to detect CD3+ T-cells, CD38+ plasma cells and CD68+ macrophages. The mean change of positive cells per mm² for each marker was determined by different operators and on different time points using DIA. Nonparametric tests were used to determine differences between the observers and assessments and to determine changes after treatment. The intraclass-correlations (ICCs) were calculated to determine the intra- and interrater reliability. Intrarater ICCs showed good reliability for measuring changes in T-lymphocytes (R=0.87), plasma cells (R=0.62) and macrophages (R=0.73). Analysis by Bland and Altman plots showed no systemic differences between the measurements. The smallest detectable changes were calculated and their discriminatory power showed good response in the prednisolone group compared to the placebo group. Similarly, interrater ICCs also revealed good reliability for measuring T-lymphocytes (R=0.68), plasma cells (R=0.69) and macrophages (R=0.72).

All measurements identified the same cell types changing significantly in the treated patients compared to the placebo group. The measurement of the change in total positive cell numbers in synovial tissue can be reproducibly determined for various cell types by digital image analysis in RA clinical trials.

**Key words:** arthroscopy, digital image analysis, rheumatoid arthritis, synovial tissue

**INTRODUCTION**

Rheumatoid arthritis (RA) is characterized by chronic and symmetrical inflammation of synovial joints (1:2). Although the etiology is still unknown, RA is thought of as an autoimmune disease with the synovial tissue (ST) being its primary target. The microscopic appearances of RA ST includes marked intimal lining layer hyperplasia due to increased numbers of fibroblast-like synoviocytes (FLS) and intimal macrophages, and the accumulation of macrophages, T-cells, B-cells, plasma cells, dendritic cells, mast cells, natural killer cells and neutrophils in the synovial sublining layer (3). Developments in synovial biopsy techniques, especially arthroscopy, have resulted in easier access to human ST. ST can now be selected from many sites within large and small joints, even in the earliest phases of disease, enhancing studies on etiology, prognosis and response to treatment (4).

Analysis of biomarkers in ST is increasingly used in the evaluation of new targeted therapies in rheumatoid arthritis patients (5). Numerous studies suggested consistent associations between the rapidity and magnitude of both the clinical and immunohistologic responses. It was shown that within the ST especially the number of infiltrating sublining macrophages can be used as a biomarker associated with clinical efficacy in relatively small studies of short duration (6:7). Therefore, the change in synovial sublining macrophages might be used as a biomarker for the evaluation of novel antirheumatic therapies. In addition to screening for possible efficacy, this approach provides insight into the mechanism of action of treatment. Within this setting reliable and validated methods for studying the ST are pivotal. The use of computerized or digital image analysis has greatly facilitated the analysis of ST. The major advantage of digital image analysis is the standardization of image acquisition and processing.
minimizing variance, and the ability to quantify the actual stained area together with staining intensity in a time efficient way (8:9). This allows the analysis of large numbers of stained sections. Strong correlations were observed between digital image analysis, semi-quantitative scoring and manual counting for the analysis of ST cellular markers, cytokines and adhesion molecules (10;11). Although the reproducibility of measuring cytokine and cell adhesion molecule staining by digital image analysis was reported to be within 10% (8), no formal studies investigating intra- and interrater variability were as yet reported. Therefore, we designed a study to determine the intra- and interrater reliability of this approach for the analysis of synovial biopsies of RA patients participating in clinical trials.

MATERIALS AND METHODS

Patients and samples. Arthroscopic synovial biopsies were obtained before and 2 weeks after treatment in 19 patients who participated in a double blind, placebo-controlled single center study with prednisolone, as reported earlier (6). All patients included had RA according to the 1987 criteria of the American College of Rheumatology (ACR)(12) and were using stable disease-modifying antirheumatic drugs (DMARDs) (methotrexate, sulphasalazine, hydroxychloroquine, leflunomide, or a combination of these) for at least 28 days prior to inclusion in the study. Ten of the 19 patients received prednisolone, 9 received placebo treatment. The needle arthroscopy of an actively inflamed joint (knee, ankle, or wrist) was performed under local anesthesia in all patients before treatment and in the same joint after treatment. The procedures for needle arthroscopy were performed as described previously in detail (13:14). During each procedure biopsies were taken from 6 or more sites throughout the joint to minimize sampling error. (15;16). These specimens were directly collected en bloc in a mold embedded in Tissue Tek OCT (Miles diagnostics, Elkhart. IN) and subsequently snap frozen by immersion in methylbutane (-80°C). The frozen blocs were stored in liquid nitrogen until processed. The study was approved by the Medical Ethics Committee of the Academic Medical Center, Amsterdam, the Netherlands, and all patients signed informed consent before start of the study.

Immunohistochemical analysis. Of each tissue sample, consisting of 6 different biopsy samples, serial sections were cut with a cryostat (5μm) and stained with the following antibodies; anti-CD68 (EMB11; Dako, Glostrup, Denmark), anti-CD38 (HB-7; Becton Dickinson) and anti-CD3 (SK7; Becton Dickinson) to analyze the major cell populations in the synovium. Sections with non-assessable tissue, defined by the absence of an intimal lining layer, were not analyzed. For control sections, the primary antibodies were omitted or irrelevant antibodies were applied. Staining for cellular markers was performed using a 3 step immunoperoxidase method as previously described (17).

Digital image analysis. After immunohistochemical staining, all coded sections were randomly analyzed by computer-assisted image analysis (Figure 1). For all markers, 18 high-power fields were analyzed. The images of the high-power fields were analyzed using the Qwin analysis system (Leica, Cambridge, UK), as described previously in detail (10:11). For the determination of the intrarater reliability one observer performed the acquisition and analysis twice with an interval of 4 weeks in between (OB1 t0 and OB1 t1, respectively). To determine the interrater reliability the acquisition of the images and the analysis was performed independently by 2 other experienced observers (OB2 and OB3).
All observers were blinded for the clinical data. For each measurement all observers independently set their own threshold levels regarding the detection of stained antigen, nuclear staining, and background staining. Subsequently after the analysis all observers independently calculated the mean change in the total number of positive cells per square mm (mm) ST for each marker.

Figure 1. *Acquisition and analysis of an immunohistochemical staining of CD3+ T-lymphocytes in synovial tissue using a digital image analysis system* [10]. Three different areas of each 6 high power field (hpf) are selected, which are representative for the whole tissue section. During the analysis, staining thresholds are set for the primary staining (i.e. CD3+ T-lymphocytes), nuclear staining and background staining. The output is generated in a spreadsheet as the total number of positive cells per square millimeter (mm) of synovial tissue. A full colour image of figure 1 is provided in the Appendix (Chapter 4, figure 1).

Statistical analysis. The nonparametric Friedman test and the Wilcoxon signed rank test were used to determine differences in the detection of the change in positive cell numbers per marker in the whole patient group, between the different observers and different assessments. The intra- and interrater reliability was quantified by means of the intraclass correlation coefficient of agreement (ICC) (18). In addition, scatter plots according to Bland and Altman (19) were made to depict the differences in the change in positive cells between the two measurements of one observer. The smallest detectable changes (SDC), representing the smallest change in scores that can be deemed as a “real” change (20), for the intra-observer variances was calculated and used to evaluate their discriminatory power. The nonparametric Mann-Whitney U test was used to determine whether each analysis detected differences in the change of positive cell numbers when the placebo group was compared with the prednisolone treated group.

RESULTS

Intrarater reliability. The mean number of CD3+ T-lymphocytes, CD38+ plasma cells and CD68+ sublining macrophages before and after intervention for 2 analyses by the same observer at different time points (OB1 t0 and OB1 t1) are depicted in Table 1. There were no
significant differences in the mean change in T-cells, plasma cells, and macrophages in the total population between the 2 measurements.

The overall correlations between the first and second analysis by the same observer were good. For the measurement of the change in CD3+ T-lymphocytes, CD38+ plasma cells and CD68+ macrophages the single rater and average of raters ICCs were calculated and depicted in Table 2. The relation between the two measurements of the observer are plotted in Figure 2. There was no systemic differences between the two measurements for each marker, but the variation was rather large. An analysis of the between-patient variances and the within-patient variances is provided in Table 2.

The SDC, averaged for the number of readings, for CD3+ lymphocytes was 182, for CD38+ plasma cells was 128 and for CD68+ macrophages was 306. When these estimates were used to discriminate the number of patients who responded to the treatment (i.e. had a reduction in positive cell numbers exceeding the SDC), for CD3+ lymphocytes 4 of the 10 patients in the prednisolone group responded compared to 0 of the 9 patients in the placebo group, for CD38+ plasma cells 4 of the 10 patients in the prednisolone group responded compared to 1 of the 9 patients in the placebo group and, for CD68+ macrophages 7 of the 10 patients in the prednisolone group responded compared to 0 of the 9 patients in the placebo group.

To determine whether the same observer identified the same differences in the synovial infiltrate after treatment at different time points, we determined whether there were significant differences in the change of T-cells, plasma cells, and macrophages between the placebo group and the prednisolone treated group for each measurement. At both time points there was on average a significant reduction in the number of CD3+ lymphocytes and CD68+ macrophages in the prednisolone treated patients compared to placebo (Table 1), while on average there were no significant changes in the number of CD38+ plasma cells.

Interrater reliability. The mean number of T-cells, plasma cells, and macrophages before and after intervention measured by the other 2 observers (OB2 and OB3) are also depicted in Table 1. There were no statistically significant differences in the mean change of positive cells between the analyses of the 3 observers (OB1, OB2 and OB3).

When the overall correlations between the analyses of the 3 observers were calculated the ICCs (single and average of raters) appeared to be good for CD3+ lymphocytes, CD38+ plasma cells and CD68+ macrophages (Table 2). An analysis of the between-patient variances and the within-patient variances is also provided in Table 2.

To determine whether all 3 observers identified the same differences in the synovial infiltrate after treatment, we determined whether there were significant differences in the change of T-cells, plasma cells and macrophages between the placebo group and the prednisolone treated group for each measurement. The measurements of all 3 observers showed on average a significant reduction in the number of CD3+ lymphocytes and CD68+ macrophages in the prednisolone treated patients compared to placebo (Table 1), while on average there were no significant changes in the number of CD38+ plasma cells.
Table 1. Mean number of CD3+ T-lymphocytes, CD38+ plasma cells and CD68+ sublining macrophages per mm² synovial tissue before and after intervention, measured by one observer at two different time points (OB1 t0 and OB1 t1) and by two other observers (OB2 and OB3) for placebo treated patients and prednisolone treated patients.

<table>
<thead>
<tr>
<th>Cell Type</th>
<th>Observer 1 t0</th>
<th>Observer 1 t1</th>
<th>Intra observer comparison</th>
<th>Observer 2</th>
<th>Observer 3 Inter observer comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Placebo</td>
<td>Prednisolone</td>
<td>P value</td>
<td>Placebo</td>
<td>Prednisolone</td>
</tr>
<tr>
<td>CD3+ T-lymphocytes (Mean ± SD)</td>
<td>Before: 192 (246)</td>
<td>558 (413)</td>
<td>$&lt;0.05$</td>
<td>Placebo: 179 (185)</td>
<td>285 (250)</td>
</tr>
<tr>
<td></td>
<td>After: 387 (391)</td>
<td>140 (150)</td>
<td></td>
<td>Placebo: 299 (356)</td>
<td>64 (62)</td>
</tr>
<tr>
<td>CD38+ Plasma cells (Mean ± SD)</td>
<td>Before: 56 (87)</td>
<td>99 (130)</td>
<td>Ns</td>
<td>Before: 246 (307)</td>
<td>397 (498)</td>
</tr>
<tr>
<td></td>
<td>After: 96 (127)</td>
<td>37 (57)</td>
<td></td>
<td>After: 119 (149)</td>
<td>42 (78)</td>
</tr>
<tr>
<td>CD68+ Macrophages (Mean ± SD)</td>
<td>Before: 804 (422)</td>
<td>973 (419)</td>
<td>$&lt;0.03$</td>
<td>Before: 937 (292)</td>
<td>1151 (254)</td>
</tr>
<tr>
<td></td>
<td>After: 972 (151)</td>
<td>553 (342)</td>
<td></td>
<td>After: 672 (686)</td>
<td>222 (278)</td>
</tr>
</tbody>
</table>

* Non-parametric, unpaired, Mann-Whitney U test for the comparison between placebo and prednisolone treatment.
** Non-parametric, paired, Wilcoxon signed rank test, for the comparison between OB1 t0 and OB1 t1 (intra observer comparison).
*** Non-parametric, paired, Friedman test, for the comparison between the three observers (OB1 t0, OB2 and OB3).
SD=Standard Deviation
Figure 2. Scatter plot of the mean change in the number of positive cells versus the difference in change of positive cells between the 2 measurements of observer 1 for A. CD3+ lymphocytes, B. CD38+ plasma cells and, C. CD68+ macrophages. The dotted line represents the mean ± 2×SD (standard deviation).
Table 2. Estimates of the variance components (between and within patients), and of the intraclass correlations (single rater and average of raters).

<table>
<thead>
<tr>
<th></th>
<th>INTRA-OBSERVER</th>
<th></th>
<th></th>
<th>INTER-OBSERVER</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>between patients</td>
<td>within patients</td>
<td>ICC</td>
<td>mean of 2 observations</td>
<td>between patients</td>
<td>within patients</td>
</tr>
<tr>
<td>CD3+ cells</td>
<td>11.59</td>
<td>1.73</td>
<td>0.87</td>
<td>0.93</td>
<td>10.13</td>
<td>4.85</td>
</tr>
<tr>
<td>CD38+ cells</td>
<td>1.35</td>
<td>0.82</td>
<td>0.62</td>
<td>0.77</td>
<td>8.65</td>
<td>3.83</td>
</tr>
<tr>
<td>CD68+ cells</td>
<td>20.32</td>
<td>7.35</td>
<td>0.73</td>
<td>0.85</td>
<td>18.92</td>
<td>7.19</td>
</tr>
</tbody>
</table>

DISCUSSION

This study investigated the intra- and interobserver reliability of the assessment of the change in ST T-cells, plasma cells, and macrophages quantified by digital image analysis. Tissue samples were obtained from RA patients participating in a monocenter placebo-controlled clinical trial with prednisolone. There were no significant differences in measurement of the mean change of T-cells, plasma cells, and macrophages between the 3 observers nor for different measurements by one observer. ICCs showed that there was good agreement between the measurements. All observers and all measurements identified on average a significant reduction in T-cells and macrophages, but not in plasma cells in the prednisolone group compared to placebo.

It can be anticipated that there will be an upsurge in randomized controlled trials investigating novel biologicals and small molecules for safety and efficacy. This clearly raises the need for sensitive, validated, and reliable measurements to screen for potential efficacy in an early phase of drug development. Clinical outcome measures have historically been used as primary endpoints, but their reliability may be limited in small proof of principle studies. For clinical measurements like the tender and swollen joint count ICCs have been reported varying between 0.15 and 0.85 for interrater variability and between 0.67 and 0.95 for intrarater variability (21). Radiographic measurements, by the use of conventional X-rays, shows good reliability in most studies, but are not useful in short term clinical trials (21). The use of magnetic resonance images is promising with acceptable interrater ICC for global synovitis scores and bone erosions, although the best scoring systems yet have to be determined (22).

In light of the need to screen various compounds for potential efficacy in small numbers of patients and because of recent technical developments, we believe that our thinking about clinical trials is about to change dramatically. Clinical studies during early phases of drug development will increasingly consist of small trials with a high density of biological data (23). Consistent with this notion, serial ST analysis with the evaluation of biomarkers was recently included in several randomized clinical trials of both disease-modifying anti-rheumatic drugs (DMARDs) and biologic agents (6:13:24-27). These and other studies showed consistent relationships between the magnitude of synovial changes and clinical response. Especially the change in infiltrating sublining macrophages was identified to be a potent and sensitive synovial biomarker (6:7).
ST can easily and safely be obtained due to the introduction of small-bore arthroscopes and the development of local and regional anesthesia protocols. Despite the heterogeneity in the ST within one joint it has been shown that representative measures of synovial inflammation can be obtained by examining a limited area of tissue (15:28:29). Previous work has also shown that digital image analysis is a sensitive, time efficient method for the quantification of both the number of stained cells and the staining intensity, with good correlations with both manual counting and semiquantitative scoring (10:11).

Although digital image analysis is described as a reliable and objective tool, little is known about variability and reliability. Measurement variation may be due to a limited number of causes using this approach. In our system the observer selects 3 different areas of each 6 high power fields from one slide which is composed of 6 biopsy samples from 6 different sites in the joint. This is done in such a way that a representative area is selected and this requires extensive training and experience with the histopathological morphology of ST. After scanning the representative high power fields, the images are analyzed by setting threshold values for the stained antigen, nuclear staining, and background staining (10). These thresholds are kept constant for all measurements with the same marker within a study, but could theoretically give rise to variation when set by different observers or by one observer at different time points. In the present study it was shown that these variables did not result in different outcomes. There were good ICCs when comparing the results of the 3 experienced observers or the results of the same observer on different time points. Analysis by Bland and Altman plots showed no systemic differences with regard to the intra-observer measurements and the SDCs showed good discriminatory power when applied to the treatment groups. In addition, all observers and all measurements identified the same cell types (T-cells and macrophages) as decreasing significantly in the active treatment group compared to placebo.

All measurements also identified a consistent trend toward reduced plasma cell numbers after corticosteroid treatment, which did not reach statistical significance, possibly due to the relative small number of patients. Although this method does show good agreement in detecting changes in histological markers, this does not necessarily mean that these results can be extrapolated to the expression of a given marker at a given time point, as used in cross sectional studies of ST. In addition, it remains to be seen if the same reliability holds true for the determination of changes in secreted proteins, like cytokines and chemokines.

**CONCLUSION**

In conclusion, the results of this study show the reliability of ST analysis using a digital image analysis system for the evaluation of serial synovial biopsy samples before and after treatment. This approach may be used for efficient quantification of synovial biomarkers in small proof of principle clinical trials.

**Abbreviations**

ACR = American College of Rheumatology; CI = Confidence Interval; CD = Cluster Domain; DIA = Digital Image Analysis; DMARD = Disease-Modifying Antirheumatic Drugs; ICC = Intra-Class Correlation; OB = Observer; RA = Rheumatoid Arthritis; ST = Synovial Tissue

**Competing Interests**
The authors have no competing interests to disclose.

**Author contributions**

JH was contributed to the experiments and was responsible for data analysis and interpretation and wrote the manuscript. 

MV, and TJMS were responsible for both the set –up and performance of the experiments. 

DMG was responsible for the inclusion of the patients and the collection of the materials and data.

- 53 -
AHZ coordinated and assisted in the statistical analysis of the data. PPT was responsible for the planning of the work and contributed to data analysis, interpretation and write-up.

Reference List


