DIFFERENT APPROACHES TO SYNOVIAL MEMBRANE VOLUME DETERMINATION BY MAGNETIC RESONANCE IMAGING: MANUAL VERSUS AUTOMATED SEGMENTATION

M. ØSTERGAARD
The Danish Research Center of Magnetic Resonance and Department of Rheumatology, Hvidovre Hospital, University of Copenhagen, Denmark

SUMMARY
Automated fast (5–20 min) synovial membrane volume determination by MRI, based on pre-set post-gadolinium-DTPA enhancement thresholds, was evaluated as a substitute for a time-consuming (45–120 min), previously validated, manual segmentation method. Twenty-nine knees [rheumatoid arthritis (RA) 13, osteoarthritis (OA) 16] and 17 RA wrists were examined. At enhancement thresholds between 30 and 60%, the automated volumes (Synx) were highly significantly correlated to manual volumes (Synman) (knees: rho=0.78–0.91, P<10–5 to <10–9; wrists: rho=0.87–0.95, P<10–4 to <10–6). The absolute values of the automated estimates were extremely dependent on the threshold chosen. At the optimal threshold of 45%, the median numerical difference from Synman was 7 ml (17%) in knees and 2 ml (25%) in wrists. At this threshold, the difference was not related to diagnosis, clinical inflammation or synovial membrane volume, e.g. no systematic errors were found. The inter-MRI variation, evaluated in three knees and three wrists, was higher than by manual segmentation, particularly due to sensitivity to malalignment artefacts. Examination of test objects proved the high accuracy of the general methodology for volume determinations (maximal error 6.3%). Preceded by the determination of reproducibility and the optimal threshold at the available MR unit, automated ‘threshold’ segmentation appears to be acceptable when changes rather than absolute values of synovial membrane volumes are most important, e.g. in clinical trials.

KEY WORDS: Magnetic resonance imaging, MRI, Gadolinium, Gadopentetate dimeglumine, Rheumatoid arthritis, Arthritis, Osteoarthritis, Synovitis, Knee, Wrist.

The synovial membrane is the site of inflammation in joints with rheumatoid arthritis (RA). On gross examination, the mass of the inflamed synovial membrane is markedly increased [1, 2]. Thus, the amount of synovial membrane is probably related to the disease activity and/or severity, and a reliable quantification may be useful in the assessment of RA patients and as an outcome measure of treatment.

A method for estimating volumes of synovial membrane and joint effusion by manual computer-assisted segmentation (outlining) based on visual analysis of pre- and post-gadolinium-DTPA MR images was introduced in previous studies [3–7]. The reproducibility, i.e. intra-observer, inter-observer and inter-MRI variation, has been described [3, 6]. Semi-automatic, but unvalidated, methods based on computerized counting of pixels fulfilling certain criteria have been described by other groups [8–11]. The synovial membrane volume is highly correlated to clinical signs of inflammation in cross-sectional studies [3, 7], as well as in longitudinal studies following the effect of various anti-inflammatory medications [5, 6, 9–11]. Furthermore, the pre-treatment synovial membrane volume was inversely correlated to the duration of clinical remission in RA knees following intra-articular methylprednisolone [6], suggesting that synovial membrane volume may have predictive value for treatment outcome. Thus, the synovial membrane volume may be useful as a marker, and perhaps a predictor, of treatment outcome in RA. The manual method is, however, time consuming (3–2 h per joint). Automation in some form is, therefore, necessary to make synovial membrane volume determination acceptable in larger clinical studies or in routine clinical practice.

The object of the present study was to evaluate whether a faster method, based on pre-set enhancement thresholds, may replace the time-consuming manual segmentation method for synovial membrane volume determination by MRI.

PATIENTS AND METHODS

Patients
Twenty-nine knees of 26 patients and 17 wrists of 16 patients were included. The underlying disease was RA in 13 knees and all wrists, and osteoarthritis (OA) in 16 knees. The patients fulfilled the ARA 1987 classification criteria of RA [12] and the ARA 1986 classification criteria of idiopathic OA of the knee [13], respectively. Patient data are summarized in Table I. The present report is divided into three protocols, as described below.

Protocol A. Twenty-six knees (11 with RA and 15 with OA) of 23 patients were included. Three patients (two RA, one OA) had both knees examined, with an interval of 10–12 months. MRI, clinical examination and standard blood tests (including erythro-
cyte sedimentation rate (ESR), serum C-reactive protein (s-CRP) and serum IgM-rheumatoid factor (IgM-RF) were performed once.

Protocol B. Three knees (two with RA and one with OA) of three patients and three wrists (all with RA) of three patients were included. MRI was performed twice within 2–5 days in order to assess reproducibility. Both examinations of the individual patient were carried out on the same MR unit. Clinical examination and standard blood tests were obtained on the days of MRI. To test the accuracy of the volume measurements, four differently shaped test objects, containing a known volume of water with an admixture of gadolinium-DTPA, were examined.

Protocol C. Fourteen wrists of 14 RA patients were included. MRI, clinical examination and standard blood tests were performed once.

MRI

The MR images were obtained using a 1.5 T Siemens MR unit (all patients in Protocols A and C, and one patient in Protocol B) or a 1.0 T Siemens Impact MR unit (five patients in Protocol B).

The knees examined were positioned neutrally rotated in a dedicated transmit–receive knee coil, with patients supine. Continuous transverse and sagittal T1-weighted spin-echo images (TR/TE/slice thickness = 500–750 ms/15 ms/5 mm) were obtained. At wrist examinations, the patient was positioned on the side with the opposite hand in front of the head, in the knee coil (Protocol C), or the patients were supine, with the arm along the side of the body and the wrist in a dedicated wrap-around receive-only wrist coil (Protocol B). Continuous coronal and transverse T1-weighted spin-echo MR images (TR/TE/slice thickness = 480–600 ms/15 ms/3 mm) were obtained. While the patient remained in the same position in the MR unit, 0.05 mmol gadolinium-DTPA/kg body weight was injected into a cubital vein through a cannula that was inserted before the examination. The T1-weighted spin-echo images were repeated. In both knees and wrists, the transverse images, which were used for volume estimations, were obtained first, starting 4–5 min after i.v. gadolinium-DTPA (acquisition time ~5 min). The recording of the sagittal (knees) or coronal (wrists) images was started 10 min after i.v. gadolinium-DTPA (acquisition time ~5 min). The transverse images were obtained with a field of view (FOV) of 180 mm (knees) or 100–120 mm (wrists) and a matrix size of 200–256 × 256. The test objects were placed in the knee coil, and transverse 5 mm T1-weighted spin-echo images were obtained. In two test objects, 3 mm transverse slices were also obtained.

Determination of synovial membrane and joint effusion volumes by 'manual' segmentation

The image-processing software package XPrime, installed on a Sun Sparc 10 computer (Unix), allowed outlining and calculation of the areas of regions of interest. The synovial membrane and the joint effusion of each transverse MRI slice were outlined manually, using a computer mouse, and the areas were calculated automatically. The outlining was done on post-gadolinium-DTPA images, based on visual analysis of pre- and post-contrast images as well as subtraction images (Fig. 1). The total volumes of synovial membrane (SynMan) and effusion (EffMan) were calculated by summation of the slices using the following formulae:

\[
\text{SynMan volume, manual method} = \Sigma \text{(Ar}_{\text{syn}}, \times ST)
\]

\[
\text{Joint effusion volume, manual method} = \Sigma \text{(Ar}_{\text{eff}}, \times ST)
\]

where ST is the slice thickness, while \(\text{Ar}_{\text{syn}}\) and \(\text{Ar}_{\text{eff}}\) represent the areas of synovial membrane and

| TABLE I | Patient data |
|---|---|---|
| | Protocol A (knees) | Protocol B (knees and wrists) | Protocol C (wrists) |
| | Total | RA | OA | RA | OA | Total |
| No. of knees | 26 | 11 | 15 | 5 | 1 | 14 |
| No. of patients | 23 | 9 | 14 | 1 | 1 | 14 |
| Age (yr) | 70 (24–85) | 65 (24–85) | 73 (60–85) | 62 (38–71) | 73 | 62 (20–76) |
| Disease duration (yr) | 8 (3–40) | 8 (3–30) | 10 (3–40) | 12 (5–24) | 11 | 4 (1–2) |
| IgM-RF (kIU/l) | 18/8 | 7/4 | 11/4 | 1/4 | 1/0 | 11/3 |
| ESR | 20 (1–74) | 44 (10–70) | 10 (1–30) | 24 (7–46) | 36 | 44 (3–62) |
| SynMan (ml) | 35 (9–95) | 56 (9–95) | 34 (5–79) | 10 (2–18) | 10 (2–18) |
| Syn50% (ml) | 46 (5–125) | 60 (5–125) | 41 (16–97) | 14 (5–27) |
| Syn60% (ml) | 39 (4–90) | 57 (4–84) | 37 (10–90) | 10 (4–25) |
| Syn70% (ml) | 28 (5–68) | 43 (5–68) | 25 (8–61) | 8 (2–21) |
| Syn80% (ml) | 10 (1–29) | 10 (1–26) | 9 (2–29) | 2 (1–7) |

Median values are given, with the range in parentheses. SynMan and Syn50% refer to synovial membrane volumes determined by the manual segmentation method and by automated 'threshold' segmentation, respectively, with a threshold of \(x\).
effusion in slice \( i \), respectively. Similarly, the test objects were outlined and their volumes determined by summation of slices.

**Determination of synovial membrane volumes by automated ‘threshold’ segmentation**

This method included two steps, both performed using the image-processing software package XPrime. Firstly, a rough manual outlining of the areas including synovial tissue was performed. Extra-articular enhancing tissues, particularly vessels, were excluded. Secondly, a segmentation algorithm was applied, by means of which XPrime showed and counted pixels (image points) fulfilling certain criteria (Fig. 1c–f).

The criteria chosen were as follows:

1. A relative post-gadolinium-DTPA signal intensity increase (enhancement) above a pre-set threshold. Five different enhancement thresholds were applied: (a) 30\% (the resulting volume was named ‘Syn\(_{30\%}\)’); (b) 40\% (Syn\(_{40\%}\)); (c) 45\% (Syn\(_{45\%}\)); (d) 50\% (Syn\(_{50\%}\)); (e) 60\% (Syn\(_{60\%}\)).

2. A post-gadolinium-DTPA absolute synovial signal intensity of >300, corresponding approximately to the mean pre-gadolinium-DTPA synovial membrane signal intensity minus 2 S.D. This criterion was included in all measurements to avoid noise from low-intensity pixels.

Based on the number of pixels, the volume of tissue fulfilling the criteria could be calculated by the following formula:

![Figure 1(a–d)](Legend opposite)
Synovial membrane volume, threshold method

\[ \text{Syn} \times \% = \frac{\text{number of pixels}}{\text{pixel size} \times \text{ST}} \]

where \( x \) indicates the enhancement threshold applied and \( \text{ST} \) is the slice thickness.

**Statistical methods**

Non-parametric methods were used. Analysis of statistical correlation was by the Spearman test of rank correlation. The Mann–Whitney test (two-sample rank sum test) was used to analyse differences between groups of patients. The ‘difference against mean plot’, proposed by Bland and Altman [14], was used for comparing the different ‘threshold volumes’ (\( \text{Syn} \times \% \)) with the ‘manual volumes’ (\( \text{SynMan} \)). The difference between \( \text{Syn} \times \% \) and \( \text{SynMan} \) (Diff \( \times \% = \text{Syn} \times \% - \text{SynMan} \)) was calculated, as were the 95%...
and the resulting volume determinations were based on visual analysis of these signal intensity differences. The synovial membrane, surrounded areas that still had low signal intensity, i.e. the joint fluid (Fig. 1).

The duration of volume determination by this procedure was 1–2 h, depending on the amount of synovial membrane and effusion to be outlined. The initial step in the ‘threshold’ method procedure was a rough outlining of an area that included all synovial tissue and excluded extra-articular enhancing pixels, particularly vessels. The number of 5 mm slices needed to cover the knee joint ranged from 11 to 25 (median 20). The rough outlining (Fig. 1b) could always be used for several adjacent images at a time, diminishing the number of outlinings needed. The number of different outlinings per knee ranged from 1 to 8 (median 3). The computer automatically calculated the areas of pixels fulfilling the criteria, and showed the result as a binary image (Fig. 1c–f).

The Spearman correlation coefficients between the ‘manual’ synovial volumes and the volumes obtained by automated segmentation were as follows: Syn$_{Man}$ vs Syn$_{30\%}$: rho = 0.86, P < 10$^{-7}$ for uncorrelated values; Syn$_{Man}$ vs Syn$_{40\%}$: rho = 0.91, P < 10$^{-9}$ for uncorrelated values; Syn$_{Man}$ vs Syn$_{45\%}$: rho = 0.84, P < 10$^{-5}$; Syn$_{Man}$ vs Syn$_{50\%}$: rho = 0.78, P < 10$^{-5}$; Syn$_{Man}$ vs Syn$_{60\%}$: rho = 0.56, P < 0.01 (Table II).

Thus, all ‘automated’ volumes were statistically significant at the 0.01 confidence limit compared to Syn$_{Man}$, while Syn$_{30\%}$ to Syn$_{60\%}$ were not significant.

### RESULTS

**Protocol A**

Before gadolinium-DTPA injection, the synovial membrane and the joint fluid could not be separated on the T1-weighted MR images, since both had low signal intensities. Following i.v. gadolinium-DTPA injection, a peripheral high-intensity rim, interpreted as the synovial membrane, surrounded areas that still had low signal intensity, i.e. the joint fluid (Fig. 1).

The ‘manual’ segmentation method, the outlining and showing the result as a binary image (Fig. 1c–f).

The Spearman correlation coefficients between the ‘manual’ synovial volumes and the volumes obtained by automated segmentation were as follows: Syn$_{Man}$ vs Syn$_{30\%}$: rho = 0.86, P < 10$^{-7}$ for uncorrelated values; Syn$_{Man}$ vs Syn$_{40\%}$: rho = 0.91, P < 10$^{-9}$ for uncorrelated values; Syn$_{Man}$ vs Syn$_{45\%}$: rho = 0.84, P < 10$^{-5}$; Syn$_{Man}$ vs Syn$_{50\%}$: rho = 0.78, P < 10$^{-5}$; Syn$_{Man}$ vs Syn$_{60\%}$: rho = 0.56, P < 0.01 (Table II).

Thus, all ‘automated’ volumes were statistically significant at the 0.01 confidence limit compared to Syn$_{Man}$, while Syn$_{30\%}$ to Syn$_{60\%}$ were not significant.

### TABLE II

<table>
<thead>
<tr>
<th>Statistical correlation (Spearman)</th>
<th>Absolute difference (ml)</th>
<th>Relative difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn$<em>{30%}$ vs Syn$</em>{Man}$</td>
<td>Median: 48</td>
<td>+99</td>
</tr>
<tr>
<td></td>
<td>Range: +4 to +120</td>
<td>+360</td>
</tr>
<tr>
<td>Syn$<em>{40%}$ vs Syn$</em>{Man}$</td>
<td>Median: 11</td>
<td>+26</td>
</tr>
<tr>
<td></td>
<td>Range: +12 to +38</td>
<td>+220</td>
</tr>
<tr>
<td>Syn$<em>{45%}$ vs Syn$</em>{Man}$</td>
<td>Median: 0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +29 to +11</td>
<td>+100</td>
</tr>
<tr>
<td>Syn$<em>{50%}$ vs Syn$</em>{Man}$</td>
<td>Median: 15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +50 to +7</td>
<td>+60</td>
</tr>
<tr>
<td>Syn$<em>{60%}$ vs Syn$</em>{Man}$</td>
<td>Median: 25.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +83 to +2</td>
<td>+40</td>
</tr>
</tbody>
</table>

Statistical correlation Absolute difference (ml) Relative difference (%)

Knees (Protocol A)

<table>
<thead>
<tr>
<th>Statistical correlation (Spearman)</th>
<th>Absolute difference (ml)</th>
<th>Relative difference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syn$<em>{30%}$ vs Syn$</em>{Man}$</td>
<td>Median: 48</td>
<td>+99</td>
</tr>
<tr>
<td></td>
<td>Range: +4 to +120</td>
<td>+360</td>
</tr>
<tr>
<td>Syn$<em>{40%}$ vs Syn$</em>{Man}$</td>
<td>Median: 11</td>
<td>+26</td>
</tr>
<tr>
<td></td>
<td>Range: +12 to +38</td>
<td>+220</td>
</tr>
<tr>
<td>Syn$<em>{45%}$ vs Syn$</em>{Man}$</td>
<td>Median: 0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +29 to +11</td>
<td>+100</td>
</tr>
<tr>
<td>Syn$<em>{50%}$ vs Syn$</em>{Man}$</td>
<td>Median: 15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +50 to +7</td>
<td>+60</td>
</tr>
<tr>
<td>Syn$<em>{60%}$ vs Syn$</em>{Man}$</td>
<td>Median: 25.5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Range: +83 to +2</td>
<td>+40</td>
</tr>
</tbody>
</table>

Syn$_{Man}$ and Syn$_{x\%}$ refer to synovial membrane volumes determined by the manual segmentation method and by automated ‘threshold’ segmentation, respectively, with a threshold of $x$. Diff$_{x\%}$ is the difference Syn$_{x\%}$ minus Syn$_{Man}$.

**Knees (Protocol A)**

Syn$_{Man}$ and Syn$_{x\%}$ refer to synovial membrane volumes determined by the manual segmentation method and by automated ‘threshold’ segmentation, respectively, with a threshold of $x$. Diff$_{x\%}$ is the difference Syn$_{x\%}$ minus Syn$_{Man}$.

### Fig. 2

The automated ‘threshold’ synovial volumes (Syn$_{x\%}$) plotted against the ‘manual’ volumes (Syn$_{Man}$). The full-drawn line represents absolute agreement. (a) Knees (Protocol B). (b) Wrists (Protocol C). It is seen that values of Syn$_{45\%}$ are closest to Syn$_{Man}$.
significantly correlated to the volumes determined by manual segmentation. However, the absolute difference between the manual and the various automated volumes differed markedly. As seen in Table II, the median absolute difference $|\text{Di}\% - \text{SynMan}|$ varied from +48 ml ($|\text{Di}\%|$) to -29 ml ($|\text{Di}\%|$). The agreement was best at a threshold of 45% (median difference = 0 ml). Correspondingly, the lowest median relative difference was found between $\text{Syn}_{45\%}$ and $\text{Syn}_{\text{Man}}$ (-1%) (Table II). The corresponding median numerical differences were 6.5 ml and 17% (Table II). In Fig. 2, the absolute values of the 'threshold' volumes are plotted against the volumes determined by manual outlining. The plot clearly confirms that $\text{Syn}_{45\%}$ is closest to $\text{Syn}_{\text{Man}}$.

In Fig. 3, 'difference against mean' plots, as proposed by Bland and Altman [14], of $\text{Syn}_{\%}$ vs $\text{Syn}_{\text{Man}}$ are given. With respect to $\text{Syn}_{45\%}$ (Fig. 3a), no general biases were observed, since the mean difference between $\text{Syn}_{\text{Man}}$ and $\text{Syn}_{45\%}$ was low.

**Fig. 3.**—Difference against mean plots [14] of automated volumes ($\text{Syn}_{\%}$), at the different thresholds, vs the manually determined volumes ($\text{Syn}_{\text{Man}}$). The average ($\text{Syn}_{\%} + \text{Syn}_{\text{Man}})/2$ is the abscissa, while the difference ($\text{Di}_{\%} = \text{Syn}_{\%} - \text{Syn}_{\text{Man}}$) is the ordinate. The mean difference $\overline{\text{Di}_{\%}}$ and the 95% confidence limits of the difference (mean difference $\pm 2$ s.d.) (- - - -) are marked. Except in 'Syn$_{45\%}$ vs Syn$_{\text{Man}}$' in knees (a) as well as wrists (b), systematic bias is found in all plots, i.e. at all other thresholds, in both knees (c−f) and wrists. See the text for further explanation.
TABLE III

Protocol B: variation at repeated MRI (inter-MRI variation)

<table>
<thead>
<tr>
<th></th>
<th>Knees</th>
<th>Wrists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Absolute difference (ml)</td>
<td>Relative difference (%)</td>
</tr>
<tr>
<td>SynMan</td>
<td>0, 4 and 6 ml</td>
<td>0, 9 and 12%</td>
</tr>
<tr>
<td>Syn30%</td>
<td>2, 2 and 4 ml</td>
<td>2, 3 and 8%</td>
</tr>
<tr>
<td>Syn40%</td>
<td>3, 6 and 10 ml</td>
<td>11, 11 and 12%</td>
</tr>
<tr>
<td>Syn45%</td>
<td>1, 4 and 12 ml</td>
<td>1, 18 and 24%</td>
</tr>
<tr>
<td>Syn50%</td>
<td>3, 8 and 11 ml</td>
<td>14, 21 and 37%</td>
</tr>
<tr>
<td>Syn60%</td>
<td>3, 3 and 10 ml</td>
<td>21, 33 and 77%</td>
</tr>
</tbody>
</table>

SynMan and Synx% refer to synovial membrane volumes determined by the manual segmentation method and by automated ‘threshold’ segmentation, respectively, with a threshold of x.

FIG. 4(a–d). (Continued opposite.)
...and the observations were spread homogeneously around the mean on the scatterplot. Nevertheless, the 95% confidence interval was still as wide as 39 ml [–21; +18]. Systematic differences between the methods were present in the other plots, particularly with respect to Syn30% and Syn60%, in which the mean difference \( d \) was numerically very high and the difference increased with increasing volumes (Fig. 3).

It was analysed whether the difference Syn\( x\% \) minus Syn\( \text{Man} \) (Diff\( x\% \)) was related to diagnosis, clinical appearance of the examined joint or the synovial membrane volume. Diff\( 45\% \) was not correlated to Syn\( \text{Man} \) (Spearman rho = –0.32; \( P = 0.11 \)), and no significant differences between RA and OA knees (Mann–Whitney, \( P = 0.18 \)) or between knees with and without tenderness/swelling (Mann–Whitney, \( P = 1.0 \)) were found. Thus, no systematic errors were found with respect to Syn\( 45\% \).

With respect to the other thresholds, statistically...
significant correlations between Diff x% and Syn Man were found: Diff 30% vs Syn Man: \( \text{rho} = 0.63, P < 0.001; \) Diff 40% vs Syn Man: \( \text{rho} = 0.52, P < 0.01; \) Diff 30% vs Syn Man: \( \text{rho} = 0.74, P < 0.001; \) Diff 60% vs Syn Man: \( \text{rho} = -0.94, P < 10^{-10}. \) In other words, the higher the synovial membrane volume, the higher was the difference between automated volumes and manually determined volumes. Thus, a systematic error, also illustrated in the Bland and Altman plots in Fig. 3, was found.

The automated volumes at different thresholds were highly significantly correlated. All mutual Spearman correlation coefficients between Syn 30%, Syn 40%, Syn 45%, and Syn 50% were >0.90 (\( P < 10^{-9} \)), except Syn 30% vs Syn 50% (\( \text{rho} = 0.82, P < 10^{-6} \)). The correlation coefficients with Syn 60% ranged from 0.64 to 0.88 (\( P < 0.01 \) to \( P < 10^{-8} \)).

**Protocol B**

The numerical relative difference between MRI-determined and true volumes of the test objects ranged from 4.2 to 6.3%, median 4.6%. No systematic differences were found between volumes determined from 3 and 5 mm slices.

Three knees and three wrists were examined by MRI twice within 2–5 days (Protocol B) (Table III). No clinical changes were observed between the MRI sessions. The variations of the volumes at the two successive MRIs evaluated by the same observer (inter-MRI variation) are given in Table III. It is seen that inter-MRI variations were generally higher by the automated method than by the manual method. The relative variation increased with increasing thresholds. Re-analysis of the images of the knee and the wrist with the highest Syn 45%, inter-MRI variations [12 ml (28%) and 4.3 ml (90%)] revealed minor malalignment artefacts contributing significantly to the measured volume at the examinations with the highest volumes. Synovial volumes in the three wrists were small (Syn Man = 3–6 ml), compared to the median Syn Man of 10 ml in Protocol C. This means that even small absolute inter-MRI variations result in considerable relative variations (Table III).

**Protocol C**

The synovial membrane volume in the 14 rheumatoid wrists ranged from 2 to 18 ml (median 10 ml) when determined by manual segmentation. The duration of volume measurements by this procedure was \( \frac{2}{3} - \frac{11}{12} \) h per wrist.

Synovial membrane volumes, determined by automated segmentation, were determined in 5–15 min. The number of 3 mm slices needed to cover the wrist joint ranged from 12 to 15 (median 13). The rough outlining (Fig. 4) could always be used for several adjacent slices at a time, diminishing the number of outlinings needed. The number of different outlinings per wrist ranged from 2 to 3 (median 2). The median volumes at the different enhancement thresholds were: Syn 30% = 21 ml, Syn 40% = 14 ml, Syn 45% = 10 ml, Syn 50% = 8 ml, Syn 60% = 2 ml, i.e. the chosen threshold markedly influenced the volumes measured (see Table I for details).

The Spearman correlation coefficients between ‘manual’ and ‘automated’ volumes were as follows: Syn Man vs Syn 30%: \( \text{rho} = 0.93, P < 10^{-6} \) for uncorrelated values; Syn Man vs Syn 40%: \( \text{rho} = 0.95, P < 10^{-6} \); Syn Man vs Syn 45%: \( \text{rho} = 0.92, P < 10^{-5} \); Syn Man vs Syn 50%: \( \text{rho} = 0.87, P < 10^{-4} \); Syn Man vs Syn 60%: \( \text{rho} = 0.50, P = 0.07 \) (Table II). Thus, except for Syn 60%, all the ‘automated’ volumes were statistically significantly correlated to the ‘manual’ volumes. However, the absolute difference between the ‘manual’ and the various automated volumes differed markedly. As seen in Table II, the median absolute difference Syn x% minus Syn Man (Diff x%) varied from +11 ml (Syn 30%) to –8 ml (Syn 60%). The agreement was best when a threshold of 45% was used (median difference = +1 ml). Correspondingly, the lowest median relative difference was found between Syn 45% and Syn Man (+15%) (Table II). The corresponding median numerical differences were 2 ml and 25% (Table II). In Fig. 2, the absolute values of the ‘threshold volumes’ are plotted against the volumes determined by manual segmentation. The plot illustrates that Syn 45% is closest to Syn Man.

‘Difference against mean’ plots [14] of Syn x% vs Syn Man were drawn. With respect to Syn 45%, no general bias was observed, since the mean difference between Syn Man and Syn 45% was low (+0.7 ml) and the observations were spread homogeneously around the mean on the scatterplot (Fig. 3b). The 95% confidence interval on the difference ranged from –4.9 to +6.3 ml. In the other plots, the difference Syn x% minus Syn Man, in almost all wrists, was either positive (Syn 30% and Syn 60%) or negative (Syn 40% and Syn 50%). Consequently, the mean difference \( d \) was numerically high in these plots, particularly with respect to Syn 30% and Syn 60%. Thus, the pattern was exactly as in the knees of Protocol A (Fig. 3c–f).

It was analysed whether the difference Syn x% minus Syn Man (Diff x%) was related to the clinical appearance of the examined joint or the synovial membrane volume. The difference between Syn 45% and Syn Man (Diff 45% = Syn 45% − Syn Man) was not correlated to Syn Man (Spearman \( \text{rho} = 0.05, P = 0.87 \)), and there was no significant difference between wrists with and without tenderness/swelling (Mann–Whitney, \( P = 0.81 \)). In other words, no systematic errors were found with respect to Syn 45%.

With respect to the other thresholds, statistically significant correlations between Diff x% and Syn Man, indicating systematic errors, were found between Diff 30% and Syn Man (\( \text{rho} = 0.64, P < 0.05 \)) and between Diff 60% and Syn Man (\( \text{rho} = -0.58, P < 0.05 \)). No significant correlations were found between Diff 40% and Syn Man (\( \text{rho} = 0.40, P = 0.14 \)) or between Diff 30% and Syn Man (\( \text{rho} = -0.11, P = 0.70 \)).

The automated volumes at different thresholds were highly significantly correlated. All mutual Spearman correlation coefficients between Syn 30%,
inflammatory activity. No general patterns in Di/C12845%

estimate synovial volumes in joints with low

'threshold'volumes and 'manual' volumes (Di/C128

brane volumes in highly inflamed joints, and under-

pronounced in less inflamed joints. If so, the

threshold method could overestimate synovial mem-

brane volume. Nevertheless, the 95% con-

fidence limits were rather wide:

f

The median numerical di/C128erence was 7 ml (17%) in knees

and 2 ml (25%) in wrists. However, such operations are carried out in a blood-

empty field provided by tourniquets, while MRI

measures vascularized, blood-filled and oedematous

tissue.

The main problem involved in delineating the

synovial membrane is generally assumed to be the
distinction from joint fluid [16,17], because the signal

intensity of the joint fluid also increases after
gadolinium-DTPA injection. However, the rate of

signal increase is much slower than in the synovium
[17–19], and a comparison of pre- and post-
aspiration volumes in knee joints indicated that mis-

interpretation of the effusion–synovium borderline is

not the major source of variation when the MR
images, as in this study, are obtained within 10–15

DISCUSSION

Two different approaches to synovial membrane
determination by MRI are discussed in the present
study. The first is a time-consuming but validated
segmentation method, which consists of manual com-

puter-assisted outlining, based on visual interpreta-

tion of pre- and post-gadolinium-DTPA MR images.

This method has been evaluated in previous studies
by our group [3–7]. The second approach involves
the use of much faster, but unvalidated, automated
segmentation methods, based on automatic comput-

erized calculation of areas fulfilling certain criteria
[8–10]. In this study, the main criterion was a post-
gadolinium-DTPA relative signal intensity increase
(enhancement) above a pre-set threshold of either 30,
40, 45, 50 or 60%. It was investigated whether the
fast, automated ‘threshold’ method may replace the
time-consuming ‘manual’ segmentation method, in

knees as well as wrists.

Highly significant correlations were found between
the manually determined synovial volumes (SynMan)
and the ‘threshold’ volumes (Synx%) when the
enhancement threshold was 30, 40, 45 or 50%, both
with respect to wrists and knees (Spearman
rho = 0.78–0.95, P from <10–9 to <10–4) (Table II).
A threshold of 40% gave the highest correlation
coefficients. The high correlation coefficients illustrate
that even though absolute volumes were highly
dependent on the thresholds (see below), the ranking
of the volumes was virtually independent of the
method applied, except when the threshold was so
high that only a few pixels were included (Syn40%).

The absolute values of synovial volume were very
different when different enhancement thresholds were
applied (Table I). Agreement with the manually
determined volumes was best when a threshold of
45% was chosen, both in knees and wrists (Table II).
The median difference between Syn45% and SynMan
was low, 0–1 ml in both knees and wrists, and the
median numerical difference was 7 ml (17%) in knees
and 2 ml (25%) in wrists. Nevertheless, the 95% con-
fidence limits were rather wide: ~±20 ml in knees
and ±5–6 ml in wrists, corresponding to ~50% of
the median synovial membrane volumes in the re-
spective joints.

One might fear that the difference between
‘threshold’ volumes and ‘manual’ volumes (Diffx)
was influenced by the inflammatory activity of the
joint, since the synovial enhancement may be less
pronounced in less inflamed joints. If so, the
threshold method could overestimate synovial mem-
brane volumes in highly inflamed joints, and under-
estimate synovial volumes in joints with low
inflammatory activity. No general patterns in Diff45%
(Syn45% – SynMan) were found: neither diagnosis,
clinical signs of inflammation nor the synovial

membrane volume were related to Diff45%. Thus, no
systematic errors were revealed with respect to
Syn45%. This is probably explained by the fact that
mainly the early synovial enhancement, determined
by dynamic imaging immediately after contrast
agent administration, is influenced by the synovial
inflammatory activity, while the correlation is
considerably weaker when static spin-echo images are
used, as in this study [15].

At other enhancement thresholds [30%, 40% (only
knees), 50% (only knees) and 60%], significant cor-
relations between Diffx% and SynMan were found.
This probably reflects the general under- and over-
estimation of synovial volumes at too high and
too low enhancement thresholds, respectively. The
absolute size of the error will logically be increased
with increased synovial volumes, e.g. with a larger
synovial volume there will be more synovial mem-
brane to miss if the enhancement threshold is too
high.

The optimal enhancement threshold (Protocols A
and C) was investigated on images obtained by the
1.5 T MR unit. The 1.0 T MR unit was only used in
Protocol B, i.e. for comparison of the reproducibility
of volume determinations by manual and automated
segmentation.

It should be emphasized that an enhancement
threshold of 45% may not be optimal on another
MRI unit since the post-gadolinium-DTPA enhance-
ment and, consequently, the optimal threshold, is
influenced by various factors, e.g. field strength, the
MRI sequence applied and the dose of gadolinium-
DTPA.

A main problem regarding the validation of
synovial membrane volume determination by MRI is
to find appropriate gold standards. The manual MRI
segmentation technique itself may represent such a
standard. Studies of test objects proved the high
accuracy of the general methodology for volume
determinations. No established references for synovial
membrane volumes are available. Postmortem studies
are not possible because the MRI technique requires
i.v. administration of a contrast agent. Even
arthroscopy or arthrotomy would not allow exact
quantification. Open surgical synovectomy, with care-
ful dissection of the synovium, may be a possibility.
However, such operations are carried out in a blood-
empty field provided by tourniquets, while MRI
measures vascularized, blood-filled and oedematous
tissue.

The main problem involved in delineating the
synovial membrane is generally assumed to be the
distinction from joint fluid [16,17], because the signal
intensity of the joint fluid also increases after
gadolinium-DTPA injection. However, the rate of
signal increase is much slower than in the synovium
[17–19], and a comparison of pre- and post-
aspiration volumes in knee joints indicated that mis-
interpretation of the effusion–synovium borderline is
not the major source of variation when the MR
images, as in this study, are obtained within 10–15
min after gadolinium-DTPA injection [4]. Thus, the most peripheral joint fluid may be included in the synovial membrane volume, but at least within the initial 10 min after gadolinium-DTPA injection other sources of variation, probably particularly coincidental variations in the outlining and partial volume artefacts, appear to be more important [4]. Similar wrist joint data do not exist. The anatomical structures are considerably smaller. Given the slice thickness of 3 mm, it must be expected that partial volume artefacts (volume-averaging effects) are of a relatively greater importance than in knees. Furthermore, it would be expected that the ‘automated’ volumes are the most sensitive to the gradual enhancement of peripheral joint fluid, since the enhancement threshold may be exceeded, while the contrast to the even higher signal intensity of the synovium may still be detected by visual analysis.

The study of pre- and post-aspiration volumes [4] also evaluated the accuracy of synovial membrane and joint effusion volume determination by the manual method. The difference between MRI-determined and syringe-determined volumes of aspirated joint fluid was 0–18%, median 7% (2 ml) of the pre-aspiration effusion volume. Synovial membrane volumes, determined before and after arthrocentesis, varied 0–17%, median 7.1% (3 ml) of pre-aspiration synovial volumes. It was concluded that effusion volumes and, in all probability, also synovial membrane volumes, can be determined by MRI with a maximal error of ~20% [4]. In another study, repeated MRI of six knees revealed a maximal intra-observer + inter-observer + inter-MRI variation below 30% [6]. Furthermore, manual segmentation by visual image analysis is theoretically favoured by the fact that the eye is a very sensitive discriminator of different tissues, because registration of signal intensities, structural analysis and anatomical knowledge are combined, while the automated methods hitherto applied only take signal intensities into account [8–11]. At present, we therefore consider the manual segmentation method the best reference available for evaluation of faster and more generally applicable methods for synovial volume estimations.

Several studies have found synovial membrane volumes to be highly correlated to clinical signs of inflammation, in cross-sectional studies [3, 7] as well as in longitudinal studies following the effect of various anti-inflammatory medications [5, 6, 9, 10]. Furthermore, the pre-treatment synovial membrane volume was inversely correlated to the duration of clinical remission in RA knees following intra-articular methylprednisolone [6]. Thus, the synovial membrane volume may be useful as a marker, and perhaps a predictor, of treatment outcome in RA, encouraging attempts to simplify the procedure and increase the applicability.

Approaches to automated synovial membrane segmentation and volume determination by MRI have so far all consisted of semi-automatic counting of pixels above a certain threshold [8–11]. Data on accuracy and reproducibility are not available. Waterton, Creamer and co-workers [8, 11] used pre-set relative post-gadolinium-DTPA enhancement thresholds, as in the present study. This method is fast, but since signal intensity changes are measured, it is mandatory that the patient does not move at all between pre- and post-contrast imaging, in order to avoid malalignment artefacts. Palmer, Polisson and co-workers [9, 10] obtained fat-suppressed T1-weighted images, and used an absolute signal intensity threshold on post-contrast images. The threshold was determined individually at each examination, based on evaluation of pre- and post-gadolinium-DTPA signal intensities in selected regions. Malalignment artefacts are avoided, but calculation of a threshold for each examination markedly increases the duration of the procedure. Determination of a suitable fixed absolute threshold for fat-suppressed post-contrast images may be a promising approach. However, it requires reliable and uniform fat suppression, since areas of unsuppressed fatty tissue will invalidate the method.

Synovial membrane visualization without the use of a contrast agent would be very attractive, due to reduced costs and the entirely non-invasive procedure. Peterfy et al. [20] reported depiction of hypertrophic synovial tissue by unenhanced pulsed saturation transfer and fat suppression techniques. The potential for synovial volume determination has not been evaluated.

The inter-MRI variation increased with increasing threshold. At the 45% threshold, it was higher than by manual segmentation. Even minimal malalignment artefacts contributed markedly to the variation, especially at high thresholds. This illustrates the high sensitivity of ‘threshold’ segmentation to even slight joint movement. Differences in synovial enhancement will naturally also influence the measured volumes, but in the present protocol (only six joints) this factor appeared less important.

In summary, examination of test objects proved the high accuracy of the general methodology for volume determinations. Automated ‘threshold’ segmentation can in 5–20 min (less than a fifth of the time spent by manual segmentation) provide estimates that are highly correlated to the manually determined synovial membrane volumes. By careful selection of thresholds, the automated estimates can come fairly close to the more reliable, but time-consuming manual volumes. However, the 95% confidence interval of the difference is wide. At the optimal enhancement threshold (45% in our setting), the difference between automated and manual volumes was not related to diagnosis, clinical inflammatory activity or synovial membrane volume, i.e. no systematic errors were found. Thus, it appears acceptable to use the threshold method when changes, rather than the absolute size, of the synovial membrane volume are most important, e.g. in clinical trials. It is, however, important beforehand to determine the reproducibility and optimal enhancement...
threshold at the available MR unit, for instance by comparison with manual segmentation. Finally, it must be remembered that the synovial membrane has a true volume, and even though various estimates may be useful in clinical trials etc., any variation from this true volume necessarily is the result of misregistration.

ACKNOWLEDGEMENTS

The Foundation of 17-12-81, the University of Copenhagen, the Danish Rheumatism Association, the Thomas & Elisabeth Frølund Nielsen Foundation, the Danish Medical Research Council, the Danish Hospital Foundation for Medical Research, Region of Copenhagen, the Faroe Islands and Greenland are acknowledged for financial support. Schering Diagnostika, Denmark, is acknowledged for providing the contrast agent. I would also like to thank Professor Ib Lorenzen and Professor Ole Henriksen for valuable scientific guidance, Michael Stoltenberg, Michael Hansen and nurse Brigitta Pedersen Zbinden for clinical management of the patients, and Poul Ring and Jens Arnth Jensen for developing the image-processing software package XPrime.

REFERENCES