Ultrasound imaging to tailor the treatment of acute shoulder pain: a randomised controlled trial in general practice

Ramon P G Ottenheijm,1 Jochen W L Cals,1 Bjorn Winkens,2 René E Weijers,3 Rob A de Bie,4 Geert-Jan Dinant1

ABSTRACT
Objective: To determine the clinical effectiveness of ultrasound tailored treatment in patients with acute subacromial disorders.
Design: Pragmatic randomised controlled trial.
Setting: Dutch general practice.
Participants: Patients aged 18–65 years with acute (duration <3 months) unilateral shoulder pain and no previous treatment, in whom the general practitioner suspected a subacromial disorder was enrolled.
Interventions: All patients underwent ultrasound imaging of the affected shoulder. Patients who were still symptomatic after a qualification period of 2 weeks with standard treatment were randomised to treatment tailored to ultrasound diagnosis (disclosure of the ultrasound diagnosis) or usual care (non-disclosure of the ultrasound diagnosis).
Primary outcome measure: Patient-perceived recovery using the Global Perceived Effect questionnaire at 1 year.
Results: 129 patients were included. 18 patients recovered during the 2-week qualification period, resulting in 111 randomised patients; 56 were allocated to ultrasound tailored treatment and 55 to usual care. After 1 year, no statistically significant differences in recovery were found between the ultrasound tailored treatment group (72.5% (37/51)) and the usual care group (60% (30/50), OR 2.24 (95% CI 0.72 to 6.89; p=0.16)). Also, healthcare use differed. Patients in the ultrasound tailored treatment group used less primary care visits and took fewer days off work.
Conclusions: This study has shown no clinically significant difference in the primary outcome measure between the ultrasound tailored treatment and usual care groups. Furthermore, there was no overall difference in healthcare resources used between groups. Although no formal cost data are included, one can only assume that the ultrasound examinations are additional costs for the intervention group, which cannot be justified in routine practice based on this trial. Based on this study, no change in current pragmatic guidelines to incorporate early ultrasound imaging can be recommended.

Strengths and limitations of this study
- This study was developed as a pragmatic trial to inform clinicians, guideline developers and policymakers to choose wisely between options for care.
- The trial does not give a conclusive answer to whether ultrasound tailored treatment improves outcome after 1 year, as our trial was underpowered. The limited size of this trial is a limitation that may have prevented the documentation of significant clinically important differences.

INTRODUCTION

General practitioners (GPs) are frequently consulted by patients with shoulder pain.1–3 Prognosis is rather poor with 40% not being recovered after 1 year4–6 and high recurrence rates.7,8 These findings suggest that shoulder pain frequently progresses to a chronic disorder.

Shoulder pain is a symptom, not a diagnosis. Subacromial disorders are the most common cause of shoulder pain seen by GPs.10 11 In general practice, accurate diagnosis of shoulder pain is difficult because findings from medical history and physical examination often poorly correlate with the underlying disorder.12–14 Therefore, British and Dutch guidelines for shoulder pain advise GPs to start treatment based on patients’ signs and symptoms rather than on the actual disorder.10 15 GPs experience the current diagnostic process as complex, and the use of diagnostic ultrasound as helpful in establishing a more accurate diagnosis.16–21
Accuracy studies showed that diagnostic ultrasound is accurate for evaluating subacromial disorders. Moreover, the full spectrum of subacromial disorders is observed in patients with shoulder pain presenting in general practice. For each of these disorders, evidence-based treatments are available. This implies that stratification of patients into diagnostic subgroups potentially allows for more tailored treatment than currently applied. For instance, in general practice, patients with low back pain, a stratified management approach in which prognostic screening and treatment targeting were combined, improved patient outcome.

In daily general practice, combining clinical information with ultrasound diagnosis is potentially helpful to tailor treatment to patients with shoulder pain. So far, no pragmatic trial has evaluated this test-treatment approach for shoulder pain in general practice. Pragmatic trials measure effectiveness in routine clinical practice and reflect variations between patients that occur in real clinical practice improving generalisability. We conducted a pragmatic, randomised controlled trial, the Maastricht Ultrasound shoulder pain Trial (MUST), to study clinical effectiveness of ultrasound tailored treatment in patients with acute subacromial disorders.

**METHODS**

**Design and participants**

The study design and rationale of MUST were described elsewhere. Patients were eligible if they had shoulder pain on abduction with painful arc; symptoms having lasted <3 months; no other episodes of shoulder pain in the previous 12 months and age between 18 and 65 years. Exclusion criteria were consultation or treatment for shoulder pain in the past 3 months; glenohumeral external rotation range of motion <45° as this is a reason to suspect a glenohumeral disorder like osteoarthritis or a frozen shoulder; history of fractures of the proximal humerus or acromion, dislocation and/or surgery of the affected shoulder; shoulder symptoms caused by rheumatic disease, suspected referred symptoms or extrinsic cause; history of depressive or anxiety disorders (negative prognostic factors) or pain catastrophising (irrational thought in believing pain is far worse than it actually is); inability to complete a questionnaire independently; unable to give informed consent (dementia or psychiatric disorders) and involved in disability or liability procedures. Initially, 21 GPs working in 11 practices in the Westelijke Mijnstreek, a region in the southern part of the Netherlands, recruited eligible patients. These GPs were asked to include sequential eligible patients within regular consultation hours. After 2 years, all 80 GPs in the aforementioned region were asked to recruit patients. All GPs received oral and written instructions.

All patients underwent ultrasound imaging of the affected shoulder by experienced musculoskeletal radiologists using a standardised protocol and criteria for pathology. Patients who were still symptomatic after a qualification period of 2 weeks with standard treatment were randomised to treatment tailored to ultrasound diagnosis (disclosure of the ultrasound diagnosis) or usual care (non-disclosure of the ultrasound diagnosis) (figure 1). The 2-week qualification period was used to perform diagnostic ultrasound and aimed to filter out patients with a favourable natural course. During this period, patients received treatment according to the shoulder pain guidelines of the Dutch College of General Practitioners; paracetamol or non-steroidal anti-inflammatory drugs (NSAIDs) in maximum dosage on a time contingent base, advice regarding activities of daily living, work, hobbies and sports.

Our reporting follows the Consolidated Standards of Reporting Trials (CONSORT) extension for pragmatic trials. The study was approved by the Medical Ethics Committee of the Maastricht University Medical Centre, and the trial was registered at the Netherlands Trial Register (NTR2403). All patients signed an informed consent.

**Interventions**

Experienced musculoskeletal radiologists of the Zuyderland Medical Centre in Sittard-Geleen performed all ultrasound examinations using a standardised protocol. The two study arms were treatment tailored to the ultrasound diagnosis or usual care (figure 1). The GP provided the allocated treatments.

**Ultrasound tailored treatment**

A key feature of this intervention was disclosure of the ultrasound diagnosis to the GP in order to tailor treatment. GPs treated patients according to the advised evidence-based, tailored treatment steps as presented in figure 1 and in detail published in the study protocol. Advised treatment modalities depending on the ultrasound diagnosis were subacromial corticosteroid injections in case of bursitis or calcific tendinitis, referral to a physiotherapist in case of tendinopathy or partial-thickness tendon tear and an orthopaedic surgeon in case of full-thickness tendon tears.

In case there were no abnormal ultrasound findings, usual care according to the guideline for shoulder pain was provided. In cases where multiple ultrasound diagnoses were present, the most relevant diagnosis was selected on the basis of the clinical findings. Motivated and within the recommendations made in the guideline for shoulder pain, GPs were allowed to deviate from the advised treatment steps.

**Usual care**

In the control group, the ultrasound diagnosis was not disclosed; therefore, usual care according to the guidelines for shoulder pain was applied. It consisted of a pragmatic, stepwise approach; a wait-and-see policy with...
advice and analgesia for another 2 weeks; corticosteroid injections and referral to a physiotherapist were advised as options in persisting cases, depending on the level of pain and functional limitations, respectively; referral to a hospital specialist was advised if conservative treatment failed.10

Randomisation and blinding
Based on a qualification assessment at 2 weeks, unrecov-
ered patients were randomly assigned by central block randomisation (blocks of 4) to one of the study arms after stratification for age (cut-off ≥50 years), using an online application developed at the centre for data and

Figure 1 Flow chart of the study. GP, general practitioner.
information management of Maastricht University. Recovery was measured by the Global Perceived Effect questionnaire (see outcomes). Neither the patient nor the GP could be blinded for allocated treatment. However, ultrasound diagnoses were only disclosed to GPs of those patients in the ultrasound tailored treatment group. Radiologists were not allowed to communicate with the patient about the ultrasound imaging results.

Outcomes
Follow-up was performed by postal questionnaires at baseline, 3, 6, 9 and 12 months. Primary outcome was patient-perceived recovery using the Global Perceived Effect questionnaire at 1 year. It consists of a one-item score concerning recovery following treatment, measured on a seven-point ordinal scale. Patients were considered recovered when they reported to be much improved or fully recovered. Secondary outcomes included experienced shoulder pain using the Shoulder Pain Score, performance of daily activities using the Shoulder Disability Questionnaire and health-related quality of life using the Euroqol five-item quality of life questionnaire (EQ-5D).

The SPS questionnaire consists of six pain symptom questions and a 10-point scale. The SPS has been proved to be a useful instrument for following the course of the disorder over time and gives an indication when a patient feels cured. The score can range from 7 to 28 with a higher score indicating more pain. The SDQ contains 16 questions and is a useful discriminative instrument, especially in the primary care setting. The SDQ score can range from 0 to 100 with a higher score indicating a more severe disability. The EQ-5D is one of the most used generic measures to quantify the health-related quality of life in participants with musculoskeletal disorders and consists of two sections. The first section comprises five questions regarding five dimensions of health. Calculation of the index score was performed according to the British recommendations and ranges from −1 to 1, higher scores indicating better quality of life.

At inclusion, patients of whom the GP doubted about pain catastrophising behaviour were given the Pain Catastrophizing Scale (PCS). The PCS is a 13-item self-report scale to measure pain catastrophising and the score can range from 0 to 52 with a higher score indicating a higher level of catastrophising. We made use of a cut-off value of 20 points. Participants were excluded by the research team if their score was ≥20.

To measure 1-year healthcare use, we obtained electronic patient records of all patients by contacting the GP and used the postal questionnaires. In case patients were referred to a physiotherapist, the number of visits was collected.

Sample size
We estimated a 20% improvement in recovery rate as the minimal clinical important difference, and therefore calculated our sample size on the ability to detect a difference in study arms of 20% (60% vs 80%) or more on recovery rate. Using these data, we estimated that we needed 81 patients per study arm to show a significant difference at the 5% level (two-sided) with 80% power (based on the test). Allowing for a 10% drop-out rate, and the expectation that the qualification period would filter out 20% of the patients we needed to include 226 patients in total.

Statistical analysis
Data were analysed on an intention-to-treat basis. In addition, a per-protocol analysis of complete cases for the primary outcome at 12 months was performed to estimate the impact on the effect of follow-up and protocol deviations. To handle single-item missing data, a sensitivity analysis based on best-case and worst case scenarios (highest and lowest scores) was performed. No multiple imputation was used if whole questionnaires were missing, since missing outcome data were dealt with using a likelihood-based approach, assuming missing at random.

For the primary outcome measure, analysis was performed using a logistic regression analysis with correction for the stratification variable age (cut-off at 50 years). In addition, a three-level logistic mixed model to correct and account for age (cut-off at 50 years) and variation at the level of GP practice, patient and repeated observations with data from four time points (3, 6, 9 and 12 months) was used. Secondary outcomes were analysed using a linear mixed model correcting and accounted for age and variation at the level of GP practice, patient and repeated observations with data from four time points (0, 3, 6 and 12 months). Shoulder-related healthcare resource during 1-year follow-up was analysed using a logistic regression analysis with correction for the stratification variable age (cut-off at 50 years).

Putative prognostic factors at baseline, as well as recruitment strategy, were added as covariates to the mixed model. Owing to the expected small number of non-recovered patients, this correction was only applied to the numerical outcomes. All analysis was performed in SPSS (V21). A p value of ≤0.05 was considered statistically significant.

RESULTS
From November 2010 to December 2013, 129 patients were included in 26 GP practices. The flow of patients through the study is presented in figure 2. Eighteen patients (14%) recovered during the 2-week qualification period, resulting in 111 randomised patients; 56 were allocated to ultrasound tailored treatment and 55 to usual care. Ten patients were lost to follow-up at
Baseline characteristics were similar in both randomised groups (table 1). Two patients were prompted by their GP to complete the Pain Catastrophizing Scale, and both were classified as not pain catastrophising.

Patient-perceived recovery

Although more patients in the ultrasound tailored treated group perceived to be recovered after 1 year according to the patients’ Global Perceived Effect assessment (72.5% (37/51) vs 60% (30/50), OR 1.86 (95% CI 0.79 to 4.36)), this difference was not significant (p=0.15). The intention-to-treat analysis of this primary outcome measure is presented in table 2. Logistic mixed model analysis adjusting for age (cut-off at 50 years) and variation at the level of the GP practice, patient and repeated observations with data from four time points (3, 6, 9 and 12 months) showed similar results (OR 2.24 (95% CI 0.72 to 6.89), p=0.16) for patient-perceived recovery after 1 year.

Secondary outcomes

After 1 year, the mean differences in the Shoulder Pain Score (1.7 points, 95% CI −3.9 to 0.5) and Shoulder Disability Questionnaire (6.9 points, 95% CI −19.9 to 6.1) were in favour of the ultrasound tailored treated group, but these differences were not significant (p=0.15 and 0.29, respectively). Also health-related quality of life, measured with the EQ-5D and EQ-5D Visual Analogue Scale, was not significantly different between groups (table 3).
Analyses adjusting for the putative prognostic factors did not alter the results. No significant differences were found in the proportion of patients referred to physiotherapy (59% of patients receiving ultrasound tailored treatment vs 64% of patients receiving usual care), their mean number of physiotherapy sessions (12.9 vs 12.4), corticosteroid injections or referrals to secondary care (39% vs 31% and 20% vs 13%). Shoulder-related healthcare use after 1 year between the two groups is summarised in table 4. No adverse events or side effects were reported.

Sensitivity analysis
Per-protocol analyses of the primary outcome measure and the sensitivity analyses with best-case and worst-case scenarios to handle single-item missing data of the secondary outcome measures produced similar effects as in the intention-to-treat analysis (table 5).

### Discussion

**Summary**
This study has shown no clinically significant difference in the primary outcome measure between the ultrasound tailored treatment and usual care groups. Furthermore, there was no overall difference in healthcare resources used between the groups. However, our trial does not give a conclusive answer to whether ultrasound tailored treatment improves outcome after 1 year, as our trial was under-enrolled. The limited size of this trial is a limitation that may have prevented the documentation of significant clinically important differences. Although no formal cost data are included, one can only assume that the ultrasound examinations are

### Table 1 Baseline characteristics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Ultrasound tailored treatment (n=56)</th>
<th>Usual care (n=55)</th>
<th>Recovered before randomisation (n=18)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>49.0 (9.9)</td>
<td>49.4 (10.9)</td>
<td>48.6 (12.4)</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>26 (46.4)</td>
<td>20 (36.4)</td>
<td>10 (55.6)</td>
</tr>
<tr>
<td><strong>Specific disease variables</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of pain in weeks (SD)</td>
<td>6.2 (3.8)</td>
<td>5.5 (3.5)</td>
<td>2.1 (1.8)</td>
</tr>
<tr>
<td>Acute onset, n (%)</td>
<td>24 (42.9)</td>
<td>28 (50.9)</td>
<td>13 (72.2)</td>
</tr>
<tr>
<td>Concomitant neck symptoms, n (%)</td>
<td>20 (35.7)</td>
<td>14 (25.5)</td>
<td>6 (33.3)</td>
</tr>
<tr>
<td>Dominant shoulder affected, n (%)</td>
<td>33 (58.9)</td>
<td>37 (67.3)</td>
<td>9 (50.0)</td>
</tr>
<tr>
<td><strong>Ultrasound findings, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcific tendonitis</td>
<td>26 (46.4)</td>
<td>28 (50.9)</td>
<td>11 (61.1)</td>
</tr>
<tr>
<td>Tendinopathy</td>
<td>16 (28.6)</td>
<td>16 (29.1)</td>
<td>5 (27.8)</td>
</tr>
<tr>
<td>Bursitis</td>
<td>14 (25.0)</td>
<td>10 (18.2)</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td>Partial-thickness tears</td>
<td>11 (19.6)</td>
<td>13 (23.6)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Full-thickness tears</td>
<td>0 (0.0)</td>
<td>3 (5.5)</td>
<td>1 (5.6)</td>
</tr>
<tr>
<td>Impingement</td>
<td>12 (21.4)</td>
<td>7 (12.7)</td>
<td>2 (11.1)</td>
</tr>
<tr>
<td><strong>Number of disorders, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No disorder</td>
<td>12 (21.4)</td>
<td>8 (14.5)</td>
<td>4 (22.2)</td>
</tr>
<tr>
<td>1 disorder</td>
<td>20 (35.7)</td>
<td>26 (47.3)</td>
<td>7 (38.9)</td>
</tr>
<tr>
<td>≥2 disorders</td>
<td>24 (42.9)</td>
<td>21 (38.2)</td>
<td>7 (38.9)</td>
</tr>
</tbody>
</table>

### Table 2 Primary outcome measure, global perceived effect

<table>
<thead>
<tr>
<th>Recovered</th>
<th>Ultrasound tailored treatment %</th>
<th>Usual care %</th>
<th>Between-group difference</th>
<th>OR* 95% CI</th>
<th>p Value</th>
<th>OR† 95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 months</td>
<td>41.5 (22/53)</td>
<td>32.1 (17/53)</td>
<td>9.4</td>
<td>1.52</td>
<td>0.67 to 3.38</td>
<td>0.30</td>
<td>2.18</td>
</tr>
<tr>
<td>6 months</td>
<td>46.8 (22/47)</td>
<td>44.7 (21/47)</td>
<td>2.1</td>
<td>1.10</td>
<td>0.47 to 2.48</td>
<td>0.82</td>
<td>1.15</td>
</tr>
<tr>
<td>9 months</td>
<td>53.1 (26/49)</td>
<td>60.4 (29/48)</td>
<td>−7.3</td>
<td>0.75</td>
<td>0.33 to 1.73</td>
<td>0.50</td>
<td>0.59</td>
</tr>
<tr>
<td>12 months</td>
<td>72.5 (37/51)</td>
<td>60.0 (30/50)</td>
<td>12.5</td>
<td>1.86</td>
<td>0.79 to 4.36</td>
<td>0.15</td>
<td>2.24</td>
</tr>
</tbody>
</table>

*Based on logistic regression analysis correcting for age (stratification variable, cut-off 50 years).
†Based on logistic mixed model analysis correcting for age (stratification variable, cut-off 50 years) and variation at the level of the GP practice, patient and repeated observations. The level-3 variance (GP practice) was equal to 0. The interaction between group and time was not significant.

Diff., difference.
Table 3  Secondary outcome measures

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Ultrasound tailored treatment</th>
<th>Usual care</th>
<th>Between-group difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>Mean (SD)</td>
<td>N</td>
</tr>
<tr>
<td>SPS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>56</td>
<td>20.6 (4.1)</td>
<td>55</td>
</tr>
<tr>
<td>3 months</td>
<td>53</td>
<td>14.3 (5.1)</td>
<td>53</td>
</tr>
<tr>
<td>6 months</td>
<td>47</td>
<td>13.1 (4.7)</td>
<td>47</td>
</tr>
<tr>
<td>12 months</td>
<td>51</td>
<td>11.1 (5.3)</td>
<td>50</td>
</tr>
<tr>
<td>SDO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>56</td>
<td>69.6 (17.3)</td>
<td>55</td>
</tr>
<tr>
<td>3 months</td>
<td>53</td>
<td>47.2 (32.4)</td>
<td>53</td>
</tr>
<tr>
<td>6 months</td>
<td>47</td>
<td>40.7 (29.7)</td>
<td>47</td>
</tr>
<tr>
<td>12 months</td>
<td>51</td>
<td>24.3 (30.2)</td>
<td>50</td>
</tr>
<tr>
<td>EQ-SD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>56</td>
<td>0.68 (0.20)</td>
<td>55</td>
</tr>
<tr>
<td>3 months</td>
<td>53</td>
<td>0.76 (0.21)</td>
<td>53</td>
</tr>
<tr>
<td>6 months</td>
<td>47</td>
<td>0.83 (0.18)</td>
<td>47</td>
</tr>
<tr>
<td>12 months</td>
<td>51</td>
<td>0.81 (0.27)</td>
<td>50</td>
</tr>
<tr>
<td>EQ-5D VAS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>56</td>
<td>69.4 (15.2)</td>
<td>55</td>
</tr>
<tr>
<td>3 months</td>
<td>53</td>
<td>72.6 (17.6)</td>
<td>53</td>
</tr>
<tr>
<td>6 months</td>
<td>47</td>
<td>78.4 (12.0)</td>
<td>47</td>
</tr>
<tr>
<td>12 months</td>
<td>51</td>
<td>78.3 (15.8)</td>
<td>50</td>
</tr>
</tbody>
</table>

*Linear mixed model analysis corrected for baseline and age (stratification variable, cut-off 50 years) and variation at the level of the GP practice, patient and repeated observations. The level-3 variance (GP practice) was equal to 0 for SPS and EQ-5D. The interaction between group and time was not significant.

EQ-5D VAS, Visual Analogue Scale 0–100 (100=best health status); EQ-5D, EuroQol five-item quality of life questionnaire tariff –1 to 1 (1=highest health-related quality of life); SDQ, Shoulder Disability Questionnaire 0–100 (100=most severe disability); SPS, Shoulder Pain Score 7–28 (28=most pain).

Table 4  Use of healthcare resources during 1-year follow-up

<table>
<thead>
<tr>
<th></th>
<th>Ultrasound tailored treatment</th>
<th>Usual care</th>
<th>OR†</th>
<th>95% CI</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP re-consultation, n (%)</td>
<td>24 (43)</td>
<td>30 (55)</td>
<td>1.05</td>
<td>0.49 to 2.25</td>
<td>0.90</td>
</tr>
<tr>
<td>No. of re-consultations (mean, SD)</td>
<td>1.7 (0.9)</td>
<td>2.1 (1.3)</td>
<td>1.02</td>
<td>0.21 to 4.92</td>
<td>0.98</td>
</tr>
<tr>
<td>Diagnostic imaging ordered by GP, n (%)‡</td>
<td>56 (100)</td>
<td>13 (24)</td>
<td>0.58</td>
<td>0.17 to 2.04</td>
<td>0.40</td>
</tr>
<tr>
<td>– Plain radiography, n (%)</td>
<td>1 (2)</td>
<td>6 (11)</td>
<td>0.21</td>
<td>0.12 to 1.30</td>
<td>0.12</td>
</tr>
<tr>
<td>– Ultrasound imaging, n (%)‡</td>
<td>56 (100)</td>
<td>12 (22)</td>
<td>0.20</td>
<td>0.10 to 0.41</td>
<td>0.001</td>
</tr>
<tr>
<td>Physiotherapist referral, n (%)</td>
<td>33 (59)</td>
<td>35 (64)</td>
<td>0.83</td>
<td>0.38 to 1.81</td>
<td>0.64</td>
</tr>
<tr>
<td>No. of physiotherapy sessions (mean, SD)§</td>
<td>12.9 (9.0)</td>
<td>12.4 (10.2)</td>
<td>1.00</td>
<td>0.73 to 3.47</td>
<td>0.25</td>
</tr>
<tr>
<td>Medication used, n (%)</td>
<td>31 (55)</td>
<td>36 (65)</td>
<td>1.59</td>
<td>0.73 to 3.47</td>
<td>0.25</td>
</tr>
<tr>
<td>– GP prescription, n (%)</td>
<td>12 (21)</td>
<td>21 (38)</td>
<td>0.85</td>
<td>0.37 to 1.95</td>
<td>0.70</td>
</tr>
<tr>
<td>– OTC, n (%)</td>
<td>25 (45)</td>
<td>27 (49)</td>
<td>1.17</td>
<td>0.55 to 2.49</td>
<td>0.50</td>
</tr>
<tr>
<td>No. of GP prescriptions (mean, SD)</td>
<td>1.6 (1.2)</td>
<td>1.4 (0.7)</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Corticosteroid injections by GP, n (%)</td>
<td>22 (39)</td>
<td>17 (31)</td>
<td>1.27</td>
<td>0.57 to 2.82</td>
<td>0.56</td>
</tr>
<tr>
<td>No. of injections (mean, SD)</td>
<td>1.4 (0.67)</td>
<td>1.5 (0.51)</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Secondary care referral, n (%)</td>
<td>11 (20)</td>
<td>7 (13)</td>
<td>2.23</td>
<td>0.73 to 6.82</td>
<td>0.16</td>
</tr>
<tr>
<td>Surgery, n (%)</td>
<td>1 (2)</td>
<td>3 (5)</td>
<td>0.74</td>
<td>0.10 to 5.54</td>
<td>0.77</td>
</tr>
</tbody>
</table>

*One patient gave no consent to obtain her patient record, and one GP did not deliver the patient record.
†Logistic regression analysis corrected for age (stratification variable, cut-off 50 years).
‡Including intervention ultrasound.
§From five patients in each group, the number is lacking.
GP, general practitioner; OTC, over-the-counter medication.
Table 5  Secondary outcome measures: linear mixed model analysis adjusting for age or all putative prognostic factors with sensitivity analysis

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>3 months</th>
<th></th>
<th>6 months</th>
<th></th>
<th></th>
<th>12 months</th>
<th></th>
<th>Overall</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean difference (95% CI)</td>
<td>p Value</td>
<td>Mean difference (95% CI)</td>
<td>p Value</td>
<td>Mean difference (95% CI)</td>
<td>p Value</td>
<td>Mean difference (95% CI)</td>
<td>p Value</td>
<td></td>
</tr>
<tr>
<td>SPS*</td>
<td>-2.02 (-4.17 to 0.13)</td>
<td>0.065</td>
<td>-1.57 (-3.73 to 0.58)</td>
<td>0.155</td>
<td>-1.59 (-3.76 to 0.58)</td>
<td>0.149</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-1.73 (-3.97 to 0.52)</td>
<td>0.130</td>
<td>-1.30 (-3.52 to 0.93)</td>
<td>0.251</td>
<td>-1.38 (-3.70 to 0.93)</td>
<td>0.239</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-1.71 (-3.96 to 0.54)</td>
<td>0.133</td>
<td>-1.27 (-3.51 to 0.96)</td>
<td>0.259</td>
<td>-1.36 (-3.68 to 0.95)</td>
<td>0.245</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Best case</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-2.05 (-4.15 to 0.06)</td>
<td>0.057</td>
<td>-1.71 (-3.84 to 0.43)</td>
<td>0.116</td>
<td>-1.70 (-3.85 to 0.45)</td>
<td>0.119</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-2.02 (-4.13 to 0.10)</td>
<td>0.061</td>
<td>-1.64 (-3.79 to 0.51)</td>
<td>0.133</td>
<td>-1.62 (-3.79 to 0.54)</td>
<td>0.139</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-1.74 (-3.96 to 0.48)</td>
<td>0.123</td>
<td>-1.35 (-3.55 to 0.86)</td>
<td>0.228</td>
<td>-1.39 (-3.69 to 0.91)</td>
<td>0.233</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-1.74 (-3.96 to 0.48)</td>
<td>0.123</td>
<td>-1.35 (-3.55 to 0.86)</td>
<td>0.228</td>
<td>-1.39 (-3.69 to 0.91)</td>
<td>0.233</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Worst case</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-1.88 (-3.99 to 0.23)</td>
<td>0.080</td>
<td>-1.89 (-4.05 to 0.27)</td>
<td>0.086</td>
<td>-1.75 (-3.89 to 0.40)</td>
<td>0.109</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-1.85 (-3.96 to 0.27)</td>
<td>0.086</td>
<td>-1.82 (-3.99 to 0.35)</td>
<td>0.099</td>
<td>-1.67 (-3.83 to 0.49)</td>
<td>0.127</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-1.74 (-3.96 to 0.48)</td>
<td>0.123</td>
<td>-1.57 (-3.81 to 0.67)</td>
<td>0.166</td>
<td>-1.44 (-3.74 to 0.86)</td>
<td>0.217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-1.74 (-3.96 to 0.48)</td>
<td>0.123</td>
<td>-1.57 (-3.81 to 0.67)</td>
<td>0.166</td>
<td>-1.44 (-3.75 to 0.86)</td>
<td>0.216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*SDQ†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-4.41 (-15.00 to 6.16)</td>
<td>0.412</td>
<td>3.03 (-8.24 to 14.31)</td>
<td>0.597</td>
<td>-6.86 (-19.81 to 6.08)</td>
<td>0.296</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-3.15 (-14.82 to 8.51)</td>
<td>0.595</td>
<td>3.12 (-8.78 to 15.03)</td>
<td>0.606</td>
<td>-6.57 (-20.31 to 7.16)</td>
<td>0.345</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-3.36 (-15.07 to 8.34)</td>
<td>0.572</td>
<td>2.94 (-8.96 to 14.84)</td>
<td>0.627</td>
<td>-6.59 (-20.16 to 6.98)</td>
<td>0.338</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Best case</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-3.38 (-13.82 to 7.05)</td>
<td>0.524</td>
<td>2.94 (-8.35 to 14.24)</td>
<td>0.609</td>
<td>-5.24 (-18.20 to 7.72)</td>
<td>0.425</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-3.24 (-13.72 to 7.24)</td>
<td>0.543</td>
<td>3.05 (-8.25 to 14.35)</td>
<td>0.596</td>
<td>-5.06 (-17.91 to 7.78)</td>
<td>0.437</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-2.56 (-14.15 to 9.02)</td>
<td>0.663</td>
<td>2.22 (-9.81 to 14.26)</td>
<td>0.716</td>
<td>-5.64 (-19.54 to 8.25)</td>
<td>0.422</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-2.56 (-14.23 to 9.09)</td>
<td>0.665</td>
<td>2.22 (-9.81 to 14.26)</td>
<td>0.716</td>
<td>-5.62 (19.26 to 8.01)</td>
<td>0.416</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*Worst case</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-3.75 (-14.25 to 6.74)</td>
<td>0.482</td>
<td>2.57 (-8.81 to 13.96)</td>
<td>0.657</td>
<td>-5.03 (-18.18 to 8.12)</td>
<td>0.450</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—age 50</td>
<td>-3.61 (-14.15 to 6.93)</td>
<td>0.501</td>
<td>2.69 (-8.71 to 14.08)</td>
<td>0.643</td>
<td>-4.85 (-17.88 to 8.18)</td>
<td>0.463</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-2.90 (-14.59 to 8.78)</td>
<td>0.625</td>
<td>1.90 (-10.27 to 14.07)</td>
<td>0.759</td>
<td>-5.43 (-19.55 to 8.69)</td>
<td>0.448</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-2.90 (-14.66 to 8.85)</td>
<td>0.627</td>
<td>1.90 (-10.26 to 14.06)</td>
<td>0.759</td>
<td>-5.41 (-19.26 to 8.43)</td>
<td>0.440</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>EQ-5D</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—all</td>
<td>0.01 (-0.07 to 0.09)</td>
<td>0.766</td>
<td>0.05 (-0.04 to 0.13)</td>
<td>0.284</td>
<td>0.004 (-0.10 to 0.09)</td>
<td>0.931</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>0.02 (-0.07 to 0.10)</td>
<td>0.715</td>
<td>0.07 (-0.03 to 0.16)</td>
<td>0.155</td>
<td>0.002 (-0.11 to 0.11)</td>
<td>0.976</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>0.02 (-0.07 to 0.10)</td>
<td>0.715</td>
<td>0.07 (-0.03 to 0.16)</td>
<td>0.155</td>
<td>0.001 (-0.11 to 0.11)</td>
<td>0.992</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*EQ-5D VAS†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITT—all</td>
<td>-0.23 (-5.26 to 4.80)</td>
<td>0.929</td>
<td>2.68 (-2.79 to 8.14)</td>
<td>0.336</td>
<td>1.39 (-5.00 to 7.78)</td>
<td>0.667</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—age 50</td>
<td>-1.30 (-7.07 to 4.47)</td>
<td>0.657</td>
<td>2.76 (-3.16 to 8.68)</td>
<td>0.359</td>
<td>2.09 (-4.53 to 8.71)</td>
<td>0.533</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PP—all</td>
<td>-1.30 (-6.94 to 4.34)</td>
<td>0.650</td>
<td>2.76 (-3.14 to 8.66)</td>
<td>0.357</td>
<td>1.95 (-5.03 to 8.92)</td>
<td>0.580</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The level-3 variance (GP practice) was equal to 0. Final model was based on lowest Akaike’s Information Criterion, which resulted in an unstructured covariance structure for repeated measures (level 1) and no random effects at levels 2 (patients) and 3 (GP practice).

†Best model (lowest Akaike’s Information Criterion) was model with random intercept and random slope (time) on patient level and random intercept at GP practice level.

Age, stratification variable (cut-off 50 years); All, all putative prognostic factors; EQ-5D, Euroqol five-item quality-of-life questionnaire tariff; ITT, intention-to-treat analysis; PP, per-protocol analysis; SDQ, Shoulder Disability Questionnaire; SPS, Shoulder Pain Score; VAS, Visual Analogue scale.
additional costs for the intervention group, which cannot be justified in routine practice based on this trial. Based on this study, no change in current pragmatic guidelines to incorporate early ultrasound imaging can be recommended.

**Strengths and limitations**

To the best of our knowledge, this is the first primary care study investigating the effect of using ultrasound imaging in the management of patients with shoulder pain to target treatment to the specific underlying patho-anatomical disorders. This study has several strengths. First, this study was developed as a pragmatic trial to inform clinicians, guideline developers and policymakers to choose wisely between options for care. Pragmatic trials measure effectiveness in routine clinical practice and reflect variations between patients that occur in real clinical practice improving generalisability. For instance, our trial reflects variations between GPs and physiotherapists who applied the treatments, and between radiologists who performed the ultrasound. Second, we filtered out patients with a favourable natural course (14%) during the 2-week qualification period to prevent non-responders in the course of the trial. Third, since blinding for ultrasound diagnosis in the usual care group had to be taken into account to prevent information bias, radiologists were not allowed to discuss their findings with patients. We evaluated this by asking patients one question in the questionnaire at 3 months; blinding of patients for the ultrasound diagnosis was violated only once. On the other hand, in five patients in the usual care group, the ultrasound diagnosis was disclosed to the GP by mistake. We incorporated these protocol violations in the per-protocol analysis. Finally, our study seems representative as the recovery rate of 60% in the usual care group corresponds to recovery rate used in the sample size calculation, which was based on previous studies.

Our study knows some limitations related to under-enrolment. First, under-enrolment hampers the interpretation of the results. We found a non-statistically significant difference in favour of ultrasound tailored treatment. However, this lack of evidence does not necessarily mean there is no effect. The original sample size was calculated on a between-group difference in recovery rate of 20% (80% vs 60% for ultrasound tailored treatment and usual care, respectively) after 1 year. However, we observed a non-significantly difference of 12.5% (72.5% vs 60%) in recovery rate. Based on a binominal distribution, we have calculated that there is a chance of 10.6% to observe a difference in effect of 20%, if we had reached the target number of patients. Recalculation of the between-group difference for a reduced sample size of 111 randomised patients indicated that a recovery rate of ~23% was necessary to reach the level of significance. Very limited literature is available on the minimal clinically important difference for the Global Perceived Effect questionnaire. We choose this outcome measure as it provides the patient’s perspective on the impact of disorder and treatment.44 One can argue that a difference of >10% is clinically important, but population and context determine what is clinically important.45 The observed difference of 12.5% would have been statistically significant if we had included 448 patients, far more than calculated and not feasible in this design. Second, although we carefully planned the recruitment aspects and GPs beforehand responded positively to the feasibility, overtime it became clear that the target number of patients would be difficult to achieve within 2 years of recruitment. To facilitate greater patient inclusion, we relocated the inclusion procedure from the GP practice to the Department of Radiology at the Zuyderland Medical Centre. GPs were asked to refer eligible patients to the radiology department for inclusion and ultrasound imaging of the shoulder. This provided two advantages: all GPs working in the region of the Zuyderland Medical Centre were capable of referring patients for inclusion, and time spent by GPs on patient recruitment decreased to a minimum. This relocation was combined with the extension of the recruitment period with an additional year. This adjusted inclusion procedure yielded 42 additional patients. As this recruitment strategy might have induced a form of selection bias, it was added as a covariate to the mixed model analysis.

Another limitation is that we are unaware of the number of eligible patients as GPs rarely registered numbers and reasons for non-participation. However, this reflects the consultation process in general practice, where recruitment is rarely straightforward, but might influence the external validity of our findings.46

**Comparison with existing literature**

Previous studies, in general practice, focused on the effectiveness of treatments in patients with subacromial impingement syndrome, for example, the effectiveness of corticosteroid injections, exercise or manual therapy.47–49 Instead of combining a diagnostic imaging test to inform GPs about underlying patho-anatomical disorders to tailor treatment to the observed underlying disorder, all these studies used the generic term subacromial impingement syndrome. We chose to label and define the shoulder disorders based on findings from physical examination followed by ultrasound imaging instead of solely findings from physical examination, as this often poorly correlates with the underlying disorder.12–14 This lack in uniformity in the way shoulder disorders are labelled and defined hampers comparisons.50–52 Therefore, content and effectiveness of this study cannot be compared to other randomised studies.

**Implications for practice and research**

Our findings do not support the ordering of ultrasound imaging at the initial visit of patients in whom the GP suspects a subacromial disorder. The current shoulder pain guidelines state that ultrasound imaging should be
considered in cases with an abnormal course or therapy resistance, an advice that still holds.\textsuperscript{10, 15} As our study was under-enrolled, larger studies are needed to provide conclusive evidence.

GPs generally tend to ignore recommendations on additional imaging in guidelines as ultrasound imaging is frequently applied in patients with acute shoulder pain.\textsuperscript{21 26 53} If GPs still consider ultrasound imaging, it might be used preferentially in patients aged 40 years or older, as over 90% of shoulder pain patients in this age group have a subacromial disorder, compared to a much lower prevalence in younger age groups.\textsuperscript{25}

In patients with persistent pain, ultrasound imaging may yield a more specific diagnosis, provide a rationale for further treatment and inform patients about the prognosis of their disorder. For daily practice, it is important that ultrasound findings are considered in the clinical context, as asymptomatic findings may be detected. A situation that can pose a real challenge.

The question whether ultrasound tailored treatment is cost-effective has not been answered. Although not calculated, healthcare costs in this study probably do not differ between groups, as mean healthcare use is similar in both study groups. However, it is suggested that episodes of shoulder pain generate additional costs for productivity losses due to sick leave.\textsuperscript{34} In an economic evaluation, these costs should be incorporated.

We showed that most patients were diagnosed with calcific tendinitis; therefore, attention can be focused predominantly on this disorder. There is growing evidence that extracorporeal shock wave therapy and barbotage are effective treatment options for calcific tendinitis, and that corticosteroid injections can be harmful,\textsuperscript{55–58} which emphasises the importance of evaluating these treatment options in general practice.

REFERENCES


Author affiliations

\textsuperscript{1}Department of Family Medicine, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands
\textsuperscript{2}Department of Methodology and Statistics, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands
\textsuperscript{3}Department of Radiology, Maastricht University Medical Centre, Maastricht, The Netherlands
\textsuperscript{4}Department of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands

Acknowledgements The authors are grateful for the participating general practices and patients, without whom this study could not have been conducted. The authors thank the management and staff of the Department of Radiology of the Zuyderland Medical Centre (Sittard-Geleen) and MCC Omnes.

Contributors All authors played a role in the study design. RPGO and BW analysed the data. All authors played a role in interpreting the data. RPGO wrote the first draft of the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final draft. RPGO is the guarantor.

Funding Medical Coordination Centre Omnes, Sittard-Geleen, the Netherlands, funded the ultrasound examinations, incentives for GPs and travel/parking expenses of patients, and were not involved in the design and conduct of the study; collection, management, analysis and interpretation of the data and preparation, review or approval of the manuscript. JWLC is supported by a Veni grant (91614078) of the Netherlands Organisation for Health Research and Development (ZonMW).

Competing interests None declared.

Ethics approval The Medical Ethics Committee of the Maastricht University Medical Centre has approved this protocol (ID 10-3-047).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/


