Author's response to reviews

Title: "Hydrodilation, corticosteroids and adhesive capsulitis: A randomized controlled trial

Authors:

Einar Kristian Tveita (e.k.tveita@medisin.uio.no)
Rana Tariq (rana.tarig@adsl.no)
Solve Sesseng (ssesseng@online.no)
Niels Gunnar Juel (NielsGunnar.Juel@ulleval.no)
Erik Bautz-Holter (erik.bautz-holter@medisin.uio.no)

Version: 4 Date: 9 January 2008

Author's response to reviews: see over
RCT Cover letter for revised version

Responses to reviewer Macdermid’s report:

Major Issues

2. Regarding selection criteria and acute frozen shoulder: This issue has been addressed in the “Discussion” section of the revised manuscript (p. 16).

3. Regarding baseline SPADI scores and group distribution: This issue has been responded to by including the following sentence in the “Randomization” section of the revised manuscript: “11 patients (30%) in the INJ group and 14 patients (36%) in the DIL group had baseline SPADI ≤ 50.”

Minor Issues

1. Regarding power: Use of regression when analyzing controlled trials is discussed in an article by Vickers and Altman who state that regression “generally has greater statistical power to detect a treatment effect”. The reference is included in the “Methods” section. Sample size calculations for our study were based on the view that the regular t-test is a type of simple regression. A pilot and study indicated that the inclusion of important variables like e.g. “baseline SPADI” would mean an improvement compared to the simple model, meaning that a smaller sample size would be appropriate. The results of the regression analysis fulfilled this expectation. What we did not expect, however, was the high SDs, which clearly reduced the ability to demonstrate treatment effects. Looking at the width of confidence intervals gives an indication of actual study power. We expected a SD of ΔSPADI of 17 points. With 44 patients in each group, this estimate would indicate a confidence interval width of (ΔSPADI group difference) of 15 points in our setting if a regular t-test was used. In the actual study, confidence interval of (ΔSPADI group difference) according to the regression analysis was (-5 to 11), so the width of the CI was slightly above 15 points. The “power issue” has been responded to in the “Discussion” section of the revised version by including the following sentences: “The waiting period may also have led to the larger variation in patient outcome (ΔSPADI) than we had expected (SDs were 21-22 compared to 17). On the other hand, the regression model appears to explain a large proportion (53%) of the variance in ΔSPADI, and it is our view that the trial is not essentially undersized for this type of study” (p. 17). We have also included the sentence “Confidence intervals do not exclude a difference in SPADI improvement larger than the 10-point difference chosen beforehand to be the minimal clinically significant difference, meaning that an important treatment effect from dilatation cannot be excluded” (p. 15).

2. Regarding physiotherapy: Physiotherapy may have a positive effect on this condition when combined with corticosteroids. Whether the specific effect of dilatation is more pronounced in combination with physiotherapy is unclear, and remains to be demonstrated in future well-designed trials.
Responses to reviewer Buchbinder’s report:

Major Compulsory Revisions

1. Regarding number of injections: The number of injections used in our study was based on pilot studies and previous studies. The majority of researchers have given only one injection, that is true. Some use three injections, while Gam and colleagues report giving up to six injections. Although grossly underpowered, Gam’s study (as the only one) suggests that dilatation may be superior to simple injection in improving ROM. Our choice was to use a number that was closer to the “usual” approach, yet was not essentially inferior to Gam’s regimen. Our results for the PROM effect of dilatation contrast with the findings of Gam et al., and we agree that possible future researchers of this topic might prefer giving only one injection.

2. Regarding injection volumes: Overall means were 7 ml and 21 ml, respectively. In the “Interventions” section, mean (SD) volume injected at each time point and overall is given as well as the number of patients in the DIL group who needed more than 20 ml (p. 9).

3. Regarding minimization: The number of participants already allocated to each intervention and the number of participants with SPADI $\leq 50$ and SPADI $> 50$ in each group were used to decide “loading” for each new participant. In all cases there would be at least 1/6 chance that the patient would be allocated to say DIL. If the two groups had the same number of participants, the chance would be at least 2/6, rising to at least 3/6 if the INJ group was largest. If the INJ group had an overweight of patients with SPADI $\leq 50$, and the patient had a SPADI score of 45, the patient would have at least 3/6 chance of being allocated to DIL. If the INJ group had the highest number of patients with SPADI $\leq 50$, and the INJ group was largest, the patient would have a 5/6 chance of being allocated to DIL. The system worked in a similar way for SPADI scores above 50. Stratification and block randomization with variable block size would have been an equally good alternative. The difference is mainly that with minimization, allocation is not based on a sequence that has been generated prior to the study. The issue has been responded to by including the following sentences in the “Randomization” section of the revised version: “Minimization was preferred to purely random allocation in order to reduce group differences at baseline. The chance of being allocated to a certain intervention group varied from 1/6 to 5/6 depending on initial SPADI [31] scores (stratifying for SPADI $\leq 50$ or SPADI $> 50$) and the number of patients already allocated to each intervention.”

4. Regarding minimization and outcome assessment procedures: The person (researcher) responsible for the allocation of patients is not an expert on the minimization approach, but he has experience from organizing randomization procedures in trials. He was also involved in outcome assessment.

5. Regarding blinding and bias: This issue has been addressed in the “Discussion” section of the revised manuscript (p. 17).

6. Regarding delay: We have included the following sentence in the “Patient flow” section: “Mean time from inclusion (baseline assessment) to the first injection was received was 8.2 weeks in the INJ group and 8.6 weeks in the DIL group”.
7. Regarding spontaneous recovery: The issue has been addressed in the “Discussion” section of the revised manuscript (p. 16).

8. Regarding other treatments: Few patients (DIL:5, INJ:8) were receiving physiotherapy at baseline, and only to a limited extent. We did not register during the “treatment” period whether they continued with their program or not. As we have stated in the revised version, no patients were prescribed new physiotherapy programs during the study (p. 9). Likewise, we did not register analgesics use during the intervention period, only at follow-up.

9. Regarding reliability of ROM measurements: Intra-rater reliability and repeatability has been assessed and is reported in a different manuscript which has been submitted to *BMC Musculoskeletal Disorders*. It is under review. Active movements are probably “better” than passive movements because movement of the scapula was not restricted during active movements.

10. Regarding adverse effects: In the revised version, side effects and adverse reactions are reported divided by effects and for each intervention group.

11. Regarding “analgesics” and “sick leave”: We have deleted the statement “The observed differences at follow-up may be a reflection of baseline figures for these variables”.

12. Regarding baseline differences and tests of imbalance: We have deleted the sentence “This may cause misinterpretation of treatment results” as we find it superfluous. Concerning baseline tests of imbalance we share the view of Roberts and Torgerson (*BMJ* 1999;319:185) that significance tests of baseline characteristics do not provide an appropriate criterion to assess the effect of imbalance on outcome. Baseline tests of imbalance are inappropriate unless the investigators suspect that there are problems with the randomization. Reporting of results: Mean (SD) values for SPADI and ROM for baseline, change and follow-up scores are provided along with mean (95% CI) values for group differences for change and follow-up scores (Table 3).

13. Regarding multiple regression: If only the variables “Treatment”, “Baseline SPADI score” and “Baseline pain medication” are included in the model, the coefficient for “Treatment” is 6 (CI: -1 to 13). This model explains 49% of the total variance. In the model reported in the manuscript, other predictors are also included. According to the recommendations by Senn (*Stat Med* 1991;10:1157-1160) and Roberts and Torgerson (see above), baseline variables of potential prognostic value were identified beforehand and then fitted in the analysis. We did not know which variables would eventually turn out as statistically significant.

14. Regarding intra-articular adhesions: The sentence “Recent studies of frozen shoulder patients report no distinct intraarticular adhesions.” has been changed to “Some more recent arthroscopic investigations of frozen shoulder patients report no distinct intraarticular adhesions of this type [6-9], and the role of such adhesions is unclear.” (p. 3).

15. Regarding ROM measurements and eligibility: No patients were excluded because of excessive pain when baseline values for ROM were to be established. However, some patients with shoulder pain were not eligible because they could not relax sufficiently for passive ROM to be assessed. It is unclear whether any of them had the condition adhesive capsulitis. We do not have the exact number of such patients.
16. Regarding drop-outs and group distribution: We feel that this issue is addressed in Figure 1 and does not need to be repeated in the text.

17. Regarding follow-up scores: Baseline score was adopted as follow-up score for one patient only. This has been clarified in the text by adding the sentence: “For the 75 other patients, original follow-up scores were used in the computations.” (p. 12).

Minor Essential Revisions

Regarding figures: Figures 2 and 3 are removed in the revised manuscript.