Reviewer's report

Title: Reliability of Measures of Impairments Associated With Patellofemoral Pain Syndrome

Version: 1 Date: 10 February 2006

Reviewer: Nadine Foster

Reviewer's report:

General

A useful contribution to the literature on this subject. The paper is mostly clear and well written. The illustrations and detailed descriptions of the tests are useful. The analyses are appropriate and there is a good balance of measures, including both ICCs and SEMs, which provide complementary information. Addressing the following points will improve the paper.

----------------------------------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Please justify or change the manuscript to reflect why intra-tester reliability is not specifically focused on. It would be more usual to include investigations of both inter and intra tester reliability. It has been suggested in the literature that if good inter-tester reliability is shown, then at least equally good intra-tester reliability can be assumed (Streiner and Norman 1995). It seems more clinically relevant to consider intra-tester reliability here, since individual patients tend to see the same health professional over time and have repeated tests from the same assessor. The authors mention ‘clinical use’ in the first conclusion - thus this seems important to explore further in the paper.

Few reliability studies include a justification of their sample size - whilst this paper does provide a justification, it is based on only one of the tests studies (out of 12 different tests). Could the authors provide some information about the adequacy of the sample size for some of the other tests studies, particularly those not providing dichotomous outcomes? Guidance for appropriate sample sizes for reliability studies have been suggested in eg. Donner and Eliasziw 1987 and Walter et al 1998.

Table 3 - It would be ideal to provide a summary of the data for each of the pairs of testers, for at least some of the tests (perhaps those with the best and the worst reliability?), so that the reader can quickly see the similarity or differences in the testers results. An additional table of this information would be useful, before Table 3. What are the means in Table 3 - the means of all 4 testers ?- this is not clear at present. I am not sure that presenting these means is useful as it tells us nothing about reliability. In addition, the 95% CIs for the SEMs would be a useful addition, as the upper limit would represent the worst expected error for each of the tests, and this would be a more conservative figure to use in the discussion on page 18.

When ICCs are used, the results can be influenced by the homogeneity of the sample, the authors acknowledge this on page 16. ICCs are influenced by the within subjects and between subjects variance (see Haas 1991, Keating and Matyas 1998, Rankin and Stokes 1998). Great variation among subjects can increase the ICC when there may be considerable difference between the measurements. Could the authors comment on this in the discussion and whether they feel this could have influenced their results? The age range alone in this study may have meant considerable diversity in the sample.
The reasonably impressive reliability results found in this study may be an over-estimate compared to real clinical practice, given that the time difference between the tests was so short. The authors comment on this on page 14 when comparing their findings to those of Hunt. Of what clinical relevance is the very short time interval in the current study and what do the authors feel is the meaning of the reliability values for clinical practice? Discussion on these issues would be very useful for the paper.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The introduction is well written, mostly. There are some sentences which are a little unclear and should be amended: Page 2 sentence which starts 'Current study' - which study is this? Page 3 sentence which starts 'Quality of movement' needs reworded.

In order to be comprehensive, I also feel it would be useful to include definitions of both intra and inter-tester reliability in the 2nd paragraph on page 1, and in the final section on page 3, explain why the focus is on inter-tester reliability, rather than intra-tester reliability, which is probably more clinically relevant when one health professional tests and re-tests a patient over time.

Check order of references - is reference 12 on page 1 correct?

The introduction mentions that reliability is tested on several occasions on stable subjects - can you clarify how you ensured that your subjects were indeed 'stable', in the methods section?

In Methods section - please provide some justification for the selection of your inclusion criteria? Do you mean 'known or suspected' internal knee derangement and ligament injury etc? The choice of subjects from age 12 to 50 is interesting - and potentially important for your results - this age range needs some justification. In particular, what might be the consequences of including adolescents as well as adults? Could something be added to the discussion on this issue? I suggest that rather than a mean age in Table 1, breaking down the proportions of the sample is each decade banding would shed more light on the ages within the sample.

Table 1 needs some footnotes to explain the range of scores for the NRS and the ADL score.

Results section - while it is undesirable to have too much repetition between the Tables and the text, some mention of the key findings in the results section on page 13 would be useful.

Page 14 - typo on word tissue

As a general point, the authors should be careful about comparing their results with those from studies using inappropriate analyses for reliability eg. Pearsons correlation co-efficients.

When the 95% CIs for the SEMs are added to the Table, could some more discussion of these be developed on page 18?

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

I have no competing interests.