Reviewer's report

Title: A systematic review of the psychometric properties of the Boston Carpal Tunnel Questionnaire

Version: 1 Date: 15 July 2006

Reviewer: Isam Atroshi

Reviewer's report:

This is an interesting study and a useful review of the carpal tunnel syndrome (CTS) questionnaire. CTS is one of the most common upper extremity disorders and the CTS questionnaire is the only disease-specific patient-reported outcomes measure currently available in CTS.

Comments

1. The first question concerns the name of this questionnaire. Unfortunately, many different names have been used in the literature to refer to the CTS questionnaire. Even the developers of the questionnaire have used different names to refer to it in their papers. I am not sure how this can be resolved; perhaps contact with the developers to discuss an appropriate name for the questionnaire might help.

2. It would be interesting and helpful for readers to mention whether or not there are any recommendations about how to manage missing responses for this questionnaire (for example how many missing item responses are allowed for a score to be calculated).

3. The authors present the results of the CTS questionnaire’s 2 scales (symptom severity and functional status) together for each study. The symptom severity and functional status scales are distinct scales that measure two entirely different dimensions. It could be much more helpful if the authors treat them as such and the 2 scales are presented separately with regard to their psychometric properties and other issues throughout the manuscript.

4. Among the studies reviewed, the authors include the studies that calculated a total score in addition to the usual symptom severity and functional status scores. A total score for the CTS questionnaire is rarely used in the literature and I am not sure if it is a valid way of using the questionnaire. It is uncertain whether it is helpful to show these results provided by only few of the studies reviewed; it may be better just to use the results of the 2 scales independently, which are provided by almost all the studies. Otherwise the authors would need to state whether they believe the total score is a valid way to use the CTS questionnaire.

5. Table 1:
   a. The study by Atroshi et al. (1998) also assessed responsiveness (as in fact shown in the text and in Table 3) but this is not marked in the Table.
   b. Electrophysiological nerve conductivity, abbreviated as ENC, is an unusual term for nerve conduction studies (NCS) or nerve conduction tests.
   c. In the last column No. of patients in the analysis, acceptability in a study is said to have been evaluated by 312 patients; it seems that it may have been the number of questionnaires rather than patients. This may need to be clarified.

6. Of the 9 studies included in the synthesis, 4 came from the same research group (references 10, 15, 18, 19) but it is not mentioned whether there was any overlap in the patient populations in these studies. If the same patients were included in different studies then the results of these studies may not differ substantially and the studies cannot be regarded as totally independent. At least the readers should be informed about any overlap in these studies.

7. Responsiveness: it is important to specify whether a study assessing the responsiveness of the CTS questionnaire calculated effect size or standardized response mean (SRM) based only on the subgroup of patients who had reported (or judged by some other criterion) that they were improved after treatment or on the whole patient population (including those who were unchanged or worse after treatment). If all patients were included then the responsiveness is underestimated because the mean change score for the whole population will be lowered and the standard deviation for the change scores will be increased by those who were unchanged or worse. If treatment does not produce change or causes deterioration in a subgroup, the
effect size or SRM for the total population will be lower even if the measure is highly responsive. Because surgery in CTS is usually very effective the effect of including unchanged/worsened patients may not be substantial and the calculated effect size or SRM will still be large. If all studies in this review calculated effect size or SRM based on all patients then even other studies that did not have responsiveness as a primary purpose but calculated effect size or SRM to assess treatment effectiveness would in fact provide the same information about responsiveness as the studies in the review (the responsiveness will be similarly underestimated).

8. Effect size and SRM do not always yield similar values (page 7, line 3); in fact they often differ depending on whether the patients were more homogenous at baseline (ie smaller baseline standard deviation) or during change (smaller SD of change scores). In some papers the term effect size might have been used despite that the SRM was actually calculated. The authors would need to check the method’s description to ensure that the correct terminology was used.

9. Table 3:
a. Responsiveness of a “total score” is not a commonly used method.
b. There seems to be a missing line in the footnote (*).
c. In the assessment interval column, it is not always obvious what the “after” refers to in term of type of treatment (e.g., surgery).
d. Responsiveness based on retrospective responses to the questionnaire may not be valid.

10. One of the studies cited in the responsiveness data partly used data from a retrospective cohort; because this methodology is of limited value, it may not be very helpful to include the results from this particular patient group in the responsiveness evaluation and justify this omission in the discussion.

11. In the section about reliability (page 7), it seems strange that reliability in the study by Greenslade et al. was assessed by applying the questionnaire “before and two weeks after surgery”; was it administered twice on each of these occasions?

12. Interpretability (page 8), it is not stated which scale the minimal important change refers to (and also the score range is 1 to 5 and not 0 to 5 as stated).

13. It is stated in the discussion (page 10) that construct validity was supported by high Cronbach alpha (internal consistency); is it commonly agreed that internal consistency is related to construct validity, maybe this could be supported by an appropriate reference.

14. The authors may consider discussing the scoring for the CTS scales, and whether the 1 to 5 scale currently used is a better option than converting the raw score to a 0 to 100 score as is the case with many other questionnaires.

15. The paper by Rosales RS et al., J Hand Surg [Am] 2002; 27(2):334-43, also examined the psychometric properties of the CTS questionnaire but it is not included or even cited.

What next?: Accept after minor essential revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: No

Declaration of competing interests:

'I declare that I have no competing interests'