Reviewer’s report

Title: Pain in elderly people with severe dementia: A systematic review of behavioural pain assessment tools.

Version: 1 Date: 3 October 2005

Reviewer: Kenneth Craig

Reviewer’s report:

An overview of nonverbal, behavioral measures of pain in the elderly with severe dementia is most welcome. A careful critical analysis is badly needed, particularly one that examines psychometric properties. I found the paper well written and clear. The authors provide a powerful statement of the need for better measures and very effectively document the nature of the problem addressed in the introduction. The methodology would appear to be consistent with standard systematic review practice and to have been successful in comprehensively identifying the key literature. Incorporating English, Dutch, German and French literatures was commendable.

My critical observations relate to a need for more critical analysis and a need to persist with the framework developed for critically evaluating the measures.

For example, the observation that self-report is accepted as the "gold standard" for pain assessment is inaccurate. Self-report has many critics. Its reliability is not as substantial as the authors would have it. It is heavily influenced by context, has difficulty reflecting the complexity of the experience, particularly when unidimensional scales are used, health professionals often question the credibility of self-report, etc. Some general critical commentary is in order, beyond the analysis of limited application to the elderly. Too often self-report measures are "fools gold".

Physiological measures are also mentioned in passing, without adequate critical analysis, other than the suggestion that they have not been studied enough. It surely should be said that virtually all the measure proposed to date are as responsive to non-noxious stress as they are to noxious events, therefore limiting their use as specific indices of pain.

The "nutshell" analyses of the various scales using a priori criteria makes for interesting, well considered, and useful reading. However, it was surprising that the rating criteria described in Table 2 were not applied in detail to the various scales, other than to generate an overall quality judgement. Thus, the "nutshell" accounts represent selective anecdotal observations, rather than the application of systematic criteria, even though the criteria were articulated. The paper would have benefited from detailed analysis using these criteria. Further, while there was an attempt to generate an overall quality score for the different measures, the constituents of these scores could be questioned. In particular, it would have been useful if inter-rater reliability for the judgements had been demonstrated. Criteria used to evaluate the various scales (Table 2) often depend upon judgement of the...
reviewer. Inter-rater reliability of these judgements needs to be demonstrated. Without this information, it is difficult to know how to interpret the overall scores. Item validity of many items in the scales seems questionable. Sensitivity-specificity should be addressed more clearly. It is not always clear that the item has been demonstrated to be responsive to pain. For example, people do sleep despite pain, verbal reactions are predicated on their impact, hence not always indicative of pain, and "problems of behavior" need to be empirically demonstrated as specifically indicative of pain. The characterization of the item ‘facial expression' for the DEGR suggests a confounding of cognitive (“concerned face”) and emotional (“frightened”) states that are not painful with pain. Or on the PAINAD, pain facial display is confounded with "sad, frightened, frowning". As well, "smiling" is scored zero. Do people have to be smiling to not be in pain. A cue for limited item validity would be the limited homogeneity of items often noted. Perhaps the problem relates to the use in the development of the many scales of the use of "possible pain cues". Without careful item analyses it will be difficult to progress toward the use of unambiguous pain cues.

The authors effectively point out the proliferation of pain scales of this type. It is not unlike the turmoil in pain assessment with infants and children where investigators start de novo rather than to benefit from existing studies. One wag observed that "pain investigators would rather use another investigator's tooth brush than their pain scale". It would seem relatively easy to devise a new scale; the hard part comes in pursuing the psychometrics to produce a reliable and valid index. The responsibility for proliferation rests not only with the investigators but with journals who publish inadequately developed scales. I would have preferred a harder hitting message of this type.

I would recommend publication following revision.