Reviewer's report

**Title:** Deconstructing patient centred communication and uncovering shared decision making: an observational study

**Authors:**
- Michel Wensing (M.Wensing@hsv.kun.nl)
- Glyn Elwyn (ElwynG@cardiff.ac.uk)
- Adrian Edwards (EdwardsAG@cardiff.ac.uk)
- Eric Vingerhoets (Vingerhh@knmg.nl)
- Richard Grol (R.Grol@hsv.kun.nl)

**Version:** 1  **Date:** 12 Nov 2001

**Reviewer:** Dr Peter Bower

**Level of interest:** A paper whose findings are important to those with closely related research interests

**Advice on publication:** Unable to decide on acceptance or rejection until I see revised version

Generally this is an interesting paper concerning the relationships between two key constructs of relevance to doctor-patient communication - 'patient-centred communication' and 'shared decision making'. The key strengths of the study are the specification of some a priori hypotheses, solid methodology, and a large sample size for the testing of effects. The findings are relatively limited in scope but will be of interest to researchers in this field.

**Major points**

1. The validity of the paper is dependent on the measurement methods used. However, I have a number of concerns in this regard. The authors point out that good measures of these constructs were not available at the time of the study, and are aware of the limitations of the measures used. However, it is not clear whether better measures were available (e.g. the Stewart measure of patient-centredness), and that they were rejected as being impractical for rating such a large number of consultations.

Limitations in their chosen measuring instrument are fundamental to the paper. The scale used is global in nature, being based on ratings made by observers on 7 point scales. Although the authors are right to point out the Cronbach's alpha for the scales are low because of the small number of items included, it might also suggest that the items used to rate the two key dimensions can be further differentiated, and that the proposed scales are not measuring highly consistent concepts.

Also, I was not sure that alpha measures are optimal reliability indices in this regard. According to the methods section, two trained raters assessed the consultations. However, it is not clear whether both raters assessed all consultations, or that the consultations were split between them. If the former, then a measure of inter-rater reliability would be required, rather than a measure of internal consistency. Where is the evidence that the MAAS scales can be applied reliably by different observers? If both raters assessed all consultations, then this would need to be done, and would increase confidence that
the results were not effected by high levels of unreliability. If ratings of each consultation were not made by both observers, then the paper is significantly weakened by the lack of such a test of reliability (the alpha measures would not seem to be of relevance here, as they may just reflect the tendency of each observer to rate consultations as similar on each of the items).

2. In the section on 'measurements', the authors state that 'we accepted that these subscales represent an approximation of the issues that would be considered by a specifically designed tool, but this seemed more advantageous than using either the Braddock or OPTION scales which were aligned with one or other of the constructs under consideration.' I was not clear of the reasoning here. Why would use of a scale specific to the construct under measurement be disadvantageous? Surely that would provide the optimal measure of the construct, and thus improve the study? I would like to see this decision justified in greater detail. It was also disappointing that the opportunity was not taken to examine the global scales alongside the more specific ones, but the resource implications may have meant that it was not possible.

Minor points

1. Figure 1 would not seem crucial to the paper, as the intracluster co-efficients quoted are clear, whereas the graphical display is more difficult to interpret in my opinion.

2. One problem with hypothesis 1 is that it is difficult to determine what a 'weak' correlation is, or a 'moderate' one. Many papers in social science and psychology would claim that a correlation of 0.25 was support for a hypothesis of a relationship between variables. I agree that the level of significance is not highly relevant with such a large sample size, and the size of the correlation shows the two constructs are not redundant. However, it should be noted that there was no a priori statement of a particular size of correlation that might be expected, and thus testing this hypothesis was always going to be slightly ambiguous. For example, what level of correlation might refute this hypothesis?

3. Hypothesis 2 also suffers from the same problem, as there is no a priori statement as to what levels of intracluster co-efficient would be deemed different. I think the authors' interpretation is legitimate, but the same problem of interpretation applies. Also, I think the theoretical basis of this hypothesis requires more explanation. I know the results may have supported their hypothesis, but I was a little unconvinced by the rationale given on reading the paper the first time. Is it clear that the skills of patient-centred interviewing are acquired during training, and those of shared decision-making acquired elsewhere? Is there not also an argument that use of patient-centred skills is also flexible, in that patient-centred doctors can act both in a patient-centred manner, and a more doctor-centred manner on occasion? A more detailed explanation might help here.

4. Hypotheses were presented for gender and education, but not for the other factors tested. Were the latter tested with two tail tests and the former with one tail? If not, in what way does the specification of the hypotheses make a difference?

5. No mention was made of the representativeness of patients and doctors in the sample. Video-based studies often suffer in this regard because of their intrusive nature, and it would be helpful to have a comment about any bias in the doctors who took part. Also, what proportion of patients refused video-recording? Is there any data on these so they can be compared with those who did participate? I think there is published data on the characteristics of patients who do and do not agree to video-recording that might be of use here.

6. The new measures mentioned in the second paragraph of the Discussion should be referenced if
they are to be published.

7. I think the authors make a good point in the discussion about shared decision making being 'located in the cognitive, rational paradigm'. However, it should be noted that patient involvement is normally seen as involving both knowledge and attitudes or values, and thus cannot be directly compared with evidence-based medicine, where knowledge alone is used to determine the decision.

8. I could not understand the meaning of the Conclusion in the Abstract, which might benefit from re-writing. Equally in the last sentence of paragraph 4 in the Background, it might be clearer to say 'more highly educated persons.'

Competing interests:

None declared.