

Author's response to reviews

Title: Quality of chronic disease care in general practice: the development and validation of a provider interview tool

Authors:

Judith G Proudfoot (j.proudfoot@unsw.edu.au)
Upali W Jayasinghe (upali.jay@unsw.edu.au)
Fernando Infante (ferchoinfante@yahoo.com.au)
Justin Beilby (justin.beilby@adelaide.edu.au)
Cheryl Amoroso (c.amoroso@unsw.edu.au)
Gawaine Powell Davies (g.powell-davies@unsw.edu.au)
Jane Grimm (gudi@ranzcr.edu.au)
Christine Holton (christine.holton@adelaide.edu.au)
Tanya Bubner (tanya.bubner@adelaide.edu.au)
Mark F Harris (m.f.harris@unsw.edu.au)

Version: 4 **Date:** 27 November 2006

Author's response to reviews: see over



THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY-2052-AUSTRALIA

DR. JUDY PROUDFOOT
SENIOR RESEARCH FELLOW
CENTRE FOR GENERAL PRACTICE INTEGRATION STUDIES

PHONE: +61 2 9382 3676
FAX: +61 2 9382 8207
E-MAIL: i.proudfoot@unsw.edu.au

25 September 2006

Professor D Saltman
Editorial Director
BMC Family Practice

Dear Professor Saltman

MS: 1010098696116719

Quality of chronic disease care in Australian general practice: the development and validation of a provider interview measurement tool

We have taken on board the very helpful comments from the Reviewers, as outlined in detail below, and we have pleasure in re-submitting our revised manuscript.

Yours sincerely

Dr Judith Proudfoot

Response to Reviewers' Comments

Reviewer Mark Atkinson

1. Title leaves room for interpretation about what is being measured.

Response: Title modified

2. Aims should be sharpened to clarify that measure is designed for completion by a physician about their overall practice

Response: Modified in Abstract, Introduction and methods. Note however that the GPCCI is an interview not a self-completion questionnaire.

3. Based on the information gathered from chart reviews, did the practice case mix differ across participants?

Response: We do not have case mix data on the practices. However we do have data on the number of patients who were audited in each category. This has been incorporated into the paper as Appendix II. This shows that there was some variability between practices in the number of patients with each condition recruited however these differences were not statistically significant. We have referred to this in the methods.

4. The number of items on the coding schedule differs from the number of items on the GPCCI and comparison of content coverage between the two is not possible without a table comparing the details of both, mapping the content of Interview and coding schedule to one another.

Response: This has been included as Appendix III and referred to in the methods. It shows that the proportion of codes for each element between the GPCCI and Medical Record Audit was similar.

5. It is unclear how the data gleaned from patient charts were aggregated.

Response: Each patient record was coded. This was aggregated at practice level to provide a mean score, within each disease category and overall, for each practice.

6. Page 10 para 1 and page 12 para 2 describe how the 9 items found to reduce the internal consistency of the scale were dropped. Which items were removed?.. Thus removal of items before determining whether they independently predict the validation criteria may be premature. I suggest the original pool of interview items be entered as independent variables into regressing models "predicting" results of the clinical chart audit. Those variable that independently predict the chart results may provide an "index" of care which is more useful and psychometrically convincing than a "trait of care approach to the construct of the measure. (For a discussion of this topic see Atkinson et al).

In line with the reviewer's suggestion, and as outlined in the paper he included (2), we re-did the analyses retaining all the items in the GPCCI (ie we didn't exclude the 9 items which reduced the internal consistency of the scale). The internal consistencies of the overall scale and the 3 sub-scales were lower, but as Professor Atkinson predicted, the overall correlation between the GPCCI and the notes audit was stronger than the correlation with the 9 items were excluded.

We cross-checked our findings by regressing each of the 9 previously excluded items against the total score for the medical notes audit – none of the excluded items correlated with the overall notes audit (see table below).

Items excluded from analysis of GPCCI against Record Audit because of poor internal consistency

Subscale	Excluded items	Multiple Regression against total score for medical record audit
Asthma	Poor control resulting in further assessment, treatment or referral	NS
	Admitted to hospital in past 12 months	NS
	Planned management vs. symptom control	NS
Diabetes	Screening method	NS
	Admitted to hospital in past 12 months	NS
	Arranged follow up and patient attended	NS
IHD/Hypertension	Self management education provided	NS
	Self management education provided by others	NS
	Admitted to hospital in past 12 months	NS

7. It is unclear why items with the same scores across raters were removed from the inter-rater reliability calculations – couldn't a % agreement statistic be used instead?

We calculated reliabilities by use of the intraclass correlation (ICC). ICC was computed using an ANOVA. In the preliminary analyses we conducted a series of reliability analyses for each item. If an item has same ratings for all cases, both mean square between patients and mean square within patients are zero. Thus, both numerator and denominator of ICC formula are zero (3). However, we use adjusted total scores which include all relevant items in the final reliability computations. The sentence was revised accordingly.

8. Without knowing how the GPCCI or coding scheme were scale (eg categorical, ordinal, interval etc) and/or aggregated it is not possible to evaluate whether the statistics used provide appropriate estimates of covariance or mean group difference.

Response: Appendix III provides additional detail. The coding was categorical for both the GPCCI and the record audits. The main analyses used were item analysis, reliability analysis and correlations. Reliability analysis can be conducted using categorical data.⁴ Both Pearson and Spearman rank correlations were computed and they were consistent with each other.

9. A table should be providing containing a description of the Interview items, their scaling/response options and their distributional characteristics (mean, median, skew, SD etc.)

Response: Appendix III provides distributional characteristics for the GPCCI and Record Audits. This is referred to in the Methods.

10. GPCCI is described as an acceptable tool. Acceptable for what purposes. The better performance of the overall scale compared to the disease specific subscales could be simple of the function of the number of items (or possibly respondents) used when estimating scale reliability (due to elimination of persons without the condition).

Response: We have deleted the reference to acceptability in the Discussion and modified the discussion of the overall scale to reflect its advantage for patients with comorbidity.

11. The internal consistency of disease specific subscales differed, with asthma and heart disease schedules being substantially lower than the diabetes schedule. The observation should be

discussed, although it is unclear if internal consistency and factorial coherence are desirable characteristics of this particular instrument.

Response: We have included some discussion of this in the Discussion page 15.

12. Observations about inter-rater reliability and correlation between the coding and GPCCI results should be discussed in terms of the potential sources of error, with attention paid to where sources of error cannot be easily disentangled in the current study (eg coding error from chart error).

Response: We have added this the discussion on page 16.

13. The current findings are presented as a strong demonstration of the validity of the GPCCI. However there are a number of types of validation, and only one or two of them have been touched on by the current study. The preliminary nature of this paper, the first in a series of publications on the validity of the measure, should be emphasized a couple of times in the abstract, body and discussion of the paper.

Response: We have added this qualification to the abstract, discussion and conclusion.

14. The term internal reliability is used throughout the document, when perhaps “internal consistency” is a more precise term. Also assessment, measure, scale and interview are used interchangeably when interview schedule and interview rating scale might be more precise terms.

Response: We have altered the text as suggested to replace internal reliability with internal consistency. We have also made some changes to the use of the terms measure, scale and interview throughout the text.

15. It would be very helpful if a table were provided containing the (summary) clinical guidelines for the disease states and the items on the GPCCI that address each of the guidelines. This would help clarify how the clinical guidelines differed across the three conditions and how this was reflected in the GPCCI.

Response: We have presented this in Appendix I.

16. Were GPCCI items reviewed by colleagues involved in producing the clinical guidelines or piloted with a group of potential respondents? If so this should be stated. Were the final items assessed for clarity and comprehension?

Response: Some colleagues who were involved in the guidelines reviewed the GPCCI. The GPCCI was piloted with 5 GPs who were asked about their clarity and comprehension. This is described in the Methods on page 6.

17. The qualifications/experience and training of the five raters should be described.

Response: The five raters were post graduate researchers with experience in conducting research in general practice and clinical experience in chronic disease management. Their training included observation of their rating of simulated records. This is referred to on page 10 of methods.

Reviewer Jacques van Eijk

18. First of all the authors use the data of the medical record of the participating GPs to construct a golden standard. I don't believe this is convincing. Without explicitly defined criteria for the registration of the relevant data there is too much room for individual GPs to record data.

Response: We were uncertain what the reviewer meant by this comment. The Audit of records used a proforma with clear criteria and the reviewers were trained in its use. We have discussed the limitations of records in the Introduction section of the paper.

19. The number of units of analysis is too small. Apart from this it is confusing that the first four columns of table 1 contain 28 GPs and the last column contains 10 practices. Obviously the authors used two different units of analysis.

Response: There were 28 GPs working in the 10 practices in this study. In table 1 the unit of measurement was 10 practices with the data for GPs aggregated at practice level. We have clarified this in the labeling of the table.

20. I am left with uncertainty about the content of the GPCCI. What are the response options? What items are used for the four distinguished dimensions and how are they distributed over the three different clinical conditions.

Response: We have clarified this by including Appendix II which contains a summary of the items and their coding for both the GPCCI and the record audit.

21. Cronbach's alpha alone is not sufficient to guarantee internal reliability in this study.

Response: The purpose of the item analysis was to find those items that form an internally consistent scale and coefficient alpha reflected internal- consistency reliability. ¹ Results showed high inter-rater reliability of data extracted from patients' medical notes. The correlations between the overall GPCCI and patients' medical notes indicated the validity of the GPCCI instrument.

References

1. Spector PE. Summated rating scale construction: an introduction (1992). Sage publications: California. Pp 29-50.
2. Atkinson MJ, Lennox RD Extending basic principles of measurement models to the design and validation of Patient Reported Outcomes *Health and Quality of Life Outcomes* 2006, 4:65
3. Taylor R, Jayasighe UW et al. (2006). Reliability and validity of arm volume measurements for assessment of lymphedema. *Physical Therapy*, 86: 205-214.
4. Winer BJ. Statistical principles in experimental design (1971). McGraw-Hill Book Company: New York. Pp