Response of the authors to comments of the reviewer, Nursing Research, May 21, 2007

We thank the reviewers for their valuable input. We have tried to address all suggestions as fully as possible. Please find below our response to each comment.

Reviewer #1:
This manuscript addresses an important aspect of health services delivery - measuring outcomes of decision making related to the allocation of health human resources, specifically at the nurse/patient interface. The manuscript is generally well written, and contributes to the field by developing a measure of "resource rationing".

However, most of the paper actually addresses reliability rather than validity, and there are some fundamental approaches to assessment of validity of measures that must be addressed.

Validity - Measurement is necessarily based on theoretical underpinnings, and therefore any discussion of validity must relate to how closely the measures (indicators) reflect the concepts in the theory. The authors have identified that they based their measure on a theory of rationing from an earlier study but have provided insufficient detail in the theory of rationing in this manuscript to be able to ascertain if in fact the instrument is a valid measure. As written, it is not clear if rationing is defined as any level of care less than ideal (everything possible to provide), less than assessed requirements (by individual or groups of nurses, or by using a standardized workload measurement system), or perhaps something else. How is rationing different or similar to triage? Is rationing always a negative thing as implied, or only when it leads to patient safety issues? With the current emphasis on patient safety internationally, this should be presented in detail.

- Thank you for your valuable input. We have read the literature you suggested, which we found very informative regarding philosophical bases of validity. However, for the current validity and reliability testing of the BERNCA, after careful consideration we have chosen to stay with the APA and NCME standards, which are based on a slightly different philosophical approach than that described in the indicated references (Borsboom et al.).

- P 5 – 6: A detailed description of the conceptual framework of implicit rationing of nursing care has been added to explain the construct of implicit rationing, related constructs and influencing factors. Further, we have tried to stress the relevance of rationing regarding patient safety.

Absolute clarity on what is meant by rationing, based on a clear theoretical framework, is required in this manuscript before the authors can present an argument for why the measure is valid, and before the reader can be convinced of the same. While the authors have indicated adherence to the APA and NCME standards for testing, new work suggests that while reliability procedures are psychometrically based, "the problem of validity cannot be solved by psychometric techniques or models alone. On the contrary, it must be addressed by substantive theory. Validity is the one problem in testing that psychology cannot contract out to methodology." p. 1062, Borsboom, D et al. (2004).

- As mentioned above, the standard we used is based on a somewhat different philosophical approach. As shown in the added summary, “Lines of validity and reliability testing Standards for educational and psychological testing” (AERA, 1999), based on this approach, our test of the psychometric properties of the BERNCA refers primarily to the instrument’s validity.

In several instances the authors refer to dimensions of rationing, but the dimensions are not clear stated or defined. Table 1 refers to reliability and validity testing by using research questions and hypotheses. These should be identified as specifically addressing reliability or validity. Post-hoc factor analysis is insufficient to ascertain validity as it relies on correlations, which do not constitute validity. Tying the research questions and the hypotheses more clearly to the text in the manuscript would strengthen the article.

- The original table 1 has been deleted and the content integrated into the text, under the new heading, "Validity and reliability testing of the BERNCA", taking into consideration whether they address validity or reliability.

Literature: It is difficult to determine if the cited literature is appropriate as much of it is not directly accessible, due to incomplete references (Kuenzi et al 2002; Aiken et al 2002); unpublished (Schubert et al, 2003); or in another language (numerous). Others are briefs or opinion papers (Aiken et al. 2003; Nocera, 2001; Ward, 2005). This literature review needs to be more rigorous and provide sufficient information to ensure that the references are accessible to readers, especially when some key papers referred to for the substance of the study are opinion papers or unpublished.

- We performed a comprehensive review of the literature. In the end, we decided to include the non-English papers because they provided, in addition to the few available English publications, the strongest support for the issue of implicit rationing of nursing care. The English papers focus mainly on the rationing of medical services.

- Reference Aiken et al. (P 5, L 21, previous version): we accidentally provided the wrong reference. This has been corrected (see P 6, L 11). The paragraph which included the reference of Nocera, 2001 has been revised and other references included (P4, L 15-19)

Methods: Given the use of the NWI-R (Aiken & Patrician, 2000) in this study to assist in validation of the BERNCA instrument, it would be appropriate to discuss implications of Cummings et al (2006) Nursing Research article that raised questions about the validity of Lake's scales and subscales when the underlying theory was tested and failed to fit the data. NWI-R

- On P 16, L 18 - 22 we have taken the questions raised about the validity of the NWI-R Lake's scale into consideration.

Results: See comments above regarding Table 1 and discussion of validity and reliability procedures.

Table 3 Instrument: The wording of the items on instrument are problematic.
- 5a) Go over is not clear. Does this mean review individually, or in discussion and collaboration with another nurse?
- 4a) …by a physician or as you felt was necessary. Whether required work is based on "feelings" or critical thinking assessments is critical to both the validity and reliability of this question.

- *4a is written in a positive sense, whereas 4 c, 4d, and 4e are reverse-coded (written in a negative sense). Yet this (nor how it was coded) is not referred to in the instrument or the manuscript.*

These problematic wordings raise questions about the results presented in Table 3. The development and presentation of a concept map that ties the theory to the indicators would strengthen this study.

- In the original table 3 only an abbreviated form and not the complete form of the questions was presented. We have changed this and the questions are now presented with their original wording in table 1. As can be gathered from the original wording of the questions, questions 4a, 4c, 4d, and 4e are written in a negative sense.

- The development of the questions was based on the literature review and experts' knowledge and experience, which is influenced by local and cultural factors specific to the respective regions / countries. This might restrict the application of some items to other countries and cultures.

Reviewer #2:
The manuscript entitled "Validation of the Basel Extent of Rationing on Nursing Care Instrument" (BERNCA) is on a timely subject. With the nursing shortage and other financial constraints in health care, having an instrument that could be used to measure the extent of rationing of different aspects nursing care among health care facilities is important.

The background of the study is well developed and the study uses appropriate measures to determine concurrent validity as well as evaluation of other aspects of validity and reliability of the instrument. However, I would suggest the statement of an explicit conceptual framework, rather than just the framework implied from the work of Aiken and colleagues and the ideas of explicit ad implicit rationing.

- Thank you too for your valuable comments and input. On P 5- 6: A description of the conceptual framework of implicit rationing of nursing care has been added

I would like more discussion of the "unexpected" finding related to nurse-patient ratios mentioned on page 13, line 16. As the author(s) state, I am not sure that the measure of "last-shift workload" accurately captures the concept of nurse-patient ratios. However, is there another explanation that indeed, rationing is not a function of nurse-patient ratios?

- The “unexpected" finding related to “nurse-patient ratios” is addressed on P 17, L 1- 15

The suggestion that the instrument needs to be validated for different cultures and countries is appropriate and an area in need of future research.

A couple minor editorial comments:
---Table 2 is referenced the 1st time in the text after Table 3.

- Table 2 is cited at P 11, L 23.

--Pages 9-11 have several single sentence paragraphs

- The single sentence paragraphs P 9 - 11(now P 10-12) have been changed.

---The discussion section has an excessive use of the pronoun "it."

- The excessive use of the pronoun "it" in the discussion section has been changed.

---Reference to Table 4 needs to be added to the Text Page 13 line 17.

- Table 4 has been deleted; we now refer only to the one-factor solution.

---The OCED reference is incomplete on the reference list.

- The OECD reference has been deleted and another reference has been inserted.

---Table 3 needs a KEY to explain that the factor scores underlined are ones that loaded on more than one factor.

- Table 3 has been deleted. The results of the most recent factor analysis are now presented in table 1.

Reviewer #4:
The focus of this manuscript is to describe testing of a 20 item questionnaire designed to measure "implicit rationing of nursing care". This review will focus on the steps of the research process and provide a critique.

Problem

The research problem appears to be that nursing care is being rationed in some countries with an impact on patient outcomes. However, as there was no tool to measure rationing, the level of rationing was not identifiable. The problem statement becomes identified on P5L16-24. Earlier in the manuscript, the problem of rationing was discussed and the cause of the rationing P2L24-9. While this reader feels that the literature provided on the rationing problem and causes appears biased, the issue is the development of a clear problem for this research study. The study problem needs to come earlier in the manuscript, as well as the definition for implicit rationing, which does not appear until P3L1-2. The reader would recommend greater problem clarity, and earlier identification in the manuscript.

- Thank you for your comments and the input. The manuscript has been reorganised. The study problem is described on P 4, L 5-13. The Definition of implicit rationing of nursing care is been included in the new paragraph “Definition of implicit rationing of nursing care” P 4 L 20-23. We have not moved the study problem and definition more forward, because in our opinion it digresses from the flow of the article. The conceptual framework, which immediately follows the definition of rationing, is been described on P 5 – 6.

Purpose

The purpose of the study is to test the reliability and validity of the 20 item questionnaire. This is clearly stated on P6L1-2. However, it is not clear why the author selected the standards reference, and why these standards were particularly important. These could be mentioned later to avoid confusion for the reader, who is likely aware of standards for reliability and

validity. Table 1 present research questions and hypotheses. The column "evidence based on" is difficult to understand and the reader suggests that only the research hypotheses be presented. H1 is not in the format of a scientific hypothesis. H2 did not include the strength of the correlation expected, while H3 did include the word "moderate". H4 is actually two hypotheses.

- We have chosen the "Standards" to test the validity and reliability of the BERNCA, although we are aware that other philosophical approaches use other systems to establish validity and reliability of measurement tools.

- The content of table 1 has been integrated into the text under the new heading "validity and reliability testing of the BERNCA". The formulations of the hypotheses have been adapted in accordance with your suggestions, except hypothesis 4. We left this as a single hypothesis, because, according to our philosophical view of validity the content is strongly related.

Review of Literature

The review of literature comes at the beginning of the article, and is involved in describing the extent of the nursing shortage. Select words and concepts are inappropriate, such as "along with substitution of outpatient for inpatient care" P2L9 as a cost saving strategy. P2L16 includes the words "necessary care" provided as often as recommended. How can care be necessary yet recommended? P4L1 uses the words "good nursing care" which infers a value judgement on a task. This is objectionable. P4L23 uses the word "neglected". This reader also objects as no nurse intends to neglect a patient.

- The selected inappropriate words / concepts (P 2, L 9 (now P 2, L 7); P 4, L 1 (now P 2, L 2)) have been changed. The paragraph which included P 2, L 16 has been revised (now P 4, L 15 -19). The Paragraph which included P 4, L 23 has been removed.

Missing from the review of literature are other ways of measuring rationing or inadequate nursing care. Specifically, much work has been done on acuity systems, one way that care is rationed.

- We have included some empirical evidence from the acute care setting which indicates the existence of implicit rationing. To our knowledge only one qualitative study focuses on other ways of measuring implicit rationing of nursing care (Morin & Leblanc, 2005). However, this study was conducted in the long-term care setting. The majority of articles available focus on the rationing of medical treatment and services. This does not provide strong support for the issues of this study (see P 2, L 13 -15).

Methodology

The study is a secondary analysis of data which is appropriate. The sample restricts the generalizability of the findings, geographically, employment area and convenience sample. It is noted that the nursing sample is very young, this may have impacted the results as their may be generational differences with respect to rationing.

The organization of this section would be improved if the development of the instrument items were discussed first, then the Sample followed immediately by data collection.

- The organization of the “methods” section has been changed: the development of the instrument is now discussed first (P 6-8). Because this study is focused on testing the validity and reliability of the BERNCA we do not explore the possible effects of the young nursing sample on the results. This will be part of a later study we are preparing on this topic.

Instrument

P7L2-6 describe the conceptual framework for the instrument, however, this framework is not described nor presented. If this framework drove the instrument then the reader needs more knowledge about this framework to judge its applicability.

- A description of the conceptual framework of implicit rationing of nursing care, which provides the framework for the instrument, has been added P 5-6.

P7L10 state "Qualified nursing specialist", yet goes on to state that some of these were students. What were the characteristics by which these nurses were qualified and/or specialists in what field? What were these experts told about implicit rationing? How many specialists were involved? Were there also nurse administrators that may have a different paradigm of rationing?

- The characteristics and number of qualified nursing experts were specified on P 7, L 6-11. On P 7, L 11 - 14 we specify what we asked the experts to do. Nurse administrators were not included in the experts’ sample.

The first scale consisted of 20 items, the final scale 20 items, it is unknown if these items were the same. This reader could, with little effort, think of additional tasks or behaviors that may be omitted if nursing care needed to be rationed, for example, preparing a patient for surgery, reviewing the chart to detect abnormal lab values, administering a medication on time, reconciliation of medications etc.

- The number of items comprising the first scale and final scale are described on P7, L 4-18.

- We acknowledge in the discussion section (P 15, L 8-20) that the BERNCA does not address the entire domain of nursing care activities which might be omitted. In the revised version of the BERNCA we will add several of tasks / behaviours you suggested above.

Some of the items are very focused, e.g. Keep a patient waiting for longer than 5 minutes, while some of the items are very broad e.g. Monitor a patient. Items under domain 2, would be delegated to the nursing assistant, and less likely to be thought of as rationing by a nurse. This reader is unclear as to what Item 4c refers to with respect to rationing. Conceptually, more information is needed as to how these items relate to and measure "implicit rationing”.

- We are aware that some of the items have different levels of abstraction. As already mentioned by reviewer 1, the development of the items was based on the literature review and experts’ knowledge, which is influenced by the local and cultural factors of the respective region or country and might restrict the applicability of some items to other countries and cultures.

The scale instructs the nurse to indicate 'how often they had been unable to carry out" each of the tasks (P8L1-2)" on a ordinal scale. This then appears to be the author's definition of nurse rationing, that is, the inability to carry out a task. This definition is different than the one stated earlier (P3L2) which includes the word "necessary". A nurse may be unable to carry a task out, but may do it because that task was not necessary. As noted earlier, further clarification of the implicit rationing concept is needed to understand the items and qualities of the instrument.

- P 8, L 1-2 and P 3, L 2 both refer to the same definition of implicit rationing of nursing care, which was presented on P 3, L 2. The English description of the scale instruction has been reformulated (P 7, L 18 - 21).

Data analysis

Factor analysis indicated 3 factors based on eigenvalue and scree plot. It is essential that the eigenvalues and plot be provided for the reader, as well as all data related to item loadings P10 Lines 18-25 are very confusing to this reader. One dominant factor has one eigenvalue (the manuscript states "each eigenvalue). If only one dominant factor resulted, then why would 5 factors be analyzed? Clinical considerations aside, perhaps the conceptual framework is inappropriate. The authors ignore the one factor solution. With the use of 5 factors, it is not clear that these 5 factors are equal in their contribution to a total score. With no support that they contribute equally, the total score, as a mean of the subscales is not valid.

- P 10, L 18-25 this paragraph has been changed (see P11 L 10-20). Based on the eigenvalue and screeplot, which have shown one dominant factor we have put the focus now only on the one-factor solution. The eigenvalue (inclusive a quotation of the value) is still described in the text. We have not provided the eigenvalue and plot, since the manuscript is already very long and focus was put on the one-factor solution (see result section P 13, L 1-5)

The means of each item (Table 3) seem to indicate a "floor effect" of the item, which can be a problem with the collection of ordinal data.

- The flow effects are described and discussed respectively on P 12, L 16 of the result section and P 16, L 1-5 of the discussion section.

Discussion

The discussion section is limited in its depth. The author tends to reinforce the findings, not explain the results or place them in the context of previously described literature. The discussion contains limitations, but not explanations. The discussion did not refer to the hypotheses.

- The discussion section has been changed and elements added, such as the value of the BERNCA instrument for clinical practice and outcome research. The results are explained and discussed based on the developed conceptual framework and the literature.

Recommendations/Conclusions

The conclusions are appropriately tentative and restricted to Swiss nursing practice. The one factor/5 factor discrepancy seems to be glossed over, with a summed mean score of the 5 factors promulgated. This reader would not accept this conclusion.

- The conclusion section has been changed. As described above, the focus has been set to a one-factor solution and the section reworked accordingly.

Organization

The organization of the manuscript, while following the research process, has sections that need review. If research questions are asked or hypotheses are to be tested then this reader expects the results and discussion section to be so organized. The problem needs to be clearly articulated before the literature is presented.

- The manuscript has been reorganized. The results and discussion section have been organized in accordance with the validity testing line and the hypotheses tested.

Standards for educational and psychological testing (p 9– 35) (AERA, APA, NCME, 1999)

Validity

Evidence based on test content refers to the extent to which test content represents the content domain it is intended to measure. This evidence is obtained by analysing the relationship between test content and the measured construct. To do this, a logical or empirical analysis is necessary regarding the adequacy with which the test content represents the content domain, as well as the relevance of the content domain to the proposed interpretation of the test score, and /or expert judgment of the relationship between parts of the test and the construct.

Evidence based on response process results from an analysis of individual responses, which are collected either by questioning test takers about their performance strategy or responses to particular items, or by monitoring the development of a response to a writing task. Analyses of the relationships between parts of the test and between the test and other variables can also develop inferences about the performance process. Such evidence can contribute to differences in opinion or interpretation of test scores across subgroups of examinees.

Evidence based on internal structure refers to the degree to which the relationships between test items and test components conform to the construct on which the proposed test score interpretation is based. This evidence therefore results from an analysis of a test's internal structure.

Evidence based on relations to other variables refers to the nature and extent of the relationships between test scores and variables measuring similar or different constructs (resulting in, respectively, convergent or discriminant evidence). This also refers to evidence of the relationship between the test score and any relevant criterion which indicates the test score's accuracy at predicting performance of that criterion (predictive or concurrent evidence).

Evidence based on consequences of testing refers to an evaluation of the consequences (intended and unintended) of test use on the concept's validity. In cases where differential consequences of testing are observed for different identifiable groups, it is important to distinguish between issues of validity and issues of social policy.

Reliability & measurement errors

Reliability refers to the consistency and precision of the measurement process as it pertains to test results. High test score quality implies high consistency and freedom from measurement error (interscorer difference, time sampling error, interitem consistency, and content heterogeneity). Critical information on reliability should include the major sources of error, summary statistics, and the degree of generalizability of scores across alternate forms, scorers, administrations or other relevant dimensions, including descriptive statistics of the examined population (reliability coefficient, standard error of measurement)