Author's response to reviews

Title: Patients' functioning as predictor of nursing workload in the acute hospital: a multi-centre cohort study.

Authors:

Martin Mueller (martin.mueller@med.uni-muenchen.de)
Stefanie Lohmann (Stefanie.lohmann@med.uni-muenchen.de)
Ralf Strobl (ralf.strobl@med.uni-muenchen.de)
Christine Boldt (christine.boldt@paranet.ch)
Eva Grill (eva.grill@med.uni-muenchen.de)

Version: 2 Date: 30 June 2010

Author's response to reviews: see over
Dear editors of BMC Health Services Research,

Thank you very much for your comments on our manuscript "Patients’ functioning as predictor of nursing workload in the acute hospital: a multi-centre cohort study " and for giving us the opportunity to revise our paper.

Please find enclosed our detailed answers and changes to the reviewers’ comments. We have revised the manuscript accordingly and we hope that the paper is now acceptable for publication.

All authors have been involved in the revision of the manuscript and agree to be co-authors of the revised manuscript.

Sincerely,

Martin Mueller


**Answers to Reviewer 1 (Sara Laxe)**

**Explain second paragraph (lines 11-18 in background) because it may seem a bit confusing.**

We rephrased and shortened the paragraph as follows:

> There are various criteria to determine if the scale and composition of nurse staffing are adequate to meet the needs of patients and staff [7]. In general, these criteria include factors such as nurses’ educational level, ward size or number of beds, acute versus chronic status of patients, and the time required for individual nursing interventions. Regression analysis has sometimes been used to account for the most important determinants, e.g. bed occupancy. Various methods are available for quantifying the nursing workload, which is arguably one of the most important management tools to optimize staffing decisions. Irrespective of the method employed for calculating staffing requirements and for accurate prediction of the consequent workload, the relevant factors determining workload must first be determined [8].

**I suggest to explain in a table the 84 items of the LEP.**

We added Table 1 explaining the 84 items.

**Review the “Results” regarding mean age of two groups, seems it is missing a part.**

The reviewer is right that this is important information. However, we felt that reporting mean age or other patient characteristics for the subgroups in the results section would be tiresome and not add to the clarity of the manuscript since patient characteristics per stratum are also reported in table 3. This is indicated at the end of the first paragraph of the results section:

> Patient characteristics stratified for condition group are presented in table 3.
Answers to Reviewer 2 (Caitlin Brennan)

On page 4, the authors clearly state that nursing workload is a measure of patients’ individual health status, functional status, need for nursing care, and severity of symptoms, and that the Barthel Index is a predictor of increased workload, in terms of the degree of patients’ functional dependence. However, a stronger argument for use of the ICF instead of the Barthel Index is needed. What are the specific limitations of the Barthel Index that the ICF would improve upon?

What is the rationale for testing the ICF, which is another measure of patients’ functioning, rather than testing an instrument that measures another aspect of workload, such as need for nursing care or severity of symptoms? The authors mention that perhaps the ICF could provide a more parsimonious measure of workload than the Barthel, but that argument needs to be more clearly stated by presenting how burdensome the Barthel Index is and how the ICF lowers that burden.

Our apologies for not making this completely clear. We amended the third paragraph of the background section as follows:

In addition, many instruments like the Barthel Index use summary scores, which summate the scores of single items, so as to describe the extent of functioning and disability. Such summation scores are known to mask the influence of particular aspects of functioning and disability reflected by particular Barthel Index items. This masking may lead to imperfect assessment and conclusions, which would decrease the usefulness of the Barthel Index for clinical practice [16]. Furthermore, the Barthel Index tends to reduce the complex picture of human functioning to a consideration only of motoric aspects of self-care, without consideration of the relevant cognitive aspects [17-18]. Therefore, it is of great practical interest to explore which aspects of functioning and disability are drivers for nursing workload, with particular consideration of aspects scarcely reflected in ADL scores, such as impairments in body functions and structures, or modifying contextual factors.

The new references are

Similarly, if the LEP measure of nursing workload explained over 80% of the variation in workload and according to the authors provides a parsimonious estimation of workload, why is testing of the ICF needed?

We found that an approximation, i.e. the average of two hospital days (LEP workload on the first and last hospital day) multiplied by length of stay was a good estimate of total workload as summed up by detailed day-to-day records of LEP workload. The reviewer is right in stating that this is an interesting result. This is, however, just an estimate of total workload which might be useful to simplify hospital records. Yet, this does not give any information about the determinants, specifically, the patient related determinants of workload which was the aim of our study.

**Why were t-tests assessed for significance at the 0.2 level, rather than the traditional cut-off of 0.05?**

This is a very reasonable comment. In fact, in correspondence to the comments made by Reviewer 3 (Prof Datta) we carried out the LASSO procedure with all covariates without serial selection based on t-tests, because the LASSO procedure is ideally suited to deal with unselected, potentially collinear covariates. Therefore, the t-tests are no longer part of the current analyses. Please see below for the corresponding modifications.

**On page 19, the face validity of ICF models was mentioned - how did the authors go about obtaining face validity?**

We agree with the reviewer that the expression “face validity” might be misleading in this context. We modified the paragraph which now reads as follows:

> We found that the selections of ICF categories predicting nursing workload differed according to the underlying health condition, and were in line with the current literature [40-41].


**In the United States, most hospitals employ unlicensed assistive personnel, who assist patients with ADLs, such as dressing, bathing, ambulating, urinating/defecating, and eating. The authors found that these activities had the most influence on nursing workload, but these activities are not performed by**
nurses in the United States. Do the hospitals where this study took place employ unlicensed assistive personnel to perform these duties or are they performed by nurses?

In Germany, Switzerland and Austria, assistance and compensation in self-care activities are tasks usually performed by registered nurses. To clarify this issue, we added the following sentence to the first paragraph of the limitations section:

Among the limitations of the present study, it must be conceded that the high face validity of the presented results from the obvious association between limitations in patients’ activities, especially in self-care tasks, and the requirement for additional nursing interventions and modifications in treatment procedures. Since duties performed by nurses may vary between health care systems, some caution should be applied in generalizing conclusions with respect to country-specific responsibilities of nurses.

On page 22, the authors state "Nevertheless, a sample size of 50 should generally be adequate for estimating linear models with three independent variables." Please include a reference that supports this statement, as typically a larger sample size is needed for models with three independent variables.

The reviewer is correct to point this out. The LASSO procedure is a cross-validation resulting in an optimal choice of variables for this sample size. A larger sample size might have resulted in a larger set of variables. We deleted the sentence and included the following paragraph instead:

However, the ideal number of independent variables was determined by cross-validating LASSO models, such that the final models with their included independent variables were validated on independent data sets, and proved to be superior to the other candidate models. As such, the number of variables was optimal under the given circumstances and given the available sample size.

All of the adjusted R-squares have non-significant p-values, but this is not addressed in the text.

We apologize for not being clear in the tables. The values in the last row of tables 4 to 6 are R-squared and adjusted R-squared (in brackets). The adjusted R-square is a measure of explained variance that is corrected for the number of variables in the model. To avoid confusion, we replaced ‘Proportion of total variance explained by the model’ by ‘R-squared (R-squared adjusted for number of variables in the model)’. We do not report p-values for R-squared.

Minor Essential Revisions

There are several grammatical errors throughout that thorough editing will improve.
Figure 1 is difficult to follow and is duplicated in the document.
We checked the text again for grammatical errors. Figure 1 was improved.
Discretionary Revisions

A more thorough explanation of the LASSO technique and a rationale for its use instead of traditional methods would be helpful.

We have rewritten the fourth paragraph to the methods section:

Step 3: In order to improve prediction accuracy, and to extract small subsets of independent variables seemingly with the strongest effects on the dependent variable, we used the least absolute shrinkage and selection operator (LASSO) [37] to identify which of the ICF categories are independently associated with the log-transformed workload. The LASSO minimizes the residual sum of squared errors, with a bound on the sum of the absolute values of the coefficients. To avoid large variance, which often arises from ordinary least square regression, the LASSO sets some regression coefficients to zero and shrinks others based on a preset regularization parameter, the so-called penalty. Thus, the method acts as a tool for identifying valid subsets of ICF categories. The penalty value was chosen by leave-one-out cross-validation [38]. By this process we obtained five subsets of predictor variables from each of the five imputation data sets.
Answers to Reviewer 3 (Susmita Datta)

Minor Comments:
1) Spelling of co-linearity should be changed to collinearity.
We have changed this accordingly.

2) All the Tables and figures should be included while describing the results.
We have reviewed the manuscript accordingly.

Major Comments:
1) "If the t-test was not significant on the 0.2-level" in Step 2, Page 13 why is the significance level so high here? Is this alpha or false discovery rate after multiple hypotheses correction for all the covariates (ICF variables)?
This is an important comment. In fact we followed the reviewer’s suggestion and omitted the t-test. Please see our response to Major Comment 3 for a detailed response.

2) How were the five imputation datasets constructed? Why 5 imputation datasets and why not 6?
Our apologies for not being completely transparent about that. To clarify this issue, we rewrote ‘step 2’ in the ‘Methods/Statistical analyses’ section:

   Step 2: Missing values in the variables ‘age’ and in the ICF categories were replaced by multiple imputation, using the function ‘mice’ of the software package ‘mice’ of R 2.11.0 [33]. This procedure uses Gibbs sampling for multiple imputations for incomplete multivariate data. Multiple imputation is a useful method for reducing bias and increasing precision when data matrices are incomplete [34]. Multiple imputation has several advantages over complete-case analyses, wherein deleted observations lead to diminished precision of the outcome measure, resulting from the reduction of sample size and consequent bias, when the missing data are not randomly distributed. Simulation studies demonstrated that with five generated data sets, multiple imputations yield valid results [35-36]. Thus, we generated five data sets with differently imputed variables. All further steps of the multiple analyses were carried out with each of the five imputation data sets.

The corresponding references are:

Authors have selected 16 ICF variables out of 45 ICF variables after doing t-test for significant associations and then used LASSO on those 16 ICF categories, which is strange. As they have used LASSO correctly to include numerous covariates with collinearity why not use them on all 45 ICF variables? If they pre-screen variables with t-test they automatically assume that the covariates are independent which are definitely true in this case. So authors are advised to use all 45 ICF variables at the same time.

I would comment similarly on the Barthel Index as well.

This is a very important comment corresponding to Major Comment 1 and the comment from Reviewer 2. The reviewer is completely right to point out that the t-test to screen for variables to enter the final model is not needed once LASSO is applied. We reanalyzed our data according the Reviewers’ suggestion. As a result, the resulting final sets of variables have slightly changed. We have modified methods, results, and discussion section accordingly.

Authors must plot the fitted value vs. observed value for the sum of the LEP variables to see the accuracy of the fit.

We added these plots as figure 2, 3 and 4 and added the sentence

Accuracy of the fit is visualized in figure ....

at the end of each 'Model with ICF categories'-paragraph in the results section.