PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Balancing nurses' workload in hospital wards: study protocol of developing a method to manage workload
AUTHORS	van den oetelaar, wilhelmina; van Stel, Henk; Van Rhenen, Willem; Stellato, R; Grolman, Wilko

VERSION 1 – REVIEW

REVIEWER	Peter Van Bogaert University of Antwerp, Belgium
REVIEW RETURNED	19-Apr-2016

GENERAL COMMENTS	The study protocol describes a very relevant research subject underpinned by a well-written introduction as well as a thoroughly described study method. I have two remarks about the study methodology: (1) maybe the authors could consider to use the well-known and validated "NASA Task load Index". This interesting 6-item workload measurement instrument developed by the NASA was used in various studies in healthcare and nursing settings*; (2) the authors will do a lot of effort to determine relevant patient characteristics to predict care time. I suggest to reflect about the transition of hospital care the next years about the changing and complex patient needs, decreasing length of stay, changing organization approaches inspired by lean management or releasing time to care, patient involvement … You need to develop a flexible system that could be adjusted not only for other specialities as described in the discussion section but also robust to relevant changes. Please check the sentence at page 11 line 50 "the method …" * Hoonakker P, Carayon P, Gurses A, et al. Measuring workload of ICU nurses with a questionnaire survey: the Nasa Task Load Index (TLX). IIE Trans Healthc Syst Eng 2011;1:131–43.

REVIEWER	Walter Sermeus KU Leuven Institute for Healthcare Policy, Belgium
REVIEW RETURNED	30-Apr-2016

GENERAL COMMENTS	The topic of nurses' workload measurement was an important topic in the 1980's and 1990's, disappeared during many years and seems to return now. There are several reasons for it. Obvious reasons are the impact on nurse staffing on quality and patient safety (that are documented more recently) and new methods of

	measurement of time of nurses given new technology (smartphone, tracking devices, sensors). The authors of this paper are returning to the old school approach of NZi of the 80s in performing time-consuming work sampling studies by an external evaluator. The problems are the accuracy of the observation (interrater reliability), the bias induced by having an external observator watching your work, privacy issues to nurses as well as patients. The choice of these systems should be well documented, given that new more sophisticated observation methods are available. The critique given by the authors is the standard critique on nurses' workload systems: the classification system is not precise enough, so they choose for a new method that will solve all problems, at final end, this new system is not solving the problem at all, and they choose to develop a new system. This is exact also the critique here. The authors should be more specific what they will add to the existing knowledge. They claim to be more prospective. This is indeed one of the major flaws of patient classification systems. It is easy to measure a patients' condition retrospectively, but what to do prospectively? It seems that patients' condition is hard to predict. My suggestions for the paper: 1. The introduction is very general, while the research is only involving surgical units (and not e.g. intensive care): be more specific 2. The paper is very focused on a specific situation of UMC Utrecht. It should be written in a more general way. UMC Utrecht can still be chosen as convenience sample. the context should be better described. In the limitations, the choice for UMC Utrecht should be made more explicit. 3. Be not too ambitious in the aims. my suggestion is that prospective use of the data is definitively is a later project 4. Why using the Nzi method (1988)? Why not using more advanced data collection methods instead of labor-intensive work sampling methods ? please clarify 5. The delphi-method is limited to representative of the 6 six units. The use of the panel is probably not a delphi-panel, but rather a steering group. Please make clear if this is a delphi-panel (experts? why only internal?) or if it is a steering group (tasks). 6. Sample size for the work sampling should be calculated. Different scenarios could be used given the fact that the number of activities are not known yet. 7. The calculation of proficiency percentage should be better documented. It seems that it will be based on expert opinion by the ward manager. Will it be on individual base? Or contextual?

VERSION 1 – AUTHOR RESPONSE

Editor Peter Van Bogaert:

Comment 1: Maybe the authors could consider to use the well-known and validated “NASA Task load Index”. This interesting 6-item workload measurement instrument developed by the NASA was used in various studies in healthcare and nursing settings*

• In our study we use an extensive and validated questionnaire (QEEW, see line 380) to measure work engagement before and after the introduction of the workload management method. This questionnaire measures a broad array of resources and demands and gives good insight on dimensions for improvement. The shortened questionnaire that we use to make a daily measurement of engagement and workload is a subset from this larger questionnaire. We would like to validate this shortened questionnaire in the hospital setting. This shortened instrument has quite some overlap with the index that you suggest, for example where measuring mental and physical demand and

temporal demand and performance. These aspects are used in our tool as well, with quite similar questions. It would be interesting though to include frustration level, however we do incorporate work-related absenteeism, tension, fretting, confidence and enthusiasm via the extensive questionnaire.
Comment2: The authors will do a lot of effort to determine relevant patient characteristics to predict care time. I suggest to reflect about the transition of hospital care the next years about the changing and complex patient needs, decreasing length of stay, changing organization approaches inspired by lean management or releasing time to care, patient involvement … You need to develop a flexible system that could be adjusted not only for other specialities as described in the discussion section but also robust to relevant changes.
• We agree with this comment: if major changes or interventions occur in working processes, treatment types, staff organization or environmental factors, the patient characteristics list will need to be updated and time study will need to be repeated. We have updated the discussion accordingly, see line 471-476. We plan to update our workload management method after the first analysis results are in. We expect these results to lead to organizational changes (adjustment of working processes that may not be efficient) and therefore we expect that we need to repeat the time study before using the results in a workload planning tool.
Editor Walter Sermeus:
Comment 1:
The introduction is very general, while the research is only involving surgical units (and not e.g. intensive care): be more specific
• We have adjusted the introduction in line with this comment (3rd paragraph, we deleted literature that was not relevant)
Comment 2:
The paper is very focused on a specific situation of UMC Utrecht. It should be written in a more general way. UMC Utrecht can still be chosen as convenience sample. the context should be better described. In the limitations, the choice for UMC Utrecht should be made more explicit.
• We have adjusted the manuscript in line with this comment (introduction 1st paragraph, discussion line 477-481)
Comment 3:
Be not too ambitious in the aims. my suggestion is that prospective use of the data is definitively is a later project
• We understand this and have adjusted the manuscript in line with this comment (introduction line 91-95, method line 318-321): this will indeed be addressed in a later phase in the project, if results have proven to be useful for prospective planning
Comment 4:
Why using the Nzi method (1988)? Why not using more advanced data collection methods instead of labor-intensive work sampling methods ? please clarify
• We have carefully considered this choice as well and we appreciate prof. Sermeus' thoughts on this. Indeed we have chosen to put in quite some work to perform the time study. We have made some choices in the set-up of the time study though that we expect to help us in acceptance of the end results. For example many of the observers in the time study will be nurses of departments that take part in the study. These nurses observe in other wards than the one they usually work in. This will help them understand other ways of working than they are used to and it also is expected to help appreciate challenges on other nursing wards that they are not familiar with now. This is a big bonus for us, since we expect to realign working processes as a result of this study. This will be easier if nurses have experienced working processes on other wards. Also, because many observers are nurses of the involved wards, we expect that they will accept the end results of the time study as accurate.
We acknowledge the bias that can be introduced by putting nurses under observation, but that bias is also present when nurses self-register or when their movements are tracked by an automatic device. Also, tracking movements or patterns does not yet determine what the nurse is actually doing: they will still need to register that themselves or have an observer do that. We are not aware of any

systems that can automate this step. As mentioned in our manuscript in line 215, self-registration has downsides as well. Because we will choose nurses and medical students as observers, we think this will reduce some of the bias (or mistakes that would be made) in determining the activities. But aside from these assumptions, we will also quantify inter-rater agreement in our study (line 286-296), so we can incorporate this in the interpretation of the results.

Comment 5:

The delphi-method is limited to representative of the 6 six units. The use of the panel is probably not a delphi-panel, but rather a steering group. Please make clear if this is a delphi-panel (experts? why only internal?) or if it is a steering group (tasks).

• We did not describe this clearly enough in our study protocol. We have adjusted this in line 143. We have performed a Delphi study to compose the list of patient characteristics. We propose to first develop a list of relevant patient characteristics for the 6 surgical units that were involved in the study. If we are successful at this, then we can extend the study to a larger setting and use the results of this Delphi study as a starting point for another study. This is why we only include representatives (experts) of the involved nursing wards for now. The process is indeed a controlled Delphi process, hence we name the participants a Delphi group.

Comment 6:

Sample size for the work sampling should be calculated. Different scenarios could be used given the fact that the number of activities are not known yet.

• Based on a previous time study of 10 day shifts in 2003 on 3 of the same wards at the UMC Utrecht, we estimate that we will define between 25 and 30 activities. This 2003 study worked with 23 activities, which are all still relevant today. We do miss items such as handover though, so we assume we will determine a few more activities in this study. The database from the 2003 study is lost unfortunately, but the reports on the overall results are still available. From the 2003 time study we know that the most observed activity accounted for +/- 9% of observations in 2003. This translates for our study into a maximum of +/- 4.860 observations per activity. The minimum is estimated to be 0.7% of observations, which translates into a minimum of +/- 378 observations per activity We have added this information to the method paragraph, line 271-278.

Comment 7:

The calculation of proficiency percentage should be better documented. It seems that it will be based on expert opinion by the ward manager. Will it be on individual base? Or contextual?

• We will ask ward management (head nurses) for their expert opinion to define a proficiency % of all less experienced nurses (<1 year experience on the ward they work in) and student nurses (for 2 education types known in the Netherlands and the status of their education: senior or junior students). They are asked to determine proficiency of these type of nurses against the 100% standard. Determination of proficiency for individual nurses in the study will be done by classifying nurses in this framework. We have added this information to the manuscript, line 348-352

VERSION 2 – REVIEW

REVIEWER	Walter Sermeus KU Leuven - University of Leuven
REVIEW RETURNED	28-Jul-2016

GENERAL COMMENTS	The authors have incorporated most of the recommendations of the first review. Funding for the study is not clear and should be specified (assumption is that UMC Utrecht is funding the study - internal funds). Limitations are well addressed. The major limitation is that the study is limited to 1 hospitals (6 surgical wards). The ambition that it will solve all major problem on measuring nurses' workload should be tampered. The variables that will be measured are well described. The final model of how these variables are related and how this information can be used for predicting nurses' workload is not specified and remains open. Although the

	importance of quality of care is mentioned in the first sentence, quality of care is not operationalised as part of the model.

VERSION 2 – AUTHOR RESPONSE

Editor prof. Sermeus:
Comment 1:
Funding for the study is not clear and should be specified (assumption is that UMC Utrecht is funding the study - internal funds).
• Indeed this study is fully funded by the UMC Utrecht. We have entered this information via the BMJ Open web application
Comment 2:
The ambition that it will solve all major problem on measuring nurses' workload should be tampered.
• We appreciate this comment and have again looked at the introduction and discussion paragraphs of our manuscript and made adjustments to rephrase our ambitions. (introduction line 91-98 and discussion lines 480 and 502)
Comment 3:
The final model of how these variables are related and how this information can be used for predicting nurses' workload is not specified and remains open.
• This is correct, we have not specified this completely yet, this is dependent on our findings. However we do have an idea on how to predict workload, based on the measurements that we plan to do: for the patient characteristics where we find a relationship with care time, we will keep a database where we save all registrations of these characteristics over time. Based on this historical data, we plan to make patient group profiles, where we will group patients according to diagnosis or treatment type. We will specify which characteristics we expect to be present on each day during the expected length of stay for this patient group. We will use the OR schedule to predict which type of patient enters the ward on which date. We have added more information on this in our manuscript (line 321-325)
Comment 4:
Although the importance of quality of care is mentioned in the first sentence, quality of care is not operationalised as part of the model.
• Studies (reference 1-4 in our manuscript) have shown a direct relation between nurses' workload and patient outcomes and nurse reported quality of care. In our study, we will not focus on finding additional information on this relation. We will do baseline measurements of patient experiences (also about quality of care) and nurse-reported quality of care and will repeat this after working for a longer period of time with the new workload management method. In this way we hope to trace the effect of our new focus on nurses' workload on perceived quality of care.