

# Testing Reliability and Validity of the Oulu Patient Classification Instrument—The First Step in Evaluating the RAFAELA System in Norway

Marit Helen Andersen<sup>1\*</sup>, Kjersti Lønning<sup>1</sup>, Lisbeth Fagerström<sup>2</sup>

<sup>1</sup>Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway

<sup>2</sup>Department of Health Sciences, Buskerud and Vestfold University College, Buskerud, Norway

Email: \*[marit.andersen@ous-hf.no](mailto:marit.andersen@ous-hf.no), [kjersti.lonning@ous-hf.no](mailto:kjersti.lonning@ous-hf.no), [lisbeth.fagerstrom@hbv.no](mailto:lisbeth.fagerstrom@hbv.no)

Received 10 February 2014; revised 25 March 2014; accepted 11 April 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

**Objective:** To study reliability and validity of the Finnish Oulu Patient Classification instrument in Norway. **Background:** The Finnish patient classification system RAFAELA consists of three parts: 1) daily patient classification of nursing intensity using the Oulu Patient Classification instrument, 2) calculation of nursing resources providing bed side care per 24 hours, and 3) Professional Assessment of Optimal Nursing Care Intensity Level. The RAFAELA system has not been tested outside of Finland. **Methods:** A prospective, descriptive study was performed at 5 clinical units at Oslo University Hospital during 2011-2012. The interrater reliability of the Oulu Patient Classification instrument was tested by parallel classification including 100 - 167 patient classifications pr. unit, and analyzed by consensus in % and using Cohen's Kappa. Convergent validity was tested by using the average Oulu Patient Classification instrument value to predict the average Professional Assessment of Optimal Nursing Care Intensity Level for the same calendar day by linear regression analysis. **Results:** The Oulu Patient Classification instrument consensus of parallel classifications varied between 70.1% - 89%. Cohen's Kappa within patient classes varied between 0.57 and 0.81, representing substantial interrater reliability. The Oulu Patient Classification instrument was valid as the instrument in average explained about 38% of the variation of the Professional Assessment of Optimal Nursing Care Intensity Level. **Conclusions:** Patient classification systems tested for psychometric properties are needed and this study provides evidence of satisfactory reliability and validity of the Oulu Patient Classification instrument as tested outside Finland, demonstrating that this instrument has international relevance within nursing.

---

\*Corresponding author.

## Keywords

**Nursing Intensity; Staffing; RAFAELA System; Oulu Patient Classification Instrument; Reliability; Validity**

---

## 1. Introduction

Lack of appropriate staff is a stumbling block to the provision of effective nursing care. Finding answers to the association between resource allocation and quality of hospital care continues to challenge health service managers [1]-[4]. Even if advances in technology to some extent have reduced the need for inpatient care, the complexity in medical and surgical interventions increases the need for a large and more sophisticated clinical workforce [2] [5].

The contribution by nurses to the treatment and care of patients, and the nursing intensity (NI) defined as the patients' need for care and the nursing intervention provided for the patients, have a direct impact on quality and outcome [6]. Limited resources have influenced the planning of nursing forces in hospitals. According to the Norwegian Parliament's Coordination Reform of 2009, lack of appropriate systems for resource allocation may be one of the reasons why proper patient care is insufficient [7]. Recent research has also shown that instruments measuring nursing practice often are imprecise [8]. Reliable and valid systems are crucial in order to secure responsible decision making within professional nursing practice [9]. This includes balancing appropriate patient care with optimal workload. Hence, there is a need for evaluating the psychometric properties of existing patient classification systems that measure NI.

The Finnish patient classification system RAFAELA, which includes the Oulu Patient Classification instrument (OPCq), the calculation of nursing resources (N), and the Professional Assessment of Optimal Nursing Care Intensity Level (PAONCIL), was designed to measure nursing intensity and allocation of nursing staff [10]. The system was developed during the 1990s and has become widely used in Finland. Since 2010 also Iceland, Sweden and the Netherlands have started to use the RAFAELA system. However, the system has never been evaluated outside of Finland. Thus, the aim of this investigation was to study reliability and validity of the OPCq as used at Oslo University Hospital in Norway. This paper represents the first report from a larger Norwegian evaluation project of the RAFAELA system.

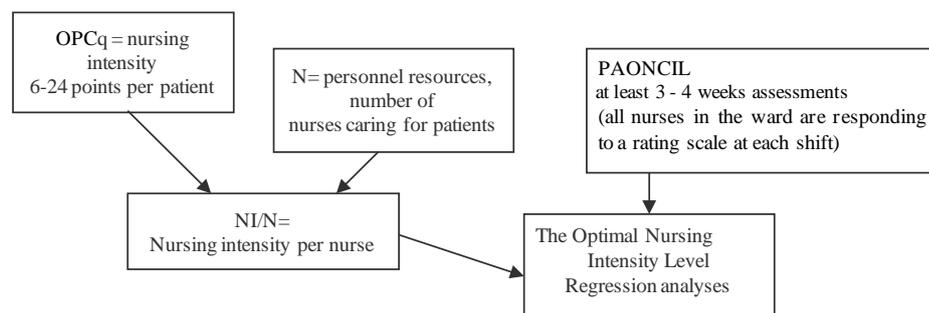
## 2. Background

The purpose of the RAFAELA system is to create a work situation where the needs and amount of patient care is in balance with personnel resources. The aim is to allocate resources in accordance with the optimal nursing care level. The system was originally developed as a three-part system for hospital settings, *i.e.* medical and surgical wards (Figure 1). The two first parts, consisting of a patient's NI, which is measured daily by the OPCq, and the daily nursing resources allocated to the patient's nursing care, are used to calculate the nurses' actual patient-related workload. The total amount of NI points on the unit (for example 350), is divided by the amount of nurses (for example 12) that have taken care of the patients during the same day (24 hours). The patient-related workload per nurse can then be expressed as NI points per nurse (NIp/N), which would be 29.2 NIp/N for this example. Assessment of optimal workload per nurse is then established for each unit by running a test over a period of at least 3 - 4 weeks where the PAONCIL instrument is used [10].

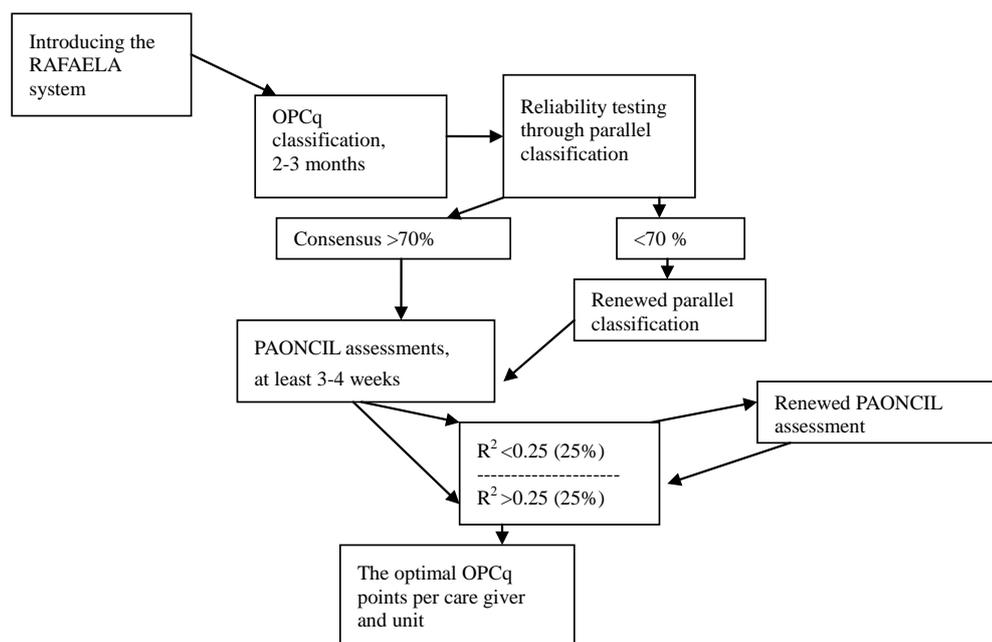
The PAONCIL method was developed as an alternative to time studies, which has been influenced by a more technological view of nursing [11]. The idea of the RAFAELA system is that the workload, which is expressed as NIp/N, is compared to the optimal nursing intensity level for the ward. When the actual NIp/N is on the optimal level, a successful resource allocation is obtained. That means that available personnel resources are in balance with the needs for patient care. The RAFAELA system has been tested for reliability and validity for use in Finland [10] [12]-[14]. The process of implementing the system into a new context is scheduled to about 6 months pr unit, which is in accordance with the RAFAELA manual (Figure 2).

## 3. Methods and Material

The OPCq and the PAONCIL instruments were translated to Norwegian by an expert group consisting of 2



**Figure 1.** The structure of the RAFAELA system.



**Figure 2.** Implementing the RAFAELA system. (Frilund, M. and Fagerström, L. 2009).

clinical nurse specialists, 2 Master nurses, and 1 PhD nurse. Established principles for good practice for translation were followed [15]. In short this process included forward translation, reconciliation and back translation, with harmonization discussions through all stages.

A pilot study was conducted at a clinical unit of Oslo university Hospital in 2010 to assess the instruments in the study setting at our hospital. Included in the pilot study were evaluations of the abilities of the nurses to classify the patients according to the Norwegian instructions, validation of the translated version of the RAFAELA system, and monitoring the RAFAELA web solution. The pilot study revealed a need for repeated training sessions for the nurses handling the patient classification, and minor adjustments regarding the Norwegian version of the instrument. No problems were identified concerning the use of the RAFAELA web solution.

### 3.1. Sample

Data were obtained from five clinical units at Oslo University Hospital. The core activity was related to organ transplantation, cancer surgery and cancer medicine, urology, rheumatology, infection medicine, and dermatology. All nurse leaders and nurses employed at the wards participating in the study received training in the use of the RAFAELA system by the responsible researchers, and took part in the daily patient classifications (Table 1).

Further, all nurses available at work in the test periods participated in the reliability and validity tests. Data for reliability tests were collected from May 2011 to March 2012 (Table 2), and data for validity tests from January to November 2012 (Table 4).

**Table 1.** Characteristics of the 5 study units.

Units	Number of beds	Number of registered nurses	Number of nursing assistants
Unit of Transplant Surgery	27	43	0
Unit of Gastro/Urology Surgery	28	44	0
Unit of Gastro/Nephrology Medicine	25	34	0
Unit of Dermatology	14	16	2
Unit of Rheumatology and Infection	13	16	4
Total	107	153	6

**Table 2.** Consensus of parallel classification tests of the OPCq patient classes.

Units	Test period reliability	Test period validity	Numbers of parallel classifications	Consensus in %
Unit of Transplant Surgery	23.05.2011-11.07.2011	14.01.2012-28.02.2012	167	70.1
Unit of Gastro/Urology Surgery	11.10.2011-25.11.2011	24.02.2012-30.03.2012	101	72.3
Unit of Gastro/Nephrology Medicine	08.11.2011-16.12.2011	15.02.2012-30.03.2012	102	80.4
Unit of Dermatology	12.01.2012-12.02.2012	12.10.2012-11.11.2012	100	83.0
Unit of Rheumatology and Infection	06.02.2012-16.03.2012	18.05.2012-19.06.2012	100	89.0

### 3.2. Instruments

The OPCq instrument measures six sub-areas of patient needs and associated nursing interventions: 1) planning and coordination of care; 2) breathing, blood circulation and symptoms of disease; 3) nutrition and medication; 4) personal hygiene and excretion; 5) activity/movement, sleep and rest, and 6) teaching, guidance and follow-up in care, and emotional support [10]. The nurse classifies the patients once per calendar day for the last 24-hour period. Only those with needs for care that are met by the nurse are taken into account. NI can vary for each sub-area between A = 1 point, B = 2 points, C = 3 points, and D = 4 points. Level A describes a patient who manages relatively well on his/her own, B describes a patient who is occasionally in need of care, C describes a patient who needs repeated care, and finally D describes a patient who cannot manage unattended at all. The criteria for the different care levels are described in a manual for each sub-area of the instrument. The NI points are added up to a total score, ranging from 6 to 24 points per patient [10]. On this basis the patients are classified into 5 patient categories, from minimum to intensive care needs.

The number of nurses allocated to attend to the needs of the patients is registered per 24 hours by the head nurse or other nurses in charge. This number is obtained for the same period as the OPCq classification. Nurses who are not involved in patient care are not included in the calculation of NI, thus administrative work and meetings outside the unit are excluded. Total NI points at the unit are divided by the number of nurses, and the patient-related workload, *i.e.* the NI/nurse for each day, is obtained [10].

The assessment of optimal workload per nurse is established by conducting a 4 - 6 week test using the PAONCIL instrument for at least 3 - 4 weeks. On a scale from -3 to +3, where -3 indicates very low and +3 very high nursing intensity, the nurses assess to which extent they have had the possibility to meet the needs of their patient group during the shift, estimated with an accuracy of 0.25. A manual that describes the seven levels on the scale is available. For example, a very low level (*i.e.* -3) describes a situation where the care needed by the patient is minimal in relation to the available nursing resources. On the opposite end of the scale, a very high level (*i.e.* 3) refers to a situation where the nurses were unable to administer all care needs; in other words the nursing care intensity was too high in relation to available resources. Consequently, only the most urgent patient needs can be taken care of. The 0-level of the scale, however, represents the optimal nursing intensity and is defined as a situation where the patients receive good total care and the balance between patient care needs and available resources is optimal [10] [16]. Before starting the PAONCIL assessment, the quality level for good

nursing care was determined for each unit through discussion meetings with the nurse group, according to the principles of staff training for the RAFAELA system. This process was led by the nurse leaders, clinical nurse experts and the researchers.

### 3.3. Data Analysis

To test reliability, a pair of nurses classified the same patients without knowing the ratings of each other by parallel classification. First we calculated the percentage of exact agreement of patient classes between two raters to the total numbers of all ratings. Then we used the un-weighted Cohen's Kappa ( $k$ ) to test interrater agreement of both the OPCq sub areas and patient classes [17]. The latter is a well established statistical measure of interrater agreement for categorical items. Since  $k$  takes the agreement occurring by chance into account, it represents an important supplement to percentage agreement and is thought to be a more robust measure. It has a maximum of 1.00 if agreement is perfect. We followed the guidelines of Landis and Koch [18] to interpret the results of the interrater analyses in our study: 0.00 - 0.20 slight, 0.21 - 0.40 fair, 0.41 - 0.60 moderate, 0.61 - 0.80 substantial, 0.81 - 1.00 almost perfect.

To test validity we used the OPCq value to predict the average PAONCIL value for the same calendar day. As previous research has shown that the OPCq and the PAONCIL to some extent measure the same phenomena [16] data were analyzed by simple linear regression analysis to quantify to what extent the predictor variable, the OPCq, explained the variations in values of the outcome variable, the PAONCIL. The coefficient of determination  $R^2$ , expresses to what extent the variation in the values of the dependent variable is accounted for by the independent variable by the aid of the regression model [19]. If for instance  $R^2$  is 0.3, the model explains 30% of the variation in the value of the outcome variable. In this study, the explanation value was expressed in percentage. All data was directly entered into the RAFAELA web solution form and automatically transferred to compatible software. The SPSS statistical package version 18 (SPSS Chicago, Illinois) was used for statistical analyses.

### 3.4. Ethical Considerations

The study was conducted in accordance to the Helsinki declaration and assessed by the Regional Committees for Medical and Health Research Ethics in Norway in 2010. The project did not directly affect patients or their care. Approval was obtained from The Institutional Review Board at Oslo University Hospital, number #2010/27572. The patients being classified and the nurses participating in the study were provided with oral and written information about the study, signed informed consent forms, and guaranteed anonymity and confidentiality as well as the right to withdraw from the study at any time.

The RAFAELA system is owned by the Association of Finnish Local and Regional Authorities, and its use is managed by FCG—Finnish Consulting Group Ltd. The actual study was initiated by Oslo University Hospital, hence MHA and KL made the first contact with FCG. The license to use the system was acquired through a standard agreement between the hospital and FCG.

## 4. Results

Background characteristics of the study units are shown in **Table 1**. As illustrated, the majority of employees were registered nurses (RNs). Only two out of five units were employing nursing assistants in addition to RNs. The data show a large diversity in number of beds and nurses, with Unit of Transplant surgery and Unit of Gastro/Urology surgery being the two largest units. For parallel classification, the percentage of consensus varied between 70.1% and 89%. This indicates satisfactory reliability in all study units (**Table 2**). The calculated Cohen's Kappa for patient classes and each sub area of the OPCq are presented in **Table 3**. The interrater reliability of the patient classes varied between 0.59 - 0.81, and of the sub areas between 0.45 and 0.90, which indicate satisfactory reliability. The lowest Cohen's Kappa was found for sub areas 1 and 6 at Unit of Transplant surgery, and sub area 6 at Unit of Gastro/Urological surgery. The nurses reached the highest level of agreement within sub areas 3 and 4, reflecting patient needs related to nutrition and medication, and personal hygiene and excretion.

When it comes to validity, the response rate of the PAONCIL test obtained more than 70% in all units (**Table 4**). We found the average explanation percentage of all study units to be 37.6, meaning that the OPCq instrument in average explains about 38% of the variation of the PAONCIL values. The lowest explanation value was at Unit of Dermatology (26.8%) and the highest at Unit of Transplant surgery (59.8%).

**Table 3.** Cohen's Kappa for patient classes and the six sub areas of the OPCq instrument.

	Unit of Transplant Surgery	Unit of Gastro/Urological Surgery	Unit of Gastro/Nephro Medicine	Unit of Dermatology	Unit of Rheumatology and Infection
Patient class	0.59	0.57	0.68	0.73	0.81
Sub area 1	0.45	0.50	0.70	0.75	0.82
Sub area 2	0.58	0.62	0.67	0.73	0.79
Sub area 3	0.62	0.68	0.73	0.90	0.71
Sub area 4	0.60	0.69	0.71	0.73	0.81
Sub area 5	0.64	0.55	0.72	0.78	0.80
Sub area 6	0.48	0.46	0.72	0.62	0.82

**Table 4.** Convergent validity with explanation values.

Units	Test Period Validity	Number of PAONCIL Assessments	Nurse Responses %	Explanation Value %
Unit of Transplant Surgery	14.01.2012-28.02.2012	512	83	59.8
Unit of Gastro/Urology Surgery	24.02.2012-30.03.2012	493	75	36.5
Unit of Gastro/Nephrology Medicine	15.02.2012-30.03.2012	525	80	35.2
Unit of Dermatology	12.10.2012-11.11.2012	205	79	26.8
Unit of Rheumatology and Infection	18.05.2012-19.06.2012	167	95	29.7

## 5. Discussion

In this paper we present the reliability and validity data of the translated Norwegian version of the OPCq based on the testing and use at Oslo University Hospital. To our knowledge this is the first study to evaluate the RAFAELA system outside of Finland. Our results of interrater reliability measured by parallel classification demonstrated satisfactory reliability and are in line with studies previously conducted in Finland [13] [20]. The Cohen's Kappa showed a substantial consensus for the OPCq classifications when used in both surgical and medical wards at the hospital. This indicates that the Norwegian version of the RAFAELA instrument provides robust and reliable information. When adapting instruments for patient classification from one country to another, it is important that the cultures of the two countries correspond to a certain extent. Hence, our results suggest that Norwegian and Finnish practice among nurses is comparable and not likely to differ considerably when evaluating patients' needs and welfare.

Our results indicated that within pairs of nurses, the level of disagreement was greatest during the ratings of sub area 1 (planning and coordination of care) and 6 (teaching, guidance/follow-up care and emotional support). This is consistent with previous findings [13] [20] and supports the assumption that two of six OPCq areas may be more difficult for nurses to assess than the rest. One way to omit this hurdle would be to provide a more detailed description of the sub areas in the RAFAELA manual. Substantial operational definitions and keywords explaining the content should be included, as previously reported [21]. Furthermore, educating, training and providing regular patient case exercises may impact positively on a common understanding of the different levels. The importance of this has been demonstrated in previous research testing the reliability of nursing scales [22]. The most consistent sub areas were number 3 (nutrition and medication) and number 4 (personal hygiene and excretion), also in line with similar findings reported [13] [20]. We believe that these areas reflect concrete and distinct patient care needs that are well known to the nurses. Further, they are described in detail in the RAFAELA system, which may facilitate the appearance of a consensus.

The most prominent consensus was achieved at the Unit of Rheumatology and Infection. Several factors may explain this. The size of the unit was small and contained less variation among the patient groups when compared to most of the other study units. Also, only 16 nurses were employed at the unit. It has previously been reported that some clinical units are easier to assess than others when performing psychometric testing of nursing scales [23]. However, the Unit of Rheumatology and Infection was the last unit included in the study, and one could also argue that the nurses at this unit received a higher quality of OPCq training as the researchers had

gained experience with the RAFAELA instrument education. This may have increased the nurses' common understanding of how to classify the patients.

When estimating the validity, our data showed a high nurse response rate of the PAONCIL test varying between 75% and 95%. The management of the clinic initiated the RAFAELA study, and the nurse leaders at the study units were committed to prepare for and follow up the research. Probably these conditions had a positive impact on both data quality and response rates.

We found the average explanation values for the five study units to be 37.6%. This finding is comparable with the results of Fagerstrøm *et al.* reporting an average percentage in their study to be 36.6% [10]. Our data support the fundamental idea of the RAFAELA system, that the OPCq and PAONCIL to a certain extent measure the same phenomena and thus can be analysed in relation to each other. Based on these findings, we conclude that there is satisfactory convergent validity of the OPCq when tested at a Norwegian university hospital.

## Limitations

Some limitations of the study should be mentioned. First, we planned to test the RAFAELA system in six units within our hospital. However, the sixth unit underwent organizational changes during study start. With the introduction of several new factors simultaneously that strongly affected the nurses as well as a shift in nurse leadership, we observed low patient classification rate when starting up. Actions were made to secure data quality, such as evening seminars and individual training sessions for the nurses, but without reaching satisfying numbers of patient classified. According to the RAFAELA manual (Figure 1) we were not able to do reliability and validity tests on these data sets and then had to exclude the unit from the study. We cannot exclude the possibility that data from this large unit, employing 77 nurses and housing 28 beds, would have resulted in higher data variation.

Second, in some units, repeated RAFAELA lectures were necessary in order to be able to conduct the PAONCIL test according to the manual. Some nurses had insufficient knowledge of how to use the scale. These were mostly nurses working night shifts and therefore missed the regular daily RAFAELA training sessions. Though the majority of these nurses received additional training by the research team, it is possible that some of them had too little knowledge to do the PAONCIL assessment. Consequently, this may have influenced the validity data in some units.

Third, some nurses at one of the units were hesitant to use the left part of the PAONCIL scale (indicating low nursing intensity) in periods with low activity, because they were afraid of sanctions from the management. As a consequence, the explanatory power remained below 25% and the test was not approved according to the PAONCIL manual. We decided to make a re-measurement after intensive sessions with education and debates in the nursing group. The result of the new PAONCIL test demonstrated a larger variation using the seven PAONCIL-levels, and thus a rise of explanatory power. During this process we experienced that the nurses were in need of extensive education and that thorough knowledge about the PAONCIL measurement impacted positively on the validity of this scale. On this background we strengthened the PAONCIL instructions before implementing the system at other study units. Hence, we believe that the nurses in general conducted the PAONCIL ratings according to the manual.

In spite of the limitations mentioned here we consider our results to be representative for surgical and medical hospital wards. Furthermore, we believe that generalization of our findings beyond our study units is possible and may provide interesting and useful information.

## 6. Conclusion

This study represents the first step in evaluating the RAFAELA system outside Finland. It provides evidence of satisfactory reliability and validity of the Oulu Patient Classification instrument as tested in a clinical hospital setting in Norway, demonstrating that this instrument has international relevance within clinical nursing practice.

## Acknowledgements

This work was funded by South-Eastern Norway Regional Health Authority. We are grateful to Ben Gøran Eriksson, Riitta Hakola and Siv Stafseth for their contribution in the process of translating the RAFAELA sys-

tem to Norwegian. Also we are grateful to Leiv Sandvik for statistical advice.

## Contributions

MHA, KL and LF were responsible for the study conception and design. MHA and KL were responsible for data collection, MHA performed the analyses. MHA, KL and LF were responsible for the manuscript.

## Conflicts of Interests

The authors declare no conflicts of interests.

## References

- [1] Tervo-Heikkinen, T., Kiviniemi, V., Partanen, P. and Vehviläinen-Julkunen, K. (2009) Nurse Staffing Levels and Nursing Outcomes: A Bayesian Analysis of Finnish-Registered Nurse Survey Data. *Journal of Nursing Management*, **17**, 986-993. <http://dx.doi.org/10.1111/j.1365-2834.2009.01020.x>
- [2] Brown, D.S., Donaldson, N., Burnes Bolton, L. and Aydin, C.E. (2010) Nursing-Sensitive Benchmarks for Hospital to Gauge High-Reliability Performance. *Journal of Health Quality*, **32**, 9-17. <http://dx.doi.org/10.1111/j.1945-1474.2010.00083.x>
- [3] Aiken, L.H., Sermeus, W., Van den Heede, K., Sloane, D.M., Busse, R., McKee, M., Bruyneel, L., Rafferty, A.M., Griffiths, P., Moreno-Casbas, M.T., Tishelman, C., Scott, A., Brzostek, T., Kinnunen, J., Schwendimann, R., Heinen, M., Zikos, D., Sjetne, I.S., Smith, H.L. and Kutney-Lee, A. (2012) Patient Safety, Satisfaction, and Quality of Hospital Care: Cross Sectional Surveys of Nurses and Patients in 12 Countries in Europe and the United States. *British Medical Journal*, **344**, e1717.
- [4] McGillis Hall, L., Wodchis, W.P., Ma, X. and Johnson, S. (2013) Changes in Patient Health Outcomes from Admission to Discharge in Acute Care. *Journal of Nursing Care Quality*, **28**, 8-16. <http://dx.doi.org/10.1097/NCQ.0b013e3182665dab>
- [5] Rechel, B., Wright, S., Edwards, N., Dowdeswell, B. and McKee, M. (2009) Investing in Hospitals of the Future. European Observatory on Health Systems and Policies. <http://www.euro.who.int/pubrequest>
- [6] Schultz, N.A., Larsen, P.N., Klarskov, B., Plum, L.M., Frederiksen, H.J., Christensen, B.M., Kehlet, H. and Hillingsø, J.G. (2013) Evaluation of a Fast-Track Programme for Patients Undergoing Liver Resection. *British Journal of Surgery*, **100**, 138-143. <http://dx.doi.org/10.1002/bjs.8996>
- [7] Report nr. 47 (2008-2009) to the Norwegian Parliament. The Coordination Reform. Proper Treatment—At the Right Place and Right Time.
- [8] Kane, R.L., Shamliyan, T.A., Mueller, C., Duval, S. and Wilt, T.J. (2007) The Association of Registered Nurse Staffing Levels and Patient Outcomes. Systematic Review and Meta-Analysis. *Medical Care*, **45**, 1195-1204. <http://dx.doi.org/10.1097/MLR.0b013e3181468ca3>
- [9] Pering, S.-J. and Yu, M.-L. (2013) Psychometric Testing of an Instrument Measuring Nurse Aides' Patient Safety Attitudes. *Journal of Nursing Management*, **21**, 1001-1007. <http://dx.doi.org/10.1111/j.1365-2834.2012.01431.x>
- [10] Fagerström, L., Rainio, A.-K., Rauhala, A. and Nojonen, K. (2000) Professional Assessment of Optimal Nursing Care Intensity Level. A New Method for Resource Allocation as an Alternative to Classical Time Studies. *Scandinavian Journal of Caring Sciences*, **14**, 97-104.
- [11] Weston, M.J. (2009) Validity of Instruments for Measuring Autonomy and Control over Nursing Practice. *Journal of Nursing Scholarship*, **41**, 87-94. <http://dx.doi.org/10.1111/j.1547-5069.2009.01255.x>
- [12] Rauhala, A. and Fagerström, L. (2004) Determining Optimal Nursing Intensity: The RAFAELA Method. *Journal of Advanced in Nursing*, **45**, 351-359. <http://dx.doi.org/10.1046/j.1365-2648.2003.02918.x>
- [13] Fagerström, L., Rainio, A.-K., Rauhala, A. and Nojonen, K. (2000) Validation of a New Method for Patient Classification, the Oulu Patient Classification. *Journal of Advanced Nursing*, **31**, 481-490. <http://dx.doi.org/10.1046/j.1365-2648.2000.01277.x>
- [14] Frilund, M. and Fagerström, L. (2009) Validity and Reliability Testing of the Oulu Patient Classification Instrument within Primary Health Care for the Older People. *International Journal of Older People Nursing*, **4**, 280-287. <http://dx.doi.org/10.1111/j.1748-3743.2009.00175.x>
- [15] Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A. and Erikson, P. (2005) Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value in Health*, **8**, 94-104. <http://dx.doi.org/10.1111/j.1524-4733.2005.04054.x>

- [16] Fagerström, L. and Rauhala, A. (2007) Benchmarking in Nursing Care by the RAFAELA Patient Classification System. *Journal of Nursing Management*, **15**, 683-692. <http://dx.doi.org/10.1111/j.1365-2934.2006.00728.x>
- [17] Tavakol, M. and Dennick, R. (2011) Making Sense of Chronbach's Alpha. *International Journal of Medical Education*, **2**, 53-55. <http://dx.doi.org/10.5116/ijme.4dfb.8dfd>
- [18] Landis, J.R. and Koch, G.G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**, 159-174. <http://dx.doi.org/10.2307/2529310>
- [19] Altman, D.G. (1997) *Practical Statistics for Medical Research*. Chapman and Hall, London.
- [20] Fagerström, L. (1999) *The Patients' Caring Needs. To Understand and Measure the Unmeasurable*. Dissertation, Åbo University, Finland.
- [21] Luiking, M.-L., van Linge, R., Bras, L., Grypdonck, M. and Aarts, L. (2012) Psychometric Properties of the Dutch Version of the American Activity Scale in an Intensive Care Unit. *Journal of Advanced Nursing*, **68**, 2750-2755. <http://dx.doi.org/10.1111/j.1365-2648.2012.06024.x>
- [22] Kottner, J., Halfens, R. and Dassen, T. (2008) An Interrater Reliability Study of the Braden Scale in Two Nursing Homes. *International Journal of Nursing Studies*, **45**, 1501-1511. <http://dx.doi.org/10.1016/j.ijnurstu.2008.02.007>
- [23] Gould, D., Goldstone, L., Kelly, D. and Gammon, J. (2004) Examining the Validity of Pressure Ulcer Risk Assessment Scales: A Replication Study. *International journal of Nursing Studies*, **41**, 331-339. <http://dx.doi.org/10.1016/j.ijnurstu.2003.10.005>