Received: 19 July 2022 | Revised: 15 September 2022 | Accepted: 29 September 2022

DOI: 10.1111/jocn.16559

EMPIRICAL RESEARCH MIXED METHODS

Journal of Clinical Nursing WILEY

How the reliability and validity of the patient classification system can be ensured in daily nursing work? A follow-up study

Jaana K. Junttila MSc, RN, Researcher[1] | Kaisa Haatainen PhD, Adjunct Professor[2] | Aija Koivu PhD, Statistics expert[3] | Pirkko Nykänen PhD, Professor Emerita[1]

[1]Faculty of Information Technology and Communication Sciences (ITC), Tampere University, Tampere, Finland

[2]Department of Nursing Science, University of Eastern Finland, Kuopio, Finland

[3]Medical Center, Kuopio University Hospital, Kuopio, Finland

Correspondence
Jaana K. Junttila, Medical Center, Kuopio University Hospital, P.O. Box 100, Kuopio 70029, Finland.
Email: jaana.junttila@kuh.fi

Funding information
Valtion tutkimusrahoitus (VTR), Grant/Award Number: Erva-research committee record 2/2021

Abstract

Aims and Objectives: This study aimed to determine the reliability and validity of the RAFAELA patient classification system (PCS) for qualified and efficient nurses.

Background: The number of patients per nurse or diagnosis-based determination of nursing workload are imprecise measures that do not consider the variation in patients' care needs. Ensuring the reliability and validity of the RAFAELA is important for the efficient allocation of nursing resources.

Methods: In this study, we investigated how the maintenance (parallel classification measurement and professional assessment of optimal nursing care intensity level measurement) of the RAFAELA was done with 9 years of follow-up data. The results were analysed using quantitative methods supplemented with qualitative audit descriptions. The STROBE checklist was used.

Results: The RAFAELA was used continuously in 44 units (40%). The length of use of the RAFAELA influenced the success of parallel classification measurements. Six per cent of units passed parallel classification measurement over 75% after 1–3 years' use, 42% after 4–6 years and 83% after 7–9 years. Among the units that used the RAFAELA PCS continuously, only four (9%) passed the professional assessment of optimal nursing care intensity level measurement.

Conclusions: This study shows that ensuring the reliability and validity of the use of the RAFAELA is laborious, requires several years of use and continuous investments in nurses' skills and motivation.

Relevance to Clinical Practice: Qualified use of PCS is challenging, and organisations should invest to maintenance, training, support and user motivation. Each patient should be classified comprehensively, and nursing resources should be calculated correctly. In addition, utilisation of the nursing intensity level should be maximised.

Clinical Trial Registration Number: Kuopio University Hospital organisation permit number 73/2014.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *Journal of Clinical Nursing* published by John Wiley & Sons Ltd.

Patient or Public Contribution: Information regarding individual patients or nurses was not available to the researchers. All materials are in the form of summary tables.

KEYWORDS
HIS, HRIS, maintenance, patient classification system, PCS, RAFAELA, reliability, validity

1 | INTRODUCTION

1.1 | Background

Nursing workload (NWL) data, which describe the diversity of patient care, are one of the key indicators of nursing work used to plan and evaluate nursing resources and management (Fagerholm, 2014; Liljamo et al., 2020;). The NWL level is defined as the ratio of the unit's patient nursing intensity (NI) to the number of nurses, which is compared to the optimal NWL (the amount of work with which sufficient quality of nursing work can be guaranteed in the unit). Determining the level of NWL that guarantees qualified care that covers patients' different care needs is a challenge (Griffiths et al., 2020). For example, Griffiths et al. (2020, 2021) and Griffiths and Saville (2019) referred to questions that are still open, such as whether the tools for measuring resources reliably identify a sufficient number of nurses and how to determine the reference level for estimating the average need for care. However, determining the optimal level of NWL is challenging because of the large variations in patients, staff and conditions of care; thus, the optimal level of NWL changes over time, and must be reassessed at certain intervals or always when major changes occur in the unit (Rauhala & Fagerström, 2004).

In time and motion studies, the number of patients per nurse alone or the diagnosis-based determination of NWL are imprecise measures that do not consider the variation in patient acuity (Duffield et al., 2011; Needleman, 2015; West et al., 2014). Compared to NWL based on measuring objective numbers, the problem of patient classification systems (PCS) based on nurses' subjective assessments is the work needed to maintain their reliability and validity (Fasoli et al., 2011). In addition, trust is needed that nurses use PCS uniformly (Rauhala, 2008). Reliable and comprehensive NI information should be available to the staff, based on the actual patient acuity, and be in line with the information recorded in the nursing documentation (Fagerström et al., 2014; Liljamo, 2018).

The Finnish RAFAELA PCS is used in more than 800 units in more than 30 public healthcare organisations in 19/21 healthcare districts in Finland. In these units, patients' NI is assessed, and nursing resources are calculated using the RAFAELA PCS to assess the daily NWL (Liljamo et al., 2017). In addition to assessing the NWL, NI categories (1–5) are used to support strategic decision-making, research, cost evaluation of specialised medical care, invoicing, and improvement of patient care quality and safety (Junttila et al., 2016, 2019; Virkkunen et al., 2015; Vuokko et al., 2017). The RAFAELA PCS is also used in Iceland, Norway and Sweden.

What does this paper contribute to the wider global clinical community?

- The assessment of nursing workload and nursing intensity related to patients' needs, and the adequacy of nursing resources is an acute and necessary topic of research and further improvement.
- There are already many patient classification systems in use, but instead of introducing new systems, the focus should be on better utilising the possibilities of existing patient classification systems and developing them further to support nurses in assessing the reliability and validity of patient classification.
- The results of this study revealed new, previously unexplored information on the challenges of the use and maintenance of the reliability and validity of patient classification systems, and identified further needs for their development.

In the RAFAELA PCS, the unit-specific optimal NWL level was determined by comparing the actual NWL with the nurses' assessment of how well the patients could be cared for during the load cycle. The NWL level is defined as the ratio of the unit's patient NI to the number of nurses, which is compared to the optimal NWL level (the amount of work with which sufficient quality of nursing work can be guaranteed in the unit). The optimal level of NWL varies not only between units, but also within units, so it should be checked at least every second year; thus, the optimal level of NWL reliably reflects the current situation in the unit.

The reliability of the instruments based on subjective assessments may be weak. Therefore, the reliability and validity of subjective evaluations need to be ensured by comparing the evaluations from parallel classifications given by two or more nurse evaluators to ensure interrater reliability. If nurse evaluators end up at different assessments, the instrument gives random results, and the outcome of the subjective evaluations cannot be utilised

(Gisev et al., 2013). Since there is no certainty whether nurses know how to use PCS correctly, the previously mentioned parallel classification method has been added as a system feature of the RAFAELA PCS.

The reliability and validity of using the RAFAELA PCS are based on the consistency of the nurses' evaluations, which is secured by continuous maintenance of the system with annual parallel classification measurement (PCM) and every second year, definition of the optimal level of NWL with the professional assessment of optimal nursing care intensity level (PAONCIL) measurement together with nurses' training. The quality assurance with regular auditing associated with the maintenance of the RAFAELA PCS is unique in systems developed for measuring NWL, making the RAFAELA PCS a reliable and valid, but also a very laborious tool for measuring patients' NI and NWL of the unit (Andersen et al., 2014).

The RAFAELA PCS comprises three elements: (1) Patient classification instruments to define NI for patients cared for in different types of units; (2) calculation of nursing resources; and (3) definition of the optimal level of NWL to enable monitoring of the unit's NWL. The patient's acuity level was assessed for the previous 24 h from the following six areas of nursing care: (1) planning and coordination of nursing care; (2) breathing, blood circulation and symptoms of illness; (3) nutrition and medication; (4) hygiene and secretion; (5) activity, functioning, sleep and rest; and (6) guidance of care/continued care/continued care and emotional support. In each sub-area is selected 0–4 points based on the patient acuity level, and a higher score indicated a higher NI.

The basic research questions of this study were related to two problems that affect the reliability and validity of the information obtained from the RAFAELA PCS. (1) If nurses are unable to classify patients uniformly in their subjective assessments, the information obtained from the RAFAELA PCS is unreliable (Kaustinen, 2011; Liljamo, 2018). The uniformity of patient classification was evaluated using the annual PCM. (2) If the unit's optimal level of NWL cannot be determined with PAONCIL measurement, it cannot be assessed whether the unit's daily NWL is in the optimal range. The optimal range of the unit's NWL changes as the factors affecting the NWL change; therefore, the optimal NWL should be checked every second year and always when the nature of the activity of the unit changes (Frilund & Fagerström, 2009).

The maintenance of reliability and validity related to the continuous use of RAFAELA PCS has not been studied previously. Liljamo (2018) studied the differences between units in parallel classifications made over 6 years. In our study, we investigated how well the reliability and validity of the RAFAELA PCS could be maintained in nursing practice.

The maintenance of the RAFAELA PCS is divided into two phases: (1) The maintenance implemented by the system supplier consists of commissioning, including training, regular maintenance training and quality assurance. (2) Maintenance activities within the organisation include commissioning, continuous training, mentoring, utilisation of PCS data, assurance of reliability and validity through PCM and PAONCIL measurements, and internal auditing (Finnish Consulting Group, 2012).

2 | METHODS

2.1 | Design

This study aimed to determine how the reliability and validity required for the qualified and efficient use of the RAFAELA PCS based on the subjective assessments of nurses can be ensured by investigating how successful in practice is the maintenance of patient classification and the optimal level of NWL for each unit. These were examined through PCM and PAONCIL measurements and the audit results. In this study, the strengthening the reporting of observational studies in epidemiology (STROBE) statement guidelines was used (Supplementary File 1).

2.2 | Research questions

1. How successful has the RAFAELA PCS's PCM been?
2. How successful has the determination of the optimal NWL of the RAFAELA PCS been?
3. How successful was the maintenance process of the RAFAELA PCS implemented in the hospital?

Continuous maintenance is required to ensure the reliability and validity of the data and information obtained from the RAFAELA PCS. In addition to training and mentoring, annual PCM and every second year PAONCIL measurement are required to define the optimal NWL for each unit. The PAONCIL measurement was also renewed when the nature of the activity of the unit changed. RAFAELA PCS maintenance also includes regular auditing of the system. (Figure 1.)

Before annual PCM can start, some steps should be performed in each unit, for example, patient classification instrument training passed and classification exercises done, and if fewer than 6 months have passed since the RAFAELA PCS was commissioned or significant changes have been made recently, the annual PCM was not performed. During the annual PCM period, the head nurse monitored the development of parallel classification consensus weekly. She collected unit-specific raw data to develop the number of parallel classifications and percentage of consensus from the hospital's Information and Communication Technology (ICT) unit in an Excel spreadsheet format. When a consensus result of equal or more than 70% was achieved during the PCM period and other PCM criteria were met, the responsible head nurse gave permission to the unit to stop the measurement; otherwise, the PCM was reported as failed (adapted from the Finnish Consulting Group [2012]). The criteria for PCM were as follows: a minimum of seven patients/nurse, all nurses in the unit have classified their patients, and a minimum of 50 patients/unit have been classified.

There are also some criteria for units before PAONCIL measurement can be started, including PAONCIL measurement maintenance training is passed, PCM is passed for fewer than 6 months, and criteria of good care are defined for the unit. If more than 6 months have passed since the PCM, the PAONCIL measurement cannot be done because interrater reliability cannot be confirmed. The head

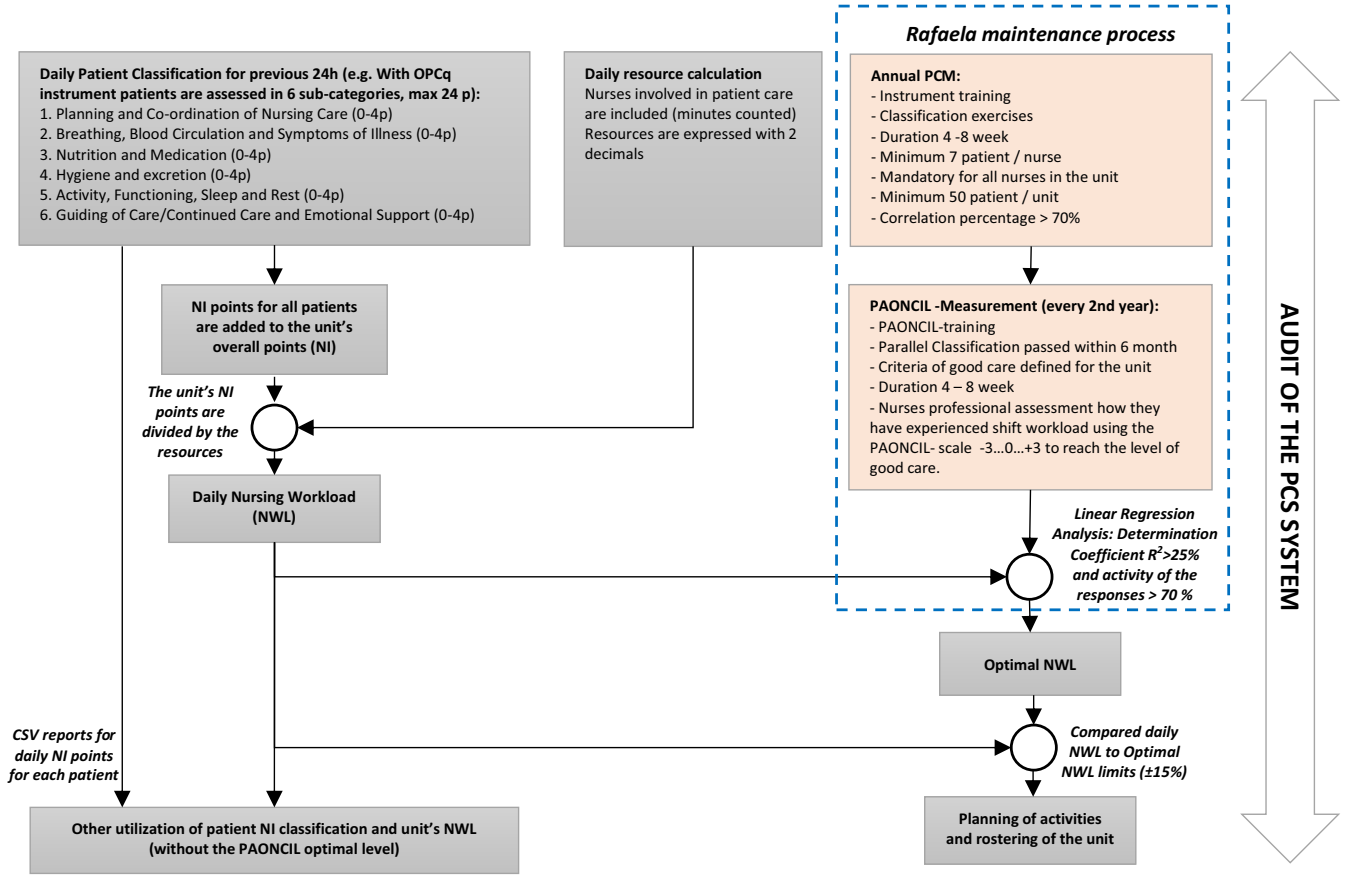


FIGURE 1 RAFAELA PCS process (adapted from Finnish Consulting Group [2012]). [Colour figure can be viewed at wileyonlinelibrary.com]

nurse monitored the development of the optimal level of NWL in the units during the PAONCIL measurement period. When the explanatory rate of the linear regression equal or more than 25% was reached, the unit was allowed to stop the measurement; otherwise, the PAONCIL measurement was reported as failed (adapted from the Finnish Consulting Group [2016]).

The RAFAELA PCS audit ensured the reliability, validity and quality of the patient classification process. To select the participants to audit, all units using the RAFAELA PCS in 2017 were given a running number, which were shuffled and randomly selected units to be audited ($N = 34$). During the audit, the patient classification process was monitored, which included daily patient classification, orientation of new nursing staff, and arrangements for training and discussion sessions. The auditors also monitored the latest results of the PCM and PAONCIL measurements of each unit, collection of benchmarking data from the RAFAELA PCS, production, and analysis of the unit's key indicators, calculation of resources, and recording and checking of the results by auditors (Figure 1).

2.3 | Data collection

The data for this study were collected from one Finnish university hospital, which is responsible for the specialised medical care of approximately 250,000 citizens (in one of 21 hospital districts). The hospital was selected because the researcher was a coordinator of the implementation project of the RAFAELA PCS and also a head nurse in charge of maintaining the PCS and collecting the results since that. The research consisted of two quantitative materials:

1. The results of the PCM and PAONCIL measurements from 109 units were collected by the researcher during the nine-year follow-up period (2011–2019).
2. Audit results of the PCS from randomly selected 34 units' head nurses and nurses ($n = 77$) were collected by auditors in 2017.

The head nurse responsible for the PCS collected the results of the annual summary reports of the PCM and PAONCIL measurements from 2011 to 2019 for the hospital's 109 units (wards, outpatient clinics and operational units) using the RARAELA PCS (have patient classification licence), as part of the maintenance process of the RAFAELA PCS.

During the monitoring period, due to changes in the unit structure (units were discontinued, new units were established, and existing units were merged), the number of units using the PCS at the same time in each year varied between 65 and 73.

The RAFAELA PCS audit was conducted by trained audit pairs consisting of the RAFAELA PCS head mentors interviewing head nurses and the RAFAELA PCS mentors interviewing nurses. The audit results collected from randomly selected 34 units which were

using the RAFAELA PCS were saved on the online audit platform in two different audit events in the spring and fall of 2017.

The extensive 42-page audit form consisted of seven main categories:

1. Management.
2. Resource calculation.
3. Skills of nurse 1.
4. Skills of nurse 2.
5. Classification of the same patients as auditor and nurse 1 and auditor and nurse 2.
6. The activities of nurse 1 during the patient classification.
7. The activities of nurse 2 during the patient classification.

- The nurse's performance was evaluated as follows:
  - Did the nurse make the classification based on nursing documentation?
  - Did the nurse check the result of the classification (patient acuity category)?
  - Did the nurse feel that they knew how to use the information obtained from the classification in their work unit and work?
  - Did the nurse use the guide for the instrument when classifying?

In each main category, the quality goals for the use of the RAFAELA PCS were also described. Free feedback was collected from head nurses and nurses participating in the audit in the feedback section of each main category. The audit form was developed and validated by the RAFAELA PCS system supplier (Finnish Consulting Group, 2012).

2.4 | Analysis

The results of the PCM and PAONCIL measurements of the units (n = 109) that used the PCS over nine years were analysed using quantitative methods. Microsoft Excel 2013 was used to calculate the ranges and averages of the results. Using SPSS 22 for Windows, the PCM results were cross-tabulated, and the chi-square test was performed to investigate whether the success of the PCM was affected by the number of years of the RAFAELA PCS use. Some of the results are presented as bar graphs to illustrate the development of the PCM and PAONCIL measurements. From the results of the RAFAELA PCS audit, the percentage distributions of the answers were calculated for each main category of questions. Respondents' views were extracted from the free-text fields of the audit results to provide a background for the processing of the research results in the reflection.

Length of service (years)	Success rate of PCMs			
	<50%	50%–75%	>75%	In total
1–3	22 (63%)	11 (31%)	2 (6%)	35 (32%)
4–6	6 (23%)	9 (35%)	11 (42%)	26 (24%)
7–9	2 (4%)	6 (13%)	40 (83%)	48 (44%)
In total	30 (28%)	27 (24%)	52 (48%)	109 (100%)

TABLE 1 Success rate of PCMs in relation to the length of time the PCS has been in use.

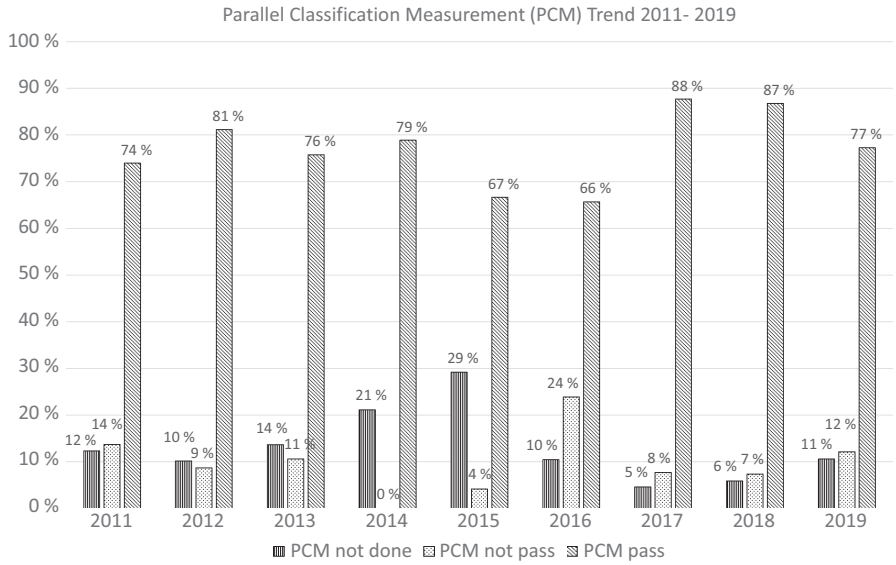


FIGURE 2 Results of PCM in 2011–2019 (N = 109 units).

3 | RESULTS

3.1 | The results of the parallel classification measurement (Research question 1)

During the nine-year follow-up period, there were much variation in PCS use. Only in 44 units (40%), the RAFAELA PCS was in use during the entire follow-up period: in 48 units (44%) for 7–9 years, 26 units (24%) for 4–6 years and 35 units (32%) for 1–3 years.

The PCM data were cross-tabulated between the length of time the RAFAELA PCS was used (years) in the unit and the success rate of PCM (Table 1). It was found that the success rate of PCM was strongly related to the length of time the unit had been using RAFAELA PCS (Chi-square = 55.64, *df* 4, $p = .000$). Statistical significance was set at $p < .05$. Approximately 6% of the units that used the RAFAELA PCS for 1–3 years passed more than 75% of the PCM, and 42% of those that used the PCS for 4–6 years, but 83% of those that used the PCS for 7–9 years passed >75% of the PCMs (Table 1).

For 30 units (28%), the PCM was successful in less than 50% of the attempts, and among those units, 20 failed in every attempt. For 27 units (24%), the PCM was successful in 50–75% of the attempts, and for 52 units (48%), the PCM was successful in more than 75% of the attempts. Of these units ($N = 44$) that participated in the PCM every time during the nine-year monitoring period, 18 units (41%) were successful at every attempt.

Among all units ($N = 109$), the proportion of those that did not attempt or failed PCMs was slightly higher at the beginning of the follow-up period (Figure 2). However, the success rate of units that participated in PCMs every year ($n = 44$) was consistently at a high level (80%–95%).

On average, 77% of the units (66%–88%) passed the PCM, but simultaneously, on average, 13% (5%–- 29%) of the units did not participate at all in the PCMs. A total of 36 units (33%) passed the PCM on every attempt during the follow-up period, while 10% of the units failed the PCM on average (0%–24%). In the early years (2011–2015) about 12%–40% and in later years (2016–2019) about 5%–10% of the units did not participate in the PCM at all (Figure 2). In approximately 10% of all PCMs, despite the efforts, sufficient consensus was not reached in the nurses' parallel classifications, so the parallel measurement failed.

Of the units ($n = 44$) that had a PCS in use throughout the whole follow-up period, the majority also passed the PCM annually (80%–95%). In these units, there was no significant change in the success rate of the PCM during the follow-up period.

3.2 | PAONCIL measurement results (Research question 2)

It was possible to participate in PAONCIL measurements for the optimal level of NWL if the PCM was successfully passed in the unit within the last 6 months. Owing to organisational changes, PAONCIL measurements were not performed in 2018. In this study, the results of PAONCIL measurements were examined only for those units ($n = 44$) that used the PCS for the entire nine-year monitoring period. In 2014, the RARAELA moved from the commissioning phase to the continuous maintenance phase, and because of that, the RAFAELA PCS's maintenance and support was reorganised, which was reflected in a temporary drop in the results of the PCM and PAONCIL measurements.

Of these 44 units, the optimal level of NWL was valid (1st or 2nd year going) on average for 57% of the units (16%–84%). PAONCIL measurement failed on average 20% of units (0%–43%). On average, 23% (9%–59%) did not participate in the PAONCIL measurements at all.

Only four units of 44 (9%) passed the PAONCIL measurements every time. Three units (7%) did not pass the PAONCIL measurements at all during the follow-up period. For these 44 units, PAONCIL measurement results improved from 48% to 84% between 2014 and

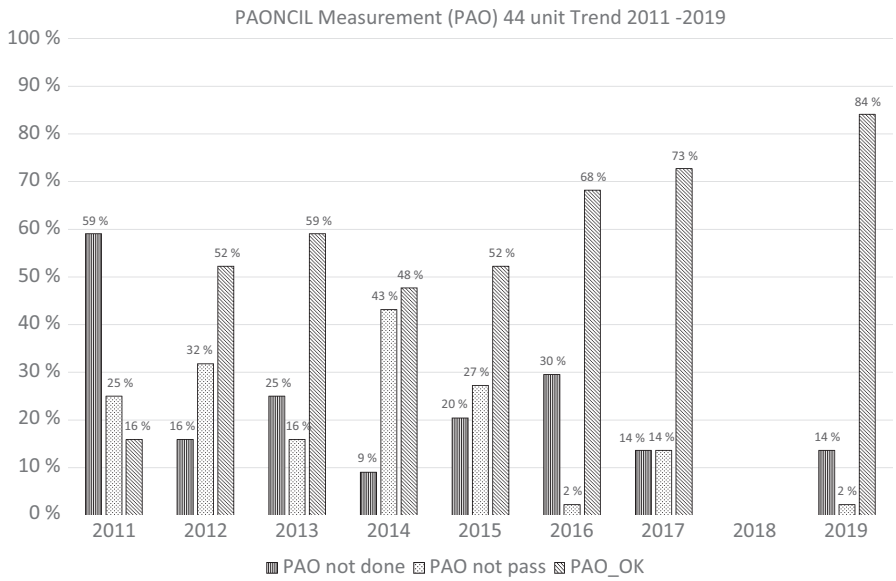


FIGURE 3 Results of the redefinition of the optimal level of NWL ($n = 44$) (in 2018, PAONCIL measurements were not done).

2019. The number of units that had not participated in or passed the PAONCIL measurements decreased steadily throughout the follow-up period (Figure 3).

3.3 | Audit results (Research question 3)

In this study, only those areas of the audit were evaluated that were related to the reliability and validity of the RAFAELA PCS. Main Categories 1–2 were related to nursing management and resource calculations. The criteria for qualified care have been defined in 94% of the units, and 91% of the units passed the PCM. The patient classification was comprehensive in <80% of the units, and approximately 62% of head nurses felt that the NWL data reflected reality. The resource calculation was performed according to the instructions at less than 85%, and the PAONCIL measurements were passed in approximately 70% of the units. Only 56% of respondents regularly reviewed the RAFAELA PCS reports with the nursing staff.

In the main Categories 3 and 4, nurses' ($n = 77$) care classification skills were discussed. The results showed that 97% of the nurses had received training for patient classification, and about 94% of the nurses thought they knew how to classify patients, but only 73% of the nurses felt that they had received sufficient training for the patient classification.

Category 5 consisted of the classification of patients ($n = 140$) by the auditor and two nurses in the units under audit. When the nurse first classified the patient and then verbally described the patient's condition to the auditor who reclassified the same patient, the patient was classified into the same NI class 78% of the time. When the auditor's classification was based on the information obtained from the nursing records, only 60% of the patients were classified into the same NI class. If the NI class was different, the nurse's NI class was higher than the auditor in 95% of the classified patients.

Categories 6 and 7 dealt with auditors' evaluations of nurses' activities during the NI classification. Only 76% of the nurses classified their patients based on their nursing records. After the classification, 75% of the nurses checked whether the NI corresponded to the patient's need for care (NI class). Less than 63% of nurses felt that they could use the information obtained from the RAFAELA PCS in their daily work. Only 20% of the nurses used the guide for patient classification instrument.

4 | DISCUSSION

This study explored how the reliability and validity of the RAFAELA PCS was maintained in the nursing practice of one Finnish hospital. The results showed that the longer the unit used the RAFAELA PCS, the more successful the unit was in ensuring the reliability and validity of the PCS. These results are consistent with those of previous research. Achieving a good level of maintenance requires several years and continuous measurements, support and motivated users. Nurses learn how to use the RAFAELA PCS over time, but it is still perceived as laborious both to use and maintain (Andersen et al., 2014). In a busy nursing practice, sometimes there is insufficient time to classify patients. Andersen et al. (2014), Hustad et al. (2015) and Oostveen et al. (2016) also stated that RAFAELA PCS is a demanding and time-consuming system for operation.

Fasoli and Haddock (2010) collected information on practices related to PCSs in their integrative literature review, looking for a 'gold standard'. However, they faced the same challenges that had been presented before, such as problems in measuring NWL and defining the content of nursing work. Griffiths et al. (2020) came to the same conclusion that the measurement of NI should be based on the assessment of a nursing professional, which reflects real nursing work considering the patient's acuity and available resources. This feature is built into the RAFAELA PCS, but there still is a question of whether nurses can classify patients' NI consistently.

If the RAFAELA PCS's reliability (success of PCM = nurses evaluate NI in a consistent way) and validity (reached optimal NWL = updated 'golden standard') cannot be verified, the information obtained with the PCS is not reliable. This is probably a problem with all the instruments based on subjective assessments. Often, in the piloting phase of the PCS, the reliability of the NI measurement is invested; however, in use, the PCS loses its reliability when for example new users are not properly trained and agreed practices are gradually modified in everyday life. Because of these factors, strict monitoring of the reliability of the RAFAELA PCS is a good thing, but is it a too laborious process?

According to the results of this study, PCMs were successfully conducted every year in most of the units where the PCS was in use throughout the whole nine-year monitoring period. In addition, it seems clear that over the years, an increasing number of these units passed the PAONCIL measurements and reached the optimal level of the unit's NWL. These audit results confirm those collected during the monitoring period. In the audit, when auditor and nurse assessed the patients based on nursing documentation, only 60% of the patients fell into the same NI category, which emphasises the importance of the quality and improvement of nursing documentation since the PCM is based on the information obtained from that.

In this study, on average 77% of units (66%–88%) passed the PCM. At the same time, on average 13% (5%–29%) of the units did not participate in all PCMs, even though the PCS licence was valid. However, in these units, nurses classified the patients daily. Maintaining the reliability and validity of the RAFAELA PCS requires continuous training and supports to motivate nursing staff to see the importance and benefits of the RAFAELA PCS; otherwise, the desired benefits cannot be achieved (Lillehol et al., 2018; Porter, 2010). Frequent failures in PCM and PAONCIL measurements may cause frustration (Liljamo et al., 2017), and thus, PCSs can be seen as a time-consuming additional task in a busy nursing work environment (Fasoli & Haddock, 2010).

PCM is often completely omitted, especially during the commissioning phase. For example, in 2015, approximately 30% of the units did not do PCMs at all. The longer the units used the RAFAELA PCS, the better they succeeded in the PCMs. Over 83% of the units that

used the RAFAELA PCS for more than seven years were successful in the PCMs. In her study, Liljamo (2018) also evaluated the reliability of NI information after 20 years of clinical use of the OPCq instrument and found significant differences between units in passing the PCM over a period of six years. According to her, the explanatory factors could be the differences in the competencies of the units, commitment to using the RAFAELA PCS and motivation of the personnel.

The integration of the PCS with the systems for nursing documentation is important, but information is useless if it is not utilised (Liljamo, 2018; Mykkänen, 2019). Most nursing managers do not fully utilise the possibilities provided by the RAFAELA PCS because of a lack of time and many problems hindering the use of the RAFAELA PCS. In her study, Pusa (2007) listed such problems as the system's reliability, comparability of the information with other units, availability of real-time information, and the connection of NI data to nursing cost accounting, as well as problems with limited computer resources and data access. In addition, using and maintaining the RAFAELA PCS take time, as reported in a focus group interview of Lillehol et al. (2018).

The benefits and costs of patient classification should be examined critically because a poorly utilised PCS is a cost without benefits (Griffiths et al., 2020; Kaustinen, 2011). In this study, the fact that some of the units did not check the reliability and validity of the patient classification even once reflected either doubt on the necessity of the RAFAELA PCS's maintenance process or the laboriousness of the assessment. In addition, the fact that PCMs did not pass confirms the vulnerability of patient classification based on subjective evaluations. Maintaining the PCS as a reliable and valid NI measure requires much maintenance, training and skills to use the PCS, audits and commitment of the management to support the use of the PCS.

Units must be able to define the optimal level of NWL every two years. Based on this follow-up study, the definition of the optimal level of NWL was more successful when the units used the RAFAELA PCS over a longer period. The units ($N = 44$) that used the RAFAELA PCS during the entire nine-year follow-up period succeeded in definition of the optimal level of NWL on average 10% better than the whole follow-up group ($N = 109$). Results of these units ($N = 44$) also improved steadily during the monitoring period, and in 2019, up to 84% of these units succeeded in defining optimal level of NWL. Fagerholm's (2014) study aimed to evaluate the relationship between personnel dimensioning and NWL. Her study found that using the RAFAELA PCS is challenging and laborious when from 192 outpatient clinics only 50% managed to reach the optimal NWL level. In another study, Fasoli et al. (2011) stated, among other things, that no standard exists for the PCSs on the basis of which they could be compared, and that there is not enough validated and objective information available about the PCSs in use. In recent years, Griffiths et al. (2020, 2021) and Griffiths and Saville (2019) have had the same concerns in their later studies.

The prerequisite for the qualified use of the RAFAELA PCS system is that the system is adequately maintained and information produced by the reports is actively utilised, but do the units implement the RAFAELA PCS maintenance process as planned and agreed upon. Only slightly above 50% of the units involved in the audit regularly reviewed with the staff the reports published by the PCS. Units where the RAFAELA PCS was used more actively also succeeded in the auditing of the RAFAELA PCS well: '*The head nurse goes through NI results with the staff regularly; NI results are discussed in nurse meetings*'. The goal is that all units would pass the annual PCM and have a valid optimal NWL level. However, these results cannot be considered sufficient in relation to the objectives: '*Not even a valid parallel classification yet*', *and* '*Tried several times*'.

The goal was that all patients are classified using the PCS, checked the nursing documentation of the previous shift about the patient and verified whether the result of the classification corresponded to the patient's NI. However, of the nurses who participated in the audit, only 75% stated that they classified all patients. Some of the nurses did not have enough time to classify all patients (*'Calls are not classified, certain nurses also have patients not classified'*). For the classification to be reliable, nurses must receive sufficient guidance and training for the patient classification instrument and how to classify patients.

In their interview study in Norway, Hustad et al. (2015) examined how nursing managers had experienced the implementation and use of the RAFAELA PCS, and among other things, the challenges in implementing and using the RAFAELA PCS system came out as '*time consuming in the implementation phase, and implementing RAFAELA was challenging, and we spent a lot of resources; it was stressed that it was important for managers to use the system actively and regularly*'.

The RAFAELA PCS should be developed so that it can deliver real-time information about the NWL level in relation to the optimal workload, allowing flexible transfer of resources between different units based on the workload situation. The RAFAELA PCS should be integrated more closely with the hospital's other information systems, such as access control, staff planning and electronic patient record (EPR) systems.

4.1 | Ethical aspects (research ethics and reliability)

Good scientific practices and ethical principles were followed in this study (Resnik, 2020). The ethics committee supported, and the relevant organisation granted a research permit for this study. Permission was obtained using the monitoring data and audit results of PCS maintenance collected by the researcher. All research materials are in the form of summary tables. No information about individual patients or nurses and no personal sensitive health-related information were available to the researchers. In Finland, only the approval of the ethics committee of a healthcare organisation is required for this type of research using statistical data.

The audit model was developed using an audit created by the PCS supplier as a basis. The monitoring material for the maintenance classification was collected from the annual maintenance reports. The audit material was manually collected during an audit performed in 2017. The auditors were the mentors/responsibilities of the RAFAELA PCS, who had been introduced to auditing. Audit instructions are available to support the audits. The audit

was conducted by two auditors, who did not evaluate their own units, which aimed to increase the reliability and credibility of the audit. The audit was performed on the hospital's premises, and the audit material was not processed outside the organisation. To increase the confirmability of the research, the collection of research materials and implementation of the audit are clearly described.

4.2 | Limitations

This study confirmed that the maintenance and use of RAFAELA PCS are challenging. However, the results should be interpreted with caution due to the small sample size of the study, as only data from one hospital were used. However, the results of other studies support those of this study (Griffiths et al., 2020; Kaustinen, 2011; Lillehol et al., 2018). For a more in-depth analysis of the reasons for the results, more detailed information on the operations of the units should be collected.

Since the registry data of this study were collected retrospectively from the PCS, there was no possibility of collecting background and comparative data.

5 | CONCLUSION

Maintaining the reliability and validity of the RAFAELA PCS is a complex, time-consuming and difficult process. Implementation of maintenance requires development by increasing the training of the nursing staff, providing support for monitoring and maintaining motivation in the utilisation of NI and NWL information.

Monitoring the quality of PCS requires a systematic evaluation. In addition to PCM and PAONCIL measurements, additional information on the PCS and utilisation of NWL and NI is needed to ensure the level, quality and development of PCS in different areas. Based on the results of this study, a plan can be developed to improve the quality and reliability of the patient classification process.

Although this study was conducted in one hospital, the results are in line with previous findings and provide a similar picture of the current support and strengths and weaknesses of the RAFAELA PCS maintenance process (Fasoli & Haddock, 2010; Griffiths et al., 2020; Griffiths & Saville, 2019; Liljamo, 2018; Lillehol et al., 2018; Mykkänen, 2019; Pusa, 2007). Head nurses and PCS mentors play important roles in ensuring that units performing PCM and PAONCIL measurements receive sufficient training, support, guidance and motivation.

In future studies, background and comparative data should be collected from several hospitals, and structural and functional differences between the units should be sorted, the duration of the PCM periods, how many patients each nurse classified during the follow-up period, whether all nurses classified their patients, how many patients in the unit were classified in total, and what was the percentage of consistency of the PCM by units and also as a survey of the units' PCM and PAONCIL measurement practices.

Based on this study, we conclude that the entire RAFAELA PCS patient classification process should be simplified. Is it possible to automatically determine care based on nursing documentation? How would it work, and what would it require in a real world?

6 | RELEVANCE TO CLINICAL PRACTICE

It is a global challenge to obtain qualified nursing care with the optimal number of nurses to ensure a high level of patient safety when there is an increasing shortage of a nursing workforce. Maintaining the reliability and validity of the PCS is a continuous and challenging task. In units where positive results were realised in this study, the information obtained from PCS could be used to plan and operate daily nursing work. Maintaining reliability and validity requires continuous training of nurses and maintaining their motivation so that each patient's NI is assessed, and nursing resources are calculated correctly. A more active use of NI and NWL data for real-time coordination of nursing resources should be encouraged, and the possibility of utilising real-time NWL data would improve the allocation of nursing resources. Integration with other hospital information systems (HISs) would enable obtaining the necessary data and information directly, thereby reducing the amount of manual work.

FUNDING INFORMATION

Valtion tutkimusrahoitus (VTR), Erva-research committee record 2/2021.

CONFLICT OF INTEREST

No conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Jaana K. Junttila https://orcid.org/0000-0003-2966-4102

REFERENCES

Andersen, M. H., Lønning, K., & Fagerström, L. (2014). Testing reliability and validity of the Oulu patient classification instrument—The first step in evaluating the RAFAELA system in Norway. *Open Journal of Nursing*, *4*(4), 303–311.

Duffield, C., Diers, D., O'Brien-Pallas, L., Aisbett, C., Roche, M., King, M., & Aisbett, K. (2011). Nursing staffing, nursing workload, the work environment and patient outcomes. *Applied Nursing Research*, *24*(4), 244–255.

Fagerholm, N. (2014). *Sairaaloiden somaattisten poliklinikoiden hoitotyön henkilöstömitoitus–tunnusluku-ja asiantuntijaperusteinen mallinnus*. Doctoral Dissertations in Health Sciences, 254. University of Eastern Finland, Kuopio.

Fagerström, L., Lønning, K., & Andersen, M. H. (2014). The RAFAELA system: A workforce planning tool for nurse staffing and human resource management. *Nursing Management*, *21*(2), 30–36.

Fasoli, D. R., Fincke, B. G., & Haddock, K. S. (2011). Going beyond patient classification systems to create an evidence-based staffing methodology. *Journal of Nursing Administration*, *41*(10), 434–439.

Fasoli, D. R., & Haddock, K. S. (2010). Results of an integrative review of patient classification systems. *Annual Review of Nursing Research*, *28*(1), 295–316.

Finnish Consulting Group. (2012). *RAFAELA Patient Classification System, Instructions OPCq version 1.4, 2012*. FCG Knowledge Management Ltd. http://www.RAFAELA.Fi/?Q=en. Accessed November 10, 2017.

Finnish Consulting Group. (2016). *RAFAELA Patient Classification System, Instructions PAONCIL version 2.1, 2016*. FCG Knowledge Management Ltd. http://www.RAFAELA.fi/?q=en. Accessed November 10, 2017.

Frilund, M., & Fagerström, L. (2009). Managing the optimal workload by the PAONCIL method – A challenge for nursing leadership in care of older people. *Journal of Nursing Management*, *17*, 426–434. https://doi.org/10.1111/j.1365-2834.2009.01013.x

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: Key concepts, approaches, and applications. *Research in Social & Administrative Pharmacy: RSAP*, *9*(3), 330–338. https://doi.org/10.1016/j.sapharm.2012.04.004

Griffiths, P., & Saville, C. (2019). *Tools & systems to determine nurse staffing requirements: An evidence review*. University of Southampton http://eprints.soton.ac.uk/id/eprint/428284

Griffiths, P., Saville, C., Ball, J., Jones, J., Pattison, N., Monks, T., & Safer Nursing Care Study Group. (2020). Nursing workload, nurse staffing methodologies and tools: A systematic scoping review and discussion. *International Journal of Nursing Studies*, *103*, 103487.

Griffiths, P., Saville, C., Ball, J. E., Jones, J., Monks, T., & Safer Nursing Care Tool study team. (2021). Beyond ratios-flexible and resilient nurse staffing options to deliver cost-effective hospital care and address staff shortages: A simulation and economic modelling study. *MedRxiv*, *117*, 103901.

Hustad, N. B., Hellesø, R., & Andersen, M. H. (2015). A qualitative study of manager experiences using the RAFAELA system. *Open Journal of Nursing*, *5*, 1224–1232.

Junttila, J. K., Koivu, A., Fagerström, L., Haatainen, K., & Nykänen, P. (2016). Hospital mortality and optimality of nursing workload: A study on the predictive validity of the RAFAELA nursing intensity and staffing system. *International Journal of Nursing Studies*, *60*, 46–53. https://doi.org/10.1016/j.ijnurstu.2016.03.008

Junttila, J. K., Koivu, A., Haatainen, K., & Nykänen, P. (2019). Nurse-assessed optimality of workload—A valid measure for the adequacy of nursing resources? *Online Journal of Nursing Informatics*, *23*(3). https://www.himss.org/resources/nurse-assessed-optimality-workload-valid-measure-adequacy-nursing-resources

Kaustinen, T. (2011). Oulu patient classification and use of working hours in nursing from the perspective of nursing care quality standards. In *Doctoral dissertation in acta universitatis Ouluensis D Medica 1134, 2011*. University of Oulu Finland. http://urn.fi/urn:isbn:9789514296437

Liljamo, P. (2018). *Correspondence between clinical and administrative care information. Reliability of care information and the possibilities of secondary use of structured care report information in determining care*. Doctoral Dissertation in Social ScienCES and Business Studies. University of Eastern Finland Kuopio, Finland. https://erepo.uef.fi/handle/123456789/19914

Liljamo, P., Kinnunen, U., Ohtonen, P., & Saranto, K. (2017). Quality of nursing intensity data: Inter-rater reliability of the patient classification after two decades in clinical use. *Journal of Advanced Nursing*, *73*(9), 2248–2259.

Liljamo, P., Kinnunen, U.-M., & Saranto, K. (2020). Assessing the relation of the coded nursing care and nursing intensity data: Towards the exploitation of clinical data for administrative use and the design of nursing workload. *Health Informatics Journal*, *26*(1), 114–128.

Lillehol, B. M., Lønning, K., & Andersen, M. H. (2018). Exploring nurse managers' perception of using the RAFAELA system as a management tool in a Norwegian hospital setting. *Nursing Open*, *5*(1), 77–83.

Mykkänen, M. (2019). *Tietorakenteet hoitotietojen käsittelyssä ja potilastiedon toisiokäytössä*. Dissertation in Social Sciences and Business Studies. University of Eastern Finland https://erepo.uef.fi/handle/123456789/21249

Needleman, J. (2015). Nurse staffing: The knowns and unknowns. *Nursing Economic$*, *33*(1), 5–7.

Oostveen, C. J., Ubbink, D. T., Mens, M. A., Pompe, E. A., & Vermeulen, H. (2016). Pre-implementation studies of a workforce planning tool for nurse staffing and human resource management in university hospitals. *Journal of Nursing Management*, *24*(2), 184–191.

Porter, M. E. (2010). What is value in health care? *The New England Journal of Medicine*, *363*(26), 2477–2481.

Pusa, A.-K. (2007). *The right nurse in the right place. Nursing productivity and utilization of the RAFAELA patient classification system in nursing management*. Dissertation in Faculty of Social Sciences. University of Kuopio.

Rauhala, A. (2008). *The Validity and Feasibility of measurement tools for human resources management in Nursing Case of the RAFAELA system*. Doctoral dissertation in Health Policy and Management. University of Kuopio.

Rauhala, A., & Fagerström, L. (2004). Determining optimal nursing intensity: The RAFAELA method. *Journal of Advanced Nursing*, *45*(4), 351–359.

Resnik, D. B. (2020). *What is ethics in research & why is it important* (Vol. *1*). National Institute of Environmental Health Sciences https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm?links=falseWijaya

Virkkunen, H., Mäkelä-Bengs, P., & Vuokko, R. (2015). *Terveydenhuollon rakenteisen kirjaamisen opas-keskeisten kertomusrakenteiden kirjaaminen sähköiseen potilaskertomukseen, osa I, versio 2015*. Finnish Institute for Health and Welfare https://www.julkari.fi/bitstream/handle/10024/126020/URN_ISBN_978-952-302479-3.pdf?sequence=1

Vuokko, R., Mäkelä-Bengs, P., Hyppönen, H., Lindqvist, M., & Doupi, P. (2017). Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data. *International Journal of Medical Informatics*, *97*, 293–303.

West, E., Barron, D. N., Harrison, D., Rafferty, A. M., Rowan, K., & Sanderson, C. (2014). Nurse staffing, medical staffing and mortality in intensive care: An observational study. *International Journal of Nursing Studies*, *51*(5), 781–794.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Junttila, J. K., Haatainen, K., Koivu, A., & Nykänen, P. (2023). How the reliability and validity of the patient classification system can be ensured in daily nursing work? A follow-up study. *Journal of Clinical Nursing*, *32*, 3720–3729. https://doi.org/10.1111/jocn.16559