JAN ORIGINAL RESEARCH

# Benchmarking nurse staffing levels: the development of a nationwide feedback tool

Koen Van den Heede, Luwis Diya, Emmanuel Lesaffre, Arthur Vleugels & Walter Sermeus

Accepted for publication 14 April 2008

Correspondence to K. Van den Heede:
e-mail: koen.vandenheede@med.kuleuven.be

Koen Van den Heede MSc RN
Research Associate
Center for Health Services and Nursing Research, Katholieke Universiteit Leuven, Leuven, Belgium

Luwis Diya MSc
Research Associate
Biostatistical Centre, Katholieke Universiteit Leuven, Leuven, Belgium

Emmanuel Lesaffre PhD
Director
Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands

Arthur Vleugels PhD MD
Director
Centre for Health Services and Nursing Research, Katholieke Universiteit Leuven, Leuven, Belgium

Walter Sermeus PhD FEANS RN
Programme Director
Centre for Health Services and Nursing Research, Katholieke Universiteit Leuven, Belgium

VAN DEN HEEDE K., DIYA L., LESAFFRE E., VLEUGELS A. & SERMEUS W. (2008) Benchmarking nurse staffing levels: the development of a nationwide feedback tool. *Journal of Advanced Nursing* **63**(6), 607–618
**doi: 10.1111/j.1365-2648.2008.04724.x**

## Abstract

**Title. Benchmarking nurse staffing levels: the development of a nationwide feedback tool.**

**Aim.** This paper is a report of a study to develop a methodology that corrects nurse staffing for nursing care intensity in a way that allows nationwide benchmarking of nurse staffing data.

**Background.** Although nurse workload measurement systems are recognized to be informative in nurse staffing decisions, they are rarely used. When these systems are used, however, it is only possible to compare units within hospitals, because currently available instruments are not standardized for comparisons beyond hospital boundaries. The Belgian Nursing Minimum Dataset (B-NMDS) contains uniformly measured data about the intensity of nursing care and nurse staffing levels for all hospitals in Belgium.

**Method.** We conducted a retrospective multilevel analysis of the B-NMDS for the year 2003. The sample included 690,258 inpatient days for 298,691 patients, recorded from 1637 acute care nursing units in 115 hospitals. We corrected the number of nursing staff by using different covariates available in the B-NMDS: intensity of nursing care, type of day (week vs. weekend), service type (general vs. intensive) and hospital type (academic vs. general).

**Findings.** The multilevel approach allowed us to explain about 70% of the variability in the number of nursing staff per nursing unit using hospital type ($P = 0{\cdot}0053$); intensity of nursing care ($P < 0{\cdot}0001$) and service type ($P < 0{\cdot}0001$) as the only covariates.

**Conclusion.** The feedback tool we developed can inform nurse managers and policymakers about nursing intensity-adjusted nurse staffing levels according to different benchmarks. Our study demonstrates that investing in large nursing datasets is appropriate for the international nursing community.

**Keywords:** benchmarking, intensity of nursing care, multilevel model, nurse staffing, nursing minimum dataset, nursing shortage, nursing workload, patient classification system

## Introduction

Recently, gaining more knowledge about nurse staffing issues has attracted much interest. This awareness is based on a global nursing shortage that threatens the sustainability of many healthcare systems. In many countries, the availability of nurses is insufficient to meet all healthcare demands (Janiszewski 2003, Rauhala *et al.* 2007). Moreover, this situation is anticipated to worsen in coming years. On the one hand, the supply of nurses continues to decrease because of improved career opportunities for women, a growing proportion of the nursing workforce about to retire, fewer funds for nursing school students, more family career obligations, societal trends towards reducing work hours and towards earlier retirement; poor working conditions (including poor pay), improved career opportunities outside the healthcare sector, and a poor image of the nursing profession. On the other hand, the demand for nurses continues to increase because of an ageing population and shortening lengths of stay in hospitals, both of which increase the intensity of nursing care (Plati *et al.* 1998, Buerhaus *et al.* 2000, Kirk 2007). Shortfalls in the nurse workforce have striking implications in light of a large and growing base of research literature demonstrating an association between nurse staffing and both nurse and patient outcomes in hospitals (Aiken *et al.* 2002a, 2003, Needleman *et al.* 2002, Kane *et al.* 2007, Rafferty *et al.* 2007). Thus, scarce nursing resources should be allocated appropriately so that excessive workload (and its negative impact on nurse and patient outcomes) is avoided (Rauhala *et al.* 2007). Therefore, it is necessary to implement efficient and effective mechanisms to address workload issues (McGillis-Hall *et al.* 2006). However, although there is general agreement that measuring nursing workload is important, in reality nurse staffing decisions are more often based on professional judgment or 'gut' feelings, which are confined within budget parameters, than on data from workload measurements (McGillis-Hall *et al.* 2006). In cases in which data are used to guide staffing decisions, these data are often limited to measuring the number of nursing hours per patient day (NHPPD). The pitfall of this approach is the underlying assumption that all patients and patient days are equal, when in fact nursing care intensity varies significantly across patients and over the length of stay of individual patients (Graf *et al.* 2003, Welton *et al.* 2006). Therefore, relying exclusively on NHPPD data generates inadequate figures of nursing resource deployment. If the goal is to benchmark nurse staffing – an area of nursing research that is fairly undeveloped – then the intensity of nursing care needs to be considered.

## Background

Confusion often exists in the nursing workload literature stemming from the loose usage of terms. Terms such as 'nursing workload', 'patient dependency', 'nursing intensity', 'patient acuity', 'complexity of care' and even 'severity of illness' are often used interchangeably. In a recent paper, Morris *et al.* (2007) clarified the nursing workload concept, describing it as encompassing both nursing intensity and non-patient care-related nursing activities. However, the concepts of patient dependency, severity of illness, complexity of care, and time taken to carry out nursing work are subsumed under the umbrella term nursing intensity. Nursing intensity comprises direct (e.g. administration of medication, assessment of blood pressure) as well as indirect (e.g. making a phone call on behalf of the patient, ordering medication) care activities that are performed on the behalf of patients. Non-patient care-related nursing activities are, for example, attending courses to enhance student education, attending staff meetings and supporting unit management (Morris *et al.* 2007).

During the last 50 years, nurse leaders and researchers have traditionally developed patient classification systems for nursing care in an attempt to measure the workload of nurses (De Groot 1989). Two main ways of categorizing patients can be distinguished: the summative task type and the critical indicator or criterion type (Seago 2002). The summative task type requires nurses to record a list of the majority, if not all, of the nursing tasks or activities that may be required to care for patients. Each activity has a corresponding weight or number of points associated with it. The total sum of these acuity points gives an estimate of a patient's nursing care requirements (Seago 2002). A widely used example is the Projet de Recherche en Nursing system (Tilquin 1987). The critical indicator or criterion type classifies patients based on a combination of critical indicators of care, which are based on conceptually distinct categories of patient care requirements rather than on individual activities, tasks or procedures. Each care indicator is ranked according to the level of care required by a patient for that particular indicator. The overall patient category is determined by scoring across and between the indicators. A well-known example is the San Joaquin Patient Classification System (Murphy *et al.* 1978).

The primary (and mostly only) use of patient classification systems is limited to constructing an instrument that allows nurse managers to compare workloads across units within hospitals so that they can more objectively document their labour budgets as they negotiate within the hospital (Seago 2002). Most nurses are not informed about or involved in the way in which patient classification systems are used. Nurses just complete a piece of paper (or the electronic equivalent), and do not realize the benefits of doing so. As a result, nurses view completing patient classification system reports as an extra, meaningless burden (McGillis-Hall *et al.* 2006). Of the numerous existing patient classification systems, of which many are commercially marketed, only a restricted number have found large-scale applications. However, none has proven to be superior (Giovannetti & Johnson 1990, Fagerström *et al.* 2000, Seago 2002). No patient classifications and definitions of nursing categories have been nationally standardized (Upenieks *et al.* 2007), and all measure only a fraction of the above-mentioned subconcepts that fall within the nursing workload concept. Furthermore, most patient classification systems are designed to be unit specific, which complicates cross-unit comparison. With the exception of a recent initiative in Finland (Fagerström & Rauhala 2007), to date researchers and managers have generally been unable to incorporate patient classification systems into large-scale nurse staffing benchmarking initiatives. Acquiring large-scale data pertinent to different specialties will only be possible if the great majority of hospitals using identical patient classification systems agree to pool their staffing data. However, Clarke (2003) does not foresee this happening in the near future. Nevertheless, Belgium is in a unique position to fill the gap in this research area, as the Belgian Nursing Minimum Dataset (B-NMDS) contains information about nurse staffing levels and the intensity of nursing care in all nursing units in all Belgian acute hospitals (Sermeus *et al.* 2007a).

## The study

### Aim

The aim of the study was to develop a methodology that corrects nurse staffing for nursing care intensity in a way that allows nationwide benchmarking of nurse staffing data.

### Design

A retrospective analysis of cross-sectionally collected data was conducted.

### Sample

Data from the year 2003 on nurse staffing and nursing activities were available from all 115 Belgian acute care hospitals. The study sample was restricted to general surgical, internal medicine, mixed internal medicine and surgical, elder care and intensive care units because the intensity of care measure was validated for these types of units (Sermeus *et al.* 2007a). The final sample included 690,258 inpatient days for 298,691 patients recorded from 1637 units in 115 acute hospitals. The nursing unit types were used to classify the sample into five groups: surgical nursing units (department type C; 232,282 inpatient days, 511 units); internal medicine nursing units (department type D; 254,139 inpatient days, 525 units); elder care nursing units (department type G; 106,829 inpatient days, 234 units); mixed surgical and internal medicine nursing units (department type H*; 60,335 inpatient days, 140 units); and intensive care nursing units (department type I; 36,673 inpatient days, 227 units).

### Data collection

Since 1988, all Belgian acute hospitals have been obliged by law to submit to the Ministry of Public Health data about nurse staffing levels and nursing activities. Data were submitted quarterly (March, June, September and December), and these data form the basis of the B-NMDS. Data intended for the B-NMDS were recorded during the first 15 days of March, June, September and December, during which one recording takes place every 24 hours. However, for 5 days (one of which always includes a weekend day) of each recording period, the data must be submitted for entry into the national database. The Ministry of Public Health randomly chooses the specific 5-day period.

### Ethical considerations

The study committee appointed by the Ministry of Science Policy gave ethics clearance for the study. The data provided by the Ministry of Public Health of Belgium did not contain information about the identity of patients or hospitals.

### Validity and reliability

Bedside nurses record 23 nursing activities for the B-NMDS (Appendix 1). Nurse managers record the number of hours that nursing staff work during the recording days. The number of nursing staff is expressed as NHPPD, which is the sum of the staffed hours of Registered Nurses (bachelor's degree prepared and second level nurses) divided by the number of inpatient days per nursing unit per observation day. The inter-rater reliability of the recording was 78·8%, as tested by the Kendall's tau-b statistic (Sermeus 1992). During regular hospital visits, a random selection of patient records is reviewed by the Ministry of Public Health to ensure that data were coded correctly. However, the results of these audits are not disclosed in the public domain (Sermeus *et al.* 2007a).

By means of non-linear principal component analysis, Sermeus *et al.* (2007a) showed that the 23 nursing activities could be aggregated into one measure (weight 1) that describes the intensity of nursing care in acute hospitals (Sermeus *et al.* 2007a). Two alternative systems (weight 2 and weight 3) are broadly used in Belgian hospitals to estimate the intensity of nursing care from the B-NMDS and have been used in the past in hospital financing systems (Sermeus *et al.* 2007b). Both systems assign to each of the 23 nursing activities (per response category) a number of points (Appendix 1) (Closon 1991, Demeyere 2001). The summed points result in an 'intensity of nursing care score' for each patient. The points were developed by using a Delphi-like approach (Closon 1991, Demeyere 2001). Yet, no further validation work was performed. Therefore, in this paper the aggregate intensity of nursing care score obtained through non-linear principal component analyses (Sermeus *et al.* 2007a) was taken as the reference, whereas weight 2 and weight 3 were used for cross-validation. The correspondence between the three intensity of nursing care variables (weight 1, weight 2 and weight 3) was evaluated using the Spearman rank correlation. A high correlation was obtained between the three different measures (weight 1, weight 2 and weight 3) used to compute the intensity of nursing from the B-NMDS (Table 1). Spearman rank correlations ranged from 0·91 to 0·98 and from 0·80 to 0·99, when all nursing units or nursing units per unit type were considered.

## Data analysis

The ultimate objective of our data analyses was to develop a model that allows the comparison of nurse staffing levels across nursing units and hospitals. To achieve fair comparisons, our model must control for different characteristics that can influence nurse staffing levels. In this study, we could only control for factors that are available in the B-NMDS: intensity of nursing care (weight 1); type of day (week vs. weekend); type of nursing unit (general acute care nursing units vs. intensive care nursing units) and type of hospital (academic vs. general hospital). During weekdays, more surgical procedures, investigations and admission/discharges/transfers take place compared with weekend days, thus justifying higher staffing levels for weekdays. Furthermore, intensive care units treat very severe ill patients for which surveillance and continuous monitoring is important. Thus, more permanent staff are required to work in these units than are reflected by the intensity of nursing care alone. Lastly, nursing staff in academic hospitals are more involved in non-patient care-related activities (e.g. research, staff training) than staff in general hospitals. The non-patient activities of nurses working in academic hospitals thus justify higher nurse staffing levels in this group.

The number of nursing staff (i.e. NHPPD) is measured for each B-NMDS observation day and depends on the characteristics of that particular day, the nursing unit, and the hospital. Thus, our dataset has a three-levelled structure: B-NMDS observation days (level 1) nested within nursing units (level 2); and nursing units, in turn, nested within hospitals (level 3). In a conventional regression model, all of the unmodelled contextual information (e.g. architecture of the nursing unit, nurse staffing model, organizational culture and nurse experience levels) ends up pooled into a single individual error term of the model. This is problematic because nursing units within a hospital are presumably correlated (e.g. nursing units within a hospital are influenced by the same organizational climate, which is one of the unmeasured factors). Ignoring this correlation induces artificially smaller standard errors, which would lead to a higher possibility of concluding that the results are statistically significant, even though they may not be (Hox 2002, Luke 2004, Park & Lake 2005). Therefore, multilevel models are increasingly employed in the nursing research field (Cho 2003, Park & Lake 2005, Adewale *et al.* 2007). These models are specifically designed to correct the dependency of observations within a cluster and to allow disentangling the components of variability pertaining to the different levels (Luke 2004).

We performed a linear multilevel modelling approach with the Proc Mixed procedure of SAS v9·1 (SAS Institute 2001). In the first step, we calculated intraclass correlation coefficients

**Table 1** Spearman rank correlations between three different nursing intensity measures deduced from the Belgian Nursing Minimum Data Set

| Correlation | All units | C | D | G | H* | I |
|---|---|---|---|---|---|---|
| Weight 1–weight 2 | 0·91 | 0·82 | 0·89 | 0·87 | 0·86 | 0·93 |
| Weight 2–weight 3 | 0·98 | 0·97 | 0·98 | 0·98 | 0·97 | 0·99 |
| Weight 1–weight 3 | 0·91 | 0·80 | 0·89 | 0·83 | 0·85 | 0·91 |

Level of measurement: nursing unit per observation day (all units, *n* = 30,722; C, *n* = 9527; D, *n* = 9947; G, *n* = 4458; H*, *n* = 2540; I, *n* = 4250).
Weight 1 (Sermeus *et al.* 2007a); weight 2 (Closon 1991); weight 3 (Demeyere 2001).

for the null model, that is, a model without covariates. The purpose of this step was to capture the variability in the number of nursing staff (NHPPD) accounted for by the different levels (Appendix 2a). In the second step, we included the covariates of each measurement level in the model: weekday vs. weekend day and intensity of nursing care (level 1); service type (level 2) and hospital type (level 3). We applied a random intercept model (Appendix 2b and c). In this type of multilevel model, the intercept of each hospital varies (i.e. the average nurse staffing levels within a hospital have a different baseline starting point, depending on the unmeasured context of that particular hospital). The residual ICC were calculated from the model with covariates (Appendix 2a). This reflects the unexplained variability remaining in these levels after controlling for the covariates. Note that the aim of study reported in this paper was not to predict the nurse staffing levels of a particular nursing unit within a hospital (taking into account the noise from that particular unit) but to obtain an 'average' or expected nurse staffing level, given the covariates.

As random intercept models take into account the noise (or unmeasured context factors) of particular nursing units, we averaged the (estimated) random intercepts of the nursing units within hospitals and between hospitals (Appendix 2d). This resulted in an (estimated) expected number of nursing staff (i.e. expected NHPPD) per unit per B-NMDS observation day. The expected NHPPD gives us a picture of how the available pool of nurses nationwide can be best shared by different nursing units. Next, we compared the expected NHPPD with the actual staffing patterns (i.e. observed NHPPD) for each nursing unit. To facilitate interpretation for nurse managers, we multiplied this index by the crude NHPPD (overall sample average):

$$\text{Nursing-intensity adjusted NHPPD}_{ijk} = \frac{\text{Observed NHPPD}_{ijk}}{\text{Expected NHPPD}_{ijk}} \times \text{Crude NHPPD} \qquad (1)$$

where the fraction (equation 1) corresponds to the $k$th observation day in the $j$th nursing unit of the $i$th hospital. A weighted census of the number of nursing staff is obtained by multiplying the fraction with the national average (crude NHPPD). The weighted census of the number of nursing staff is controlled for by the intensity of nursing care and other covariates.

*Feedback tool development*

We developed a feedback tool based on our statistical model so that feedback can be obtained from two levels. Nurse staffing data can be aggregated at the level of the hospital or the nursing unit. At the hospital level, hospitals will be able to compare data about their hospital with different benchmarks (all hospitals or a selection of hospitals of similar size or type). Next, hospitals can focus on the nursing unit level (i.e. compare data about one of their nursing units with all Belgian nursing units, with all Belgian nursing units of the same type, or with all nursing units within the same hospital). The feedback tool provides both uncorrected (observed NHPPD) as well as corrected (i.e. nursing-intensity adjusted NHPPD) data. The 'turnip' plot (The Center for the Evaluative Clinical Sciences at Dartmouth 1999) is then used to illustrate in a graphical way the variability of nursing-intensity adjusted NHPPD among Belgian hospitals. The position of a particular hospital/nursing unit is indicated by an asterisk. This graphical information is complemented by descriptive data for the corrected as well as the uncorrected data.

## Results

### General descriptive results

The average observed NHPPD in the study sample was 2·74. Table 2 summarizes NHPPD as a function of hospital, nursing unit and time characteristics. Academic hospitals had a higher median number of nursing staff than general hospitals (0·5–0·7 NHPPD for general acute units, and 1·2–2·1 NHPPD for intensive care units). In addition, staffing in intensive care units was considerably higher (4·5 times more) than that in acute general units. Finally, staffing patterns on weekdays and weekend days did not differ substantially.

### Model results

The null model showed that variability in nursing staff numbers was mostly attributed to nursing unit level (86·56%), followed by observation day level (12·07%) and hospital level (1·37%). The inclusion of the set of covariates and their interaction reduced the overall variance by 69·72%. Table 3 shows that unexplained variability was mainly attributed to nursing unit level (0·0656% or 53·8%) and observation day level (0·0479% or 39·3%) rather than to hospital level (0·0084% or 6·9%). Table 3 also presents the fixed effects coefficient estimates for the multilevel regression model. The association of the regressors hospital type ($P = 0{\cdot}0053$), intensity of nursing care ($P < 0{\cdot}0001$), and service type ($P < 0{\cdot}0001$) on the response NHPPD were highly statistically significant. As type of day interacts with intensity of nursing care and hospital type, this covariate was retained in the model.

**Table 2** Distribution of nursing hours per patient day

| | | General acute care units (C, D, E, H*) | | Intensive care units (I) | |
|---|---|---|---|---|---|
| Hospital type | Type of day | P50 | (P25–P75) | P50 | (P25–P75) |
| Academic | Weekdays | 2·72 | (2·29–3·30) | 12·11 | (10·18–13·83) |
| | Weekend days | 2·82 | (2·36–3·44) | 11·08 | (9·49–12·57) |
| General | Weekdays | 2·22 | (1·87–2·68) | 10·00 | (8·38–12·33) |
| | Weekend days | 2·13 | (1·74–2·63) | 9·84 | (8·38–12·40) |

Level of measurement: nursing unit per Belgian Nursing Minimum Data Set observation day. P25 = 25th percentile; P50 = 50th percentile or median; P75 = 75th percentile; C = surgical units; D = internal medicine units; E = acute elder care units; H* = mixed units; I = intensive care units.

**Table 3** Results of multilevel model with nursing hours per patient day as a response

| Effect | Estimate (SE) | *t* value | *P* value |
|---|---|---|---|
| *Fixed effects* | | | |
| Intercept | 0·6999 (0·0154) | 45·45 | < 0·0001 |
| Type of day (0 = weekday; 1 = weekend day) | –0·0194 (0·0109) | –1·78 | 0·0748 |
| Service type (0 = general units; 1 = intensive care units) | 1·2628 (0·0463) | 27·28 | < 0·0001 |
| Log of weight 1 | 0·1936 (0·0149) | 13·03 | < 0·0001 |
| Hospital type (0 = general; 1 = academic) | 0·1358 (0·0478) | 2·84 | 0·0053 |
| Type of day × service type | 0·0438 (0·0147) | 2·97 | 0·0030 |
| Type of day × log of weight 1 | –0·0445 (0·0152) | –2·92 | 0·0035 |
| Type of day × hospital type | 0·1301 (0·0217) | 6·00 | < 0·0001 |
| Service type × log of weight 1 | 0·0785 (0·0322) | 2·44 | 0·0147 |
| Service type × hospital type | –0·5534 (0·1474) | –3·75 | 0·0002 |
| log of weight 1 × hospital type | 0·1024 (0·0359) | 2·85 | 0·0044 |
| Type of day × log of weight 1 × hospital type | –0·0850 (0·0242) | –3·52 | 0·0004 |
| Service type × log of weight 1 × hospital type | 0·2199 (0·0985) | 2·23 | 0·0256 |
| *Covariance parameters* | | | *RICC* |
| Hospital | 0·0084 (0·0018) | | 0·0689 |
| Nursing Unit | 0·0656 (0·0025) | | 0·5384 |
| Observation day level | 0·0479 (0·0004) | | 0·3927 |
| Observations | | 30722 | |
| Model deviance | | –682·4 | |
| Change in deviance from null | | 2948·9 | |

Weight 1: nursing intensity measure (Sermeus *et al.* 2007a).
RICC, residual intraclass correlation coefficient.

### Nursing-intensity adjusted NHPPD

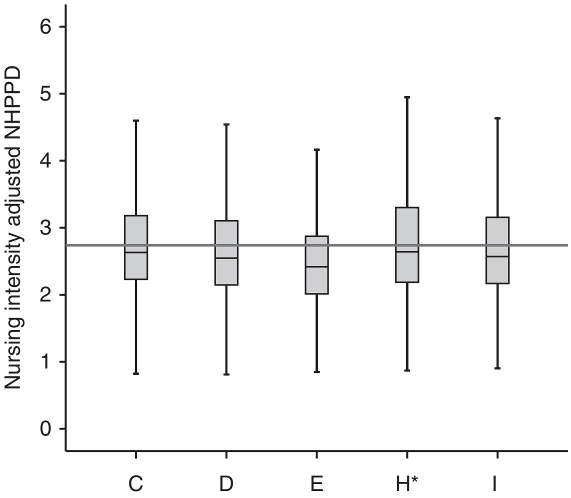
As can be seen in Figure 1, the majority of nursing units (per observation day) were staffed below the national average (indicated by bold line) when a correction for intensity of nursing care was included.

Of nursing unit types (Figure 1), acute elder care units clearly deviated most from the national staffing average. Sixty-nine per cent of observation days in elder acute care units were staffed below the national NHPPD average of 2·74. For surgical nursing units, internal medicine nursing units, mixed nursing units and intensive care nursing units, 56%, 60%, 55% and 58% of the observation days, respectively, were staffed below the national NHPPD average.

### Feedback tool

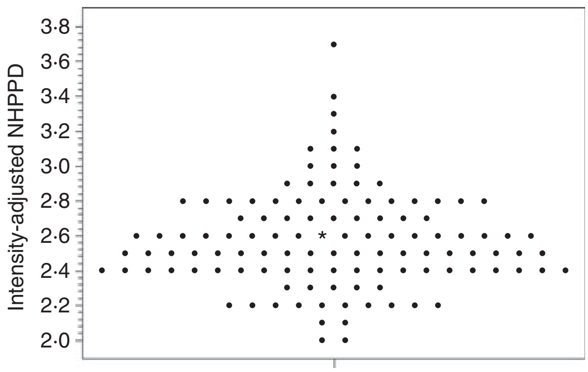
Our statistical modelling resulted in the development of a feedback tool, which will be made available online by the Ministry of Public Health on a secure website. Hospitals can generate different printouts for different analysis levels and for different benchmarks. Nurse staffing data can be aggregated at hospital level or nursing unit level. For the hospital level, different benchmarks can be chosen: all hospitals; a selection of hospitals with a similar size ( < 200 beds, 200–299 beds, 300–449 beds, ≥450 beds); or hospital type (general, academic). For the nursing unit level, printouts can be obtained for the following benchmarks: all Belgian nursing units; all Belgian nursing units of the same type




| | C | D | E | H* | I |
|---|---|---|---|---|---|
| P25 | 2·23 | 2·15 | 2·01 | 2·19 | 2·17 |
| P50 | 2·63 | 2·55 | 2·42 | 2·64 | 2·57 |
| P75 | 3·18 | 3·11 | 2·87 | 3·30 | 3·17 |

**Figure 1** Distribution of nursing-intensity adjusted nursing hours per patient day (NHPPD), measured for each nursing unit per observation day for five different unit types. The horizontal bold line in the figure corresponds to the average observed NHPPD in the overall sample. C = surgical; D = internal medicine; E = acute elder care; H* = mixed internal and surgical units; I = intensive care units.




| | Hospital value | P10 | P25 | P50 | P75 | P90 | P90:P10 ratio |
|---|---|---|---|---|---|---|---|
| Uncorrected NHPPD | 2·70 | 2·21 | 2·40 | 2·58 | 2·81 | 3·10 | 1·40 |
| Intensity-adjusted NHPPD | 2·60 | 2·22 | 2·42 | 2·54 | 2·75 | 2·93 | 1·32 |

**Figure 2** Feedback tool results for the number of nursing staff (NHPPD) at the hospital level of one hospital benchmarked with all other Belgian hospitals. Each dot represents one of 115 Belgian acute hospitals. We used a multilevel model to correct nurse staffing levels for differences in intensity of nursing care, type of service, type of hospital and type of day. The asterisk represents the position of the specific hospital for which the data were benchmarked. P10 = 10th percentile; P25 = 25th percentile; P50 = 50th percentile or median; P75 = 75th percentile; P90 = 90th percentile.

(surgical units, internal medicine units, mixed units, elder care units and intensive care units); all nursing units from the same type or size of hospitals; or all units of the same hospital. Both uncorrected and corrected data are provided in the feedback tool.

Figure 2 shows sample results from using the feedback tool to benchmark 'one hospital' against 'all other hospitals'. The position of the hospital using the feedback tool is represented by an asterisk. In this particular case, both the corrected and uncorrected NHPPDs fall between the median and the 75th percentile.

## Discussion

The average NHPPD in Belgian acute hospitals is 2·74, a figure that is situated towards the lower boundaries of nurse staffing levels in countries such as the United States of America (USA) (Landon *et al.* 2006). The potential origins of the large differences in nurse staffing levels between countries are numerous (e.g. different nursing roles, different hospital case mix and different social status of the nursing profession) and need to be further explored. Nevertheless, we found that variability in nurse staffing levels mainly occurs at nursing unit level. This finding necessitates the measurement of nurse staffing levels at nursing unit level, which agrees with Blegen (2006), who suggested that comparing nurse staffing levels should preferably be performed at nursing unit level rather than at hospital level. Therefore, in the present study we explored a novel strategy to benchmark nurse staffing data on a large scale (in this case, nationwide), allowing comparisons at both nursing unit and hospital levels. The goal of this approach was to correct the number of nursing staff for the intensity of nursing care and other covariates (i.e. type of day, type of service and type of hospital) using a multilevel regression model. The included set of covariates reduced the overall unexplained variance in number of nursing staff by 69·72%. Applying the study approach to all Belgian acute hospitals, we found that the majority of nursing units (acute elder care nursing units in particular) is staffed below the national NHPPD average of 2·74. However, the figures generated from the feedback tool do not indicate the optimal or evidence-based nurse staffing levels.

### Study limitations

The most important limitation of this study is that we only redistributed the national available nurse staffing pool according to factors available in the data sources at hand (intensity of nursing care, hospital type, service type and type of day). An external criterion is preferred to evaluate whether nursing units are under- or over-staffed relative to what is required. One possible way to determine whether a unit is under- or over-staffed is to survey nurses. For example, nurses could be asked whether enough nurses are available to provide optimal care. In Finland, a nurse survey showed that, in about 50% of the observed days, the optimal nursing care intensity level was exceeded (Fagerström & Rauhala 2007). In the USA, Canada, Scotland, England and Germany, 62–71% of nurses indicated that nurse staffing levels were insufficient for providing high-quality care (Aiken *et al.* 2002b). By contrast, in Belgium, only 41·3% of the nearly 10,000 hospital nurses surveyed indicated that they were concerned that they could not provide the level of care they wished (Milisen *et al.* 2006). Thus, surveys can complement the information generated by our feedback tool by supplying an external criterion. Nevertheless, information obtained by surveys should also be interpreted carefully as this information is, by definition, subjective and is thus not necessarily evidence based.

There are some other limitations inherent to the methodology applied in this study. First, the nursing-intensity adjusted NHPPD, although identical for two hospitals/nursing units, does not always yield a similar amount of workload. Despite the fact that large amounts of nursing data are uniformly recorded in all nursing units of all acute hospitals, an important limitation of this measure is the large number of unmeasured factors, such as staffing models, technical support, medical staff organization, the experience level of the staff, the numbers and types of non-direct care clinical supports and operational staff, the amount of technology and the physical layout of the unit. These factors are so complex and numerous, and differ so much from organization to organization, that it is unlikely that one or more current (and probably also future) instruments could uniformly capture them across a large number of hospitals. Furthermore, the majority of nurses' working hours is still allocated to nursing interventions related to individual patient care (Rauhala *et al.* 2007), which makes it worthwhile to benchmark nurse staffing data that are corrected for the intensity of nursing care.

Second, the measurement instrument for the B-NMDS used in this study is nearly 20 years old. Thus, we need to investigate how our present model can be applied to the revised B-NMDS, which will be implemented in all Belgian hospitals from 2008 onwards (Sermeus *et al.* 2005). The revised B-NMDS is expected to capture the intensity of nursing care from contemporary nursing practice better because it includes nursing activities that are adjusted to current nursing practice standards and pertains to an increased number of possible nursing activities.

### Study strengths

Despite these limitations, our study does have several strengths. It is, to our knowledge, the first attempt to apply a sound methodology that adjusts the number of nursing staff for the intensity of nursing care on a large scale with uniformly measured data. The feedback tool goes beyond most patient classification systems, which are designed only to be unit specific and do not facilitate cross-unit or cross-hospital comparison. As a result, the feedback tool has great potential for nursing management and policy applications. The Ministry of Public Health can use this instrument to monitor general patterns in nurse staffing data. For example, our finding that elder care units have the highest proportion of under-staffed units (relative to average Belgian figures) could be further explored. Do these units have more clinical support staff than other units or is it necessary to increase the minimum staffing ratios for elder care units? Additionally, chief nurse executives or unit managers can use the feedback tool as a barometer to help them assess the current nurse staffing allocation patterns (or budget) of their hospital or nursing units. Information obtained from the feedback tool can be merged with data from the hospital's management information system to complete each nursing care unit's profile by adding other attributes (e.g. experience level of nursing staff, admissions/discharges/transfers).

An important strength of our study is that it will help fill an important gap in research on nurse staffing and patient outcomes. This is especially so when our findings are linked to Belgian patient outcomes data via the Belgian Hospital Discharge Dataset. Over the last 5 years, the research literature describing possible ties between nurse staffing and patient outcomes has steadily grown. Whilst some claim that the findings are conclusive, others claim that they are tentative at best because of the many unresolved technical research issues involved (Clarke 2003). An important technical issue involves the correction of nurse staffing levels for differences in intensity of nursing care. As nursing care intensity increases, the number of nursing staff required to care properly for patients will increase (Unruh & Fottler 2006). Several researchers, such as Needleman *et al.* (2002) and Unruh and Fottler (2006), have attempted to integrate intensity of nursing care into their nurse staffing and patient outcomes research. Unruh and Fottler (2006) used patient turnover as a proxy measure of intensity of nursing care and hypothesized that as patient turnover increases, a similar increase in nursing care must be delivered in a shorter period of time during each patient stay. Reducing the length of stay presumably eliminates lower resource-use patient-care days but retains higher resource-use patient-care days (Unruh &

### What is already known about this topic

- Nursing shortages and the established link between nurse staffing levels and patient/nurse outcomes demand the appropriate work allocation of available nurses so that they are assigned to where the need for nurses is greatest.
- Although patient classification systems can be used to measure nursing workload, most are designed for specific unit types and only a few can be applied on a large scale.
- The Belgian Nursing Minimum Dataset contains information about nursing care intensity and nurse staffing levels that are uniformly measured in all acute care units in all Belgian hospitals.

### What this paper adds

- The variability in nurse staffing levels in Belgian acute hospitals occurs at the nursing unit level.
- A feedback tool that allows nationwide benchmarking of nurse staffing levels, corrected for the intensity of nursing care across a variety of disciplines.
- The feedback tool can be used by nurse managers and policymakers to guide nurse staffing decisions.

Fottler 2006). However, as described above, patient turnover remains a poor proxy measure of the intensity of nursing care. Needleman *et al.* (2002) used Nursing Intensity Weights, which are estimates of the relative level of nursing care needed by patients in each diagnosis-related group. However, the assumption that all patients with a similar medical diagnosis receive the same care is highly criticized by Welton *et al.* (2006). The nursing-intensity adjusted NHPPD calculated in this study can potentially solve this technical issue identified in the nurse staffing and patient outcomes literature because, since 2000, the B-NMDS and the Belgian Hospital Discharge Dataset (which contains data for calculating patient outcomes) were linked.

## Conclusions

Hospitals are increasingly challenged to balance ideal investments in care quality with current nurse labour constraints. In this context, it is of the utmost importance that available nurse staffing resources are allocated as accurately as possible, using the best evidence at hand. Despite the fact that our study model does not take into account many factors, given the data at hand, we succeeded in developing a benchmarking tool that enables hospital managers (at different organizational levels) to gain information about their nurse staffing allocation policy by comparing their data with those of others. Furthermore, the feedback tool can also be used by the Ministry of Public Health to assist in evaluating the current nurse staffing ratios for the different types of nursing units in acute hospitals. We hope that our findings will encourage researchers and policymakers in other countries to study and invest in systematic and uniform collection of data about the number of nurses involved in direct patient care and data about the intensity of nursing care.

## Acknowledgement

This research was commissioned by the Belgian Federal Science Policy Office, in collaboration with the Belgian Federal Ministry of Public Health, in the framework of the research programme Agora.

## Author contributions

KvdH, LD, EL, AV and WS were responsible for the study conception and design. KvdH, AV and WS performed the data collection. KvdH and LD performed the data analysis. KvdH was responsible for the drafting of the manuscript. KvdH, LD, EL, AV and WS made critical revisions to the paper for important intellectual content. LD, EL and WS provided statistical expertise. KvdH, AV and WS obtained funding. EL, AV and WS supervised the study.

## References

Adewale A.J., Hayduk L., Estabrooks C.A., Cummings G.G., Midodzi W.K. & Derksen L. (2007) Understanding hierarchical linear models: applications in nursing research. *Nursing Research* **56**(4 Suppl.), S40–S46.

Aiken L.H., Clarke S.P., Sloane D.M., Sochalski J. & Silber J.H. (2002a) Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association* **288**(16), 1987–1993.

Aiken L.H., Clarke S.P. & Sloane D.M. (2002b) Hospital staffing, organization, and quality of care: cross-national findings. *International Journal for Quality in Health Care* **14**(1), 5–13.

Aiken L.H., Clarke S.P., Cheung R.B., Sloane D.M. & Silber J.H. (2003) Educational levels of hospital nurses and surgical patient mortality. *Journal of the American Medical Association* **290**(12), 1617–1623.

Blegen M.A. (2006) Patient safety in hospital acute care units. *Annual Review of Nursing Research* **24**, 103–125.

Buerhaus P.I., Staiger D.O. & Auerbach D.I. (2000) Implications of an aging registered nurse workforce. *Journal of the American Medical Association* **283**(22), 2948–2954.

Cho S.H. (2003) Using multilevel analysis in patient and organizational outcomes research. *Nursing Research* **52**(1), 61–65.

Clarke S.P. (2003) Balancing staffing and safety. *Nursing Management* **34**(6), 44–48.

Closon M.C. (1991) *Financing Hospitals Based on Pathology Information*, Doctoral Thesis (in French). Ecole de Santé Publique UCL, Bruxelles.

De Groot H.A. (1989) Patient classification system evaluation. Part 1: Essential system elements. *Journal of Nursing Administration* **19**(6), 30–35.

Demeyere F. (2001) Nurse staffing allocation: an example (in Dutch). In *Report on Staff Allocation* (Verhaeghe R., ed.), Vlerick-Leuven-Gent Management School, Gent, pp. 1–2.

Fagerström L. & Rauhala A. (2007) Benchmarking in nursing care by the RAFAELA patient classification system – a possibility for nurse managers. *Journal of Nursing Management* **15**(7), 683–692.

Fagerström L., Rainio A.K., Rauhala A. & 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**(2), 97–104.

Giovannetti P. & Johnson J.M. (1990) A new generation patient classification system. *Journal of Nursing Administration* **20**(5), 33–40.

Graf C.M., Millar S., Feilteau C., Coakley P.J. & Erickson J.I. (2003) Patients' needs for nursing care: beyond staffing ratios. *Journal of Nursing Administration* **33**(2), 76–81.

Hox J. (2002) *Multilevel Analysis: Techniques and Applications*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.

Janiszewski G.H. (2003) The nursing shortage in the United States of America: an integrative review of the literature. *Journal of Advanced Nursing* **43**(4), 335–343.

Kane R.L., Shamliyan T., Mueller C., Duval S. & Wilt T. (2007) *Nursing Staffing and Quality of Patient Care. Evidence Report/Technology Assessment No. 151 (Prepared by the Minnesota Evidence-Based Practice Center Under Contract No. 290-02-0009) (Rep. No. AHRQ-Publication No. 07-E005)*. Agency for Healthcare Research and Quality, Rockville, MD.

Kirk H. (2007) Towards a global nursing workforce: the 'brain circulation'. *Nursing Management (Harrow)* **13**(10), 26–30.

Landon B.E., Normand S.L., Lessler A., O'Malley A.J., Schmaltz S., Loeb J.M. & McNeil B.J. (2006) Quality of care for the treatment of acute medical conditions in US hospitals. *Archives of Internal Medicine* **166**(22), 2511–2517.

Luke D.A. (2004) *Multilevel Modeling*. Sage, Thousand Oaks, CA.

McGillis-Hall L., Pink L., Lalonde M., Murphy G.T., O'Brien-Pallas L., Laschinger H.K., Tourangeau A., Besner J., White D., Tregunno D., Thomson D., Peterson J., Seto L. & Akeroyd J. (2006) Decision making for nurse staffing: Canadian perspectives. *Policy, Polititics & Nursing Practice* **7**(4), 261–269.

Milisen K., Abraham I., Siebens K., Darras E. & Dierckx de Casterlé B. (2006) Work environment and workforce problems: a cross-sectional questionnaire survey of hospital nurses in Belgium. *International Journal of Nursing Studies* **43**(6), 745–754.

Morris R., MacNeela P., Scott A., Treacy P. & Hyde A. (2007) Reconsidering the conceptualization of nursing workload: literature review. *Journal of advanced Nursing* **57**(5), 463–471.

Murphy L.N., Dunlap M.S., Williams M.A. & McAthie M. (1978) *Methods for Studying Nurse Staffing in a Patient Unit, a Manual to aid Hospitals in Making use of Personnel (Rep. No. HRA 78-3)*. DHEW, Hyattsville, MD.

Needleman J., Buerhaus P., Mattke S., Stewart M. & Zelevinsky K. (2002) Nurse staffing levels and the quality of care in hospitals. *New England Journal of Medicine* **346**(22), 1715–1722.

Park S. & Lake E.T. (2005) Multilevel modeling of a clustered continuous outcome: nurses' work hours and burnout. *Nursing Research* **54**(6), 406–413.

Plati C., Lemonidou C., Katostaras T., Mantas J. & Lanara V. (1998) Nursing manpower development and strategic planning in Greece. *Journal of Nursing Scholarship* **30**(4), 329–333.

Rafferty A.M., Clarke S.P., Coles J., Ball J., James P., McKee M. & Aiken L.H. (2007) Outcomes of variation in hospital nurse staffing in English hospitals: cross-sectional analysis of survey data and discharge records. *International Journal of Nursing Studies* **44**(2), 175–182.

Rauhala A., Kivimaki M., Fagerstrom L., Elovainio M., Virtanen M., Vahtera J., Raino A.K., Ojaniemi K. & Kinnunen J. (2007) What degree of work overload is likely to cause increased sickness absenteeism among nurses? Evidence from the RAFAELA patient classification system *Journal of Advanced Nursing* **57**(3), 286–295.

SAS Institute. (2001) *The SAS System for Windows*. SAS Institute, Cary, NC.

Seago J.A. (2002) A comparison of two patient classification instruments in an acute care hospital. *Journal of Nursing Administration* **32**(5), 243–249.

Sermeus W. (1992) *Variabiliteit van Verpleegkundige Verzorging in Algemene Ziekenhuizen*. Katholieke Universiteit Leuven, Leuven.

Sermeus W., Van den Heede K., Michiels D., Delesie L., Thonon O., Van Boven C., Codognotto J. & Gillet P. (2005) Revising the Belgian Nursing Minimum Dataset: from concept to implementation. *International Journal of Medical Informatics* **74**(11–12), 946–951.

Sermeus W., Delesie L., Van den Heede K., Diya L. & Lesaffre E. (2007a) Measuring the intensity of nursing care: making use of the Belgian Nursing Minimum Data Set. *International Journal of Nursing Studies*. Epub ahead of print.

Sermeus W., Gillet P., Tambeur W., Gillain D., Grietens J., Laport N., Michiels D., Thonon O., Vanden Boer G., Van Herck P., Swartenbroek N. & Ramaekers D. (2007b) *Financing Nursing Care in Hospitals (Rep. No. KCE Reports vol. 53 A)*. Federaal Kenniscentrum voor de gezondheidszorg (KCE), Brussels.

The Center for the Evaluative Clinical Sciences at Dartmouth (1999) Measures of variation and association. In *The Quality of Medical Care in the United States: a Report on the Medicare Program. The Dartmouth Atlas of Health Care* (Wennberg J.E. & Cooper M.M., eds), American Hospital Publishing, Chicago, IL, p. 286.

Tilquin C. (1987) *PRN 87: Measure of Patients' Required Nursing Care in a 24-Hour Period* (in French). EROS, Montréal.

Unruh L.Y. & Fottler M.D. (2006) Patient turnover and nursing staff adequacy. *Health Services Research*, **41**(2), 599–612.

Upenieks V.V., Akhavan J., Kotlerman J., Esser J. & Ngo M.J. (2007) Value-added care: a new way of assessing nursing staffing ratios and workload variability. *Journal of Nursing Administration* **37**(5), 243–252.

Welton J.M., Unruh L. & Halloran E.J. (2006) Nurse staffing, nursing intensity, staff mix, and direct nursing care costs across Massachusetts hospitals. *Journal of Nursing Administration* **36**(9), 416–425.

**Appendix 1** Nursing activities recorded for the Belgian Nursing Minimum Data Set with corresponding intensity-of-nursing-care weights

| Nursing activity | Response categories | Weight 2 | Weight 3 |
|---|---|---|---|
| 1. Care relating to hygiene | a. No assistance | 0 | 1 |
| | b. Supportive assistance | 2 | 3 |
| | c. Partial assistance | 4 | 8 |
| | d. Complete assistance | 7 | 13 |
| 2. Care relating to mobility | a. No assistance | 0 | 0·5 |
| | b. Supportive assistance | 1 | 2 |
| | c. Partial assistance | 4 | 4 |
| | d. Complete assistance | 6 | 9 |
| 3. Care relating to elimination | a. No assistance | 0 | 1 |
| | b. Supportive assistance | 2 | 4 |
| | c. Partial assistance | 5 | 7·5 |
| | d. Complete assistance | 9 | 15 |
| 4. Care relating to feeding | a. No assistance | 1 | 4 |
| | b. Supportive assistance | 3 | 6 |
| | c. Partial assistance | 5 | 9 |
| | d. Complete assistance | 15 | 25 |
| 5. Tube feeding | a. No | 0 | 0 |
| | b. Yes | 8 | 11·5 |
| 6. Mouth care | Frequency/24 hours | 2 × frequency | 1 × frequency |
| 7. Prevention of pressure sores: changing position | Frequency/24 hours | 2 × frequency | 4 × frequency |
| 8. Assistance in getting dressed | a. No | 0 | 0 |
| | b. Yes | 4 | 7 |
| 9. Care of patient with tracheotomy or endotracheal tube | a. No | 0 | 0 |
| | b. Present without artificial ventilation | 5 | 7 |
| | c. Present with artificial ventilation | 15 | 23 |
| 10. Nursing admission assessment | a. No | 0 | 0 |
| | b. Yes | 3 | 8·5 |
| 11. Training in activities of daily living | a. No | 0 | 0 |
| | b. Occasional | 3 | 10 |
| | c. According to programme | 5 | 12·5 |
| 12. Emotional support | a. No | 0 | 0 |
| | b. Yes | 5 | 8·3 |
| 13. Care of a disoriented patient | a. No | 0 | 0 |
| | b. Passive | 3 | 5 |
| | c. Reality orientation training | 7 | 10 |
| 14. Isolation to prevent contamination | a. No | 0 | 0 |
| | b. Yes | 7 | 14·5 |
| 15. Monitoring vital signs | Highest frequency/24 hours | 1 × frequency | 1 × frequency |
| 16. Monitoring clinical signs | Highest frequency/24 hours | 1 × frequency | 1 × frequency |
| 17. Cast care | a. No | 0 | 0 |
| | b. Yes | 4 | 1·5 |
| 18. Taking blood samples | Frequency/24 hours | 1 × frequency | 2·5 × frequency |
| 19. Medication management (intramuscular, subcutaneous) | Number of doses/24 hours | 2 × frequency | 1 × frequency |
| 20. Medication management (intravenous) | Number of doses/24 hours | 2 × frequency | 2 × frequency |
| 21. Infusion therapy | Number of intravenous lines | 4 × i.v.-lines | 6 × i.v -lines |
| 22. Surgical wound care | Number of interventions/24 hours | 3 × frequency | 5 × frequency |
| 23a. Traumatic wound care (surface) | a. <20% of body surface | 3 | 2 |
| | b. 21–45% of body surface | 6 | 4 |
| | c. 46–70% of body surface | 8 | 6 |
| | d. >70% of body surface | 10 | 8 |
| 23b. Traumatic wound care (interventions) | number of interventions/24 hours | Score 23a × frequency | 9 × frequency |

*Weight 1:* First component that results from non-linear principal component analysis of the 23 nursing activities using Proc PRINQUAL of SAS (Sermeus *et al.* 2007a).

*Hours of nursing care staffed per nursing unit*

Working hours of Registered Nurses having at least a Bachelor's degree

Working hours of other Registered Nurses

Weight 2 (Closon 1991); weight 3 (Demeyere 2001).

## Appendix 2 Statistical model details

The intraclass correlation coefficient (ICC) is the fraction of the total variability that is due to the group level. ICC represents to what degree units belonging to the same group resemble (correlate to) each other; that is, ICC represents the correlation between two randomly chosen units in one randomly drawn group.

$$ICC_k = \frac{\sigma_k^2}{\sum_{i=1}^{m} \sigma_i^2} \quad (2a)$$

In equation (2a), $ICC_k$ represents the intraclass correlation for the $k$th level; $\sigma_k^2$ is the variance of the $k$th level; and $\Sigma_{i=1}^{m}\sigma_i^2$ is the total variability for all the $m$ levels.

With regard to the correlation between two individuals in the same group, part of the correlation is explained by its covariates, whereas another part is unexplained. The unexplained correlation is called the residual intraclass correlation coefficient (RICC). RICC is calculated using equation (2a) but after controlling for the covariates.

The multilevel model accounts for the variability attributed to the covariates: 'Lweight1' is the logarithmic function of the intensity of care; 'Wk' indicates the difference between weekdays (0) and weekend days (1); 'htype' indicates the hospital type (academic = 1, general = 0); and 'serv' indicates the service type (0 = general units, 1 = intensive care units). The multilevel model also accounts for some unexplained variability emanating from the nursing units within hospitals and from the hospitals themselves by applying a random intercept model:

$$Y_{ijk} = \mu_{ijk} + \varepsilon_{ijk} \quad (2b)$$

$$\begin{aligned}\mu_{ijk} = {} & \beta_0 + \alpha_i + \eta_{ij} + \beta_1 \text{Lweight1}_{ijk} \\ & + \beta_2 serv_{ij} + \beta_3 wk_{ijk} + \beta_4 htype_i + \beta_5 \omega'_{ijk}\end{aligned} \quad (2c)$$

In equation (2b), $Y_{ijk}$ is the log of the NHPPD for the $k$th day of the $j$th unit in the $i$th hospital, and $\alpha_i$ and $\eta_{ij}$ are random effects pertaining to the $i$th hospital and the $j$th nursing unit, respectively, within the $i$th hospital. $\beta_1$, $\beta_2$, $\beta_3$ and $\beta_4$ are the fixed effect coefficients for $\text{Lweight1}_{ijk}$, $serv_{ij}$, $wk_{ijk}$ and $htype_i$, respectively. $\beta_0$ is the fixed effect intercept coefficient. The vector $\omega_{ijk}$ represents the set of all possible interaction terms, and $\beta_5$ represents their corresponding fixed effects coefficients. $\varepsilon_{ijk}$ is the measurement error pertaining to the observation day level.

To obtain the population average model, we average the random intercepts of the nursing units within hospitals and hospitals by integration:

$$\begin{aligned}E(NHPPD_{ijk}) &= E_{\alpha_i}(E_{\eta_{ij}}(NHPPD|\hat{\alpha}_i\hat{\eta}_{ij})) \\ &= \int\int \exp(\hat{\mu}_{ijk})d\hat{\alpha}_i d\hat{\eta}_{ij}\end{aligned} \quad (2d)$$

where $\hat{\mu}_{ijk}$ is given by equation (2c) whereby all parameters are replaced by their estimates. $\alpha_i$ and $\eta_{ij}$ are replaced by their Empirical Bayes estimates; that is, $\hat{\alpha}_i$ and $\hat{\eta}_{ij}$ respectively. The integration (equation 2d) is realized through the sampling approach, by averaging over 1000 draws of $\hat{\alpha}_i$ and $\hat{\eta}_{ij}$ for each covariate combination.