# A Predictive Model of Nurse Workload from Routine Hospital Data

## Abstract

**Background:** Managing nurse staffing is complex due to fluctuating demand based on ward occupancy, patient acuity and dependency. Monitoring staffing adequacy in real-time has the potential to inform safe and efficient deployment of staff. Patient classification systems are being used for per shift workload measurement, but they add a frequent administrative task for ward nursing staff.

**Objective:** To explore whether an algorithm could estimate ward workload using existing routinely recorded data.

**Methods:** Anonymised admission records and assessments from a patient classification system (PCS) supporting the Safer Nursing Care Tool (SNCT) were used to determine nursing care demand in medical and surgical wards in a single UK hospital between Feb 2017 and Feb 2020. Records were linked by ward and time. The data was split into a training set (75%) and a test set (25%). We built a predictive model of ward workload (as measured by the PCS) using routinely recorded administrative data and admission National Early Warning Score (NEWS). The outcome variable was ward workload derived from the patient classifications, measured as the number of whole-time equivalent (WTE) nursing staff per patient.

**Results:** In a test set of 11,592 ward assessments from 42 wards with a mean WTE per patient of 1.64, the model’s mean absolute error was 0.078, a mean percentage error of 4.9%. A Bland-Altman plot of the differences between the predicted values and the assessment values showed 95% of them within 0.21 WTE per patient.

**Conclusion:** Predictions of nursing workload from a relatively small number of routinely collected variables showed moderate accuracy for general wards in one English hospital. This demonstrates the potential for automating assessments of nurse staffing requirements from routine data, reducing time spent on this non-clinical overhead and improving monitoring of real-time staffing pressures.

**Keywords:**

Workload; Staffing; Nursing staff; Safer Nursing Care Tool

## Introduction

Nursing workload management is complex and involves the processes of forecasting, scheduling, staffing and monitoring [1]. Monitoring a ward’s nursing workload in real-time or near-time is difficult because of patient movements and because of changes in patient acuity and care needs. This is important because there is a body of evidence that shows nurse understaffing is associated with adverse outcomes for patients and for staff [2,3]. Monitoring supports the identification of staffing deficiencies and informs decisions concerning the deployment or redeployment of resources to address them.

Common structured methods of monitoring staffing requirements make use of patient classification systems. Through these systems, care intensity is determined for each patient and is then summed to calculate the total care demand for the ward. Examples are the acuity/dependency assessments in the Safer Nursing Care Tool (SNCT) [4], the Oulu Patient Classification qualisan (OPCq) used in RAFAELA [5] and the Perroca Patient Classification System [6]. Assessing and documenting the nursing intensity for every patient on the ward is an additional task for ward nurses, often repeated every shift. Automating the nursing intensity and workload monitoring using data already recorded has the potential to free up valuable staff time in the provision of patient care.

There is little published work on automating patient classification systems used to measure nursing intensity and workload. However a model was developed to determine each patient’s workload category as defined by the Perroca Patient Classification Instrument (PPCI) from routine electronic health record data in a Brazilian hospital [7]. A random forests classification model used electronic variables chosen as indicators for each of the nine care areas in the tool. The classifier was 72% accurate when categorising the patients overall with a c-statistic of 0.82. An individual patient misclassification may not have much effect on the ward workload estimation as a whole but there were no details of how to translate the patient workload categories into nurse staffing requirements for wards.

Our aim was to build a predictive model using routine electronic data from one hospital which might be known in real-time to estimate nurse staffing requirements for wards. Our objective was to match the recorded values determined by using the patient classification instrument from the SNCT, the most widely used tool in England for determining staffing requirements based on patient need [8]. The SNCT originated as a way of estimating nurse staffing establishment requirements (whole-time equivalent posts) by means of twice-a-year audits of ward workload. It included a patient classification system which is now in use once, twice or three times a day in many hospitals in England to monitor ward workload.

## Methods

### Ethics

This research was a secondary analysis of data collected for a study which considered the costs and consequences of different nursing staff configurations (Trial Registration: ClinicalTrials.gov NCT04374812)[9]. The ethical review undertaken by the faculty research ethics committee at the University of Southampton (ERGO 52957) for that study was extended for this secondary analysis. Data set construction and all analyses were performed using R version 4.4.0 (R Core Team, Vienna, Austria); further details of the software and packages used can be found in the supplementary material.

### Data Sources and Linkage

We analysed data from adult inpatient wards in one hospital trust for admissions from February 2017 to February 2020. The data consisted of anonymised admission records and patient classification assessments made using the SNCT acuity instrument. SNCT acuity assessments had been undertaken in admission units, general wards, high care wards but not in intensive care units. The admission units in this hospital were the medical emergency and surgical emergency assessment units which assess patients before transferring them to a general or high ward as required. In the UK, a general ward is an inpatient adult ward not specialising in maternity or psychiatric care. A high-care ward has some high-care beds which are defined as providing level 2 care according to the UK intensive care society, deemed as needing 0.5 WTE of nursing care per bed. A UK intensive care unit on the other hand provides the highest level of care, level 3, and this requires one-to-one nursing care. The SNCT acuity assessments data provided counts of patients in a ward for each acuity/dependency category at a point in time. The admission records included each patient’s first national early warning score (NEWS) [10]. NEWS is a patient assessment method in routine use in UK hospitals based on vital signs observations which nurses use regularly to identify patients at risk of deterioration.

The patient categorisation instrument used for ward workload assessments in SNCT involves matching patients to one of five ordinal categories according to descriptions of acuity and dependency needs. Level 0 is for stable patients with needs met in general wards, level 1a patients are acutely ill patients requiring intervention or those with a greater potential to deteriorate, level 1b patients are stable but are more dependent on nursing care, and level 2 patients are unstable, at risk of deteriorating and require specialised care. Level 3 patients, who require multiple organ support, did not exist in our dataset of assessments. Each category has a value (called a multiplier) corresponding to a staffing whole-time equivalent (WTE) needed to provide 24-hour care after considering annual leave and sickness absence. Multipliers were developed from ward observational studies [11]. One set of multipliers is used for general wards and another for admission units to take into account the extra workload associated with the higher throughput of patients. The multiplier values for the patients in a ward are summed to produce a point-in-time WTE estimate for the ward [4].

All admission records that could be linked by ward and by assessment date and time were included for wards that were routinely carrying out the assessments. Since the assessment records did not record any patient identifiers, linkage to patient records was by ward and time. Linked records were analysed where the number of patients in the ward assessment matched the number of occupants in the ward according to the patient administration system at the time of the assessment. A comparison of linked and unlinked assessments was conducted to check for bias resulting from this selection process.

### Study Variables

The dependent variable for the prediction model was the whole-time equivalent (WTE) calculated by the assessment tool divided by the number of assessed patients. Patients’ data for the assessed patients as determined by the linkage described above was aggregated to ward level by taking means for continuous variables and proportions for binary variables as shown in the table. The choice of candidate predictors (Table 1) was made according to data availability, published nurse workload models and clinical protocols regarding nurse assessments for new ward arrivals. Multi-variable models using all the candidate predictors were built since there was evidence or conceptual reasons for including them and the objective of the modelling was predictive accuracy rather than quantifying effects of individual predictors. Univariable regressions were performed on the candidate predictors out of interest. The predictors include the use of two derived variables: the Charlson comorbidity index [12] and the summary hospital mortality indicator (SHMI) risk [13]. The diagnostic predictors were proportions of patients with a recorded diagnosis belonging to a particular diagnostic group: respiratory conditions, cardiac conditions, renal conditions, cancer, frailty syndromes [14], a self-harm diagnosis and a history of poisoning. The appendix details the diagnostic groups. Lastly, an indicator was added for admission units to reflect the alternate set of patient category multipliers used for assessments in them.

### Analysis

A generalised additive regression model (GAM) [15] with an identity link function was chosen for the modelling since this could handle non-linear relationships between the individual predictors and the WTE per patient response variable.

The data was randomly split into training and test sets (75% vs. 25%) and a regression model created on the training set. All the predetermined candidate predictors were used after checking for high pairwise correlations. Removal of any predictor was found to worsen model fit judging by the Akaike Information Criterion (AIC), so all were retained. A calibration was computed to optimise the predictions by means of a smoothing of the model estimates on the assessed ones [16]. The model and the calibration were applied to the test set.

A Bland-Altman plot [17] was constructed and used to look at the agreement between the model prediction and the assessed WTE per patient. When examining this plot by ward subset, two wards had more than a third of their predictions outside the limits of agreement. This was a markedly higher proportion of assessments than for the other wards so these wards were excluded from the datasets and the GAM model was rebuilt.

Table 1: Model predictors

|  |  |  |
| --- | --- | --- |
| **Category** | **Predictor** | **Ward Aggregation** |
| Demographic |  |   |
|   | Age (md-point of 5 yr age group) | Mean |
|   | Male gender |   |
|   |  |   |
| Clinical |  |   |
|   | Admission NEWS | Mean |
|   |  |   |
| Pathway |  |   |
|   | Elective admission | Proportion |
|   | Prior length of stay | Median |
|   | Number of ward transfers | Mean |
|   | Number of consultant transfers | Mean |
|   | From admission unit (same day transfer) | Proportion |
|   | From theatres (same day transfer) | Proportion |
|   | From high or intensive care (same day transfer) | Proportion |
|   | New hospital admisson (same day, no prior ward) | Proportion |
| Diagnostic |  |   |
|   | Renal condition | Proportion |
|   | Cardiac condition | Proportion |
|   | Cancer condition | Proportion |
|   | Frailty syndrome: dementia or delirium | Proportion |
|   | Frailty syndrome: dependence or care | Proportion |
|   | Frailty syndrome: incontinence | Proportion |
|   | Frailty syndrome: falls or fractures | Proportion |
|   | Frailty syndrome: pressure ulcers or weight loss | Proportion |
|   | Frailty syndrome: anxiety or depression | Proportion |
|   | Frailty syndrome: mobility problems | Proportion |
|   | Respiratory condition | Proportion |
|   | History of Poisoning | Proportion |
|   | Self-harm diagnosis | Proportion |
|   |  |   |
| Organisational |  |   |
|   | Admission Unit | Binary |
|   |  |   |
| Derived |  |   |
|   | SHMI Mortality Risk | Mean |
|   | Charlson Comorbidity Index | Mean |

## Results

The data came from 44 wards including 8 admission units and 4 high care wards, covering 1,481,801 patient categorisations by care level from 71,027 ward-level assessments relating to 125,595 admissions. Most patient assessments in general wards (59.3%) were recorded as level 1b (Table 2).

Table 2: % of Assessed Patient Categorisations by care level

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Patient Categorisations** | **Level 0** | **Level 1a** | **Level 1b** | **Level 2** |
| General or High Care Ward | 1364690 | 17.00% | 16.20% | 59.30% | 7.50% |
| Admission Unit  | 117111 | 25.00% | 38.00% | 36.50% | 0.50% |

The distribution of WTE per patient for all assessments (median 1.63, IQR 1.48 to 1.72) shows a distinct peak near the value of the multiplier for level 1b (Figure 1).

Figure 1: Distribution of WTE per patient values from SNCT ward assessments

Requiring the number of ward occupancy records to match the number of patients assessed reduced the data set by 32% to 48,493 assessments relating to 108,558 admissions, however the WTE per patient distribution was little changed (median 1.65, IQR 1.51 to 1.73) suggesting no systematic bias. The ward occupancy and therefore the number of patients per assessment in the matched dataset had a median of 20 with an IQR of 11 to 29.

A patient might be assessed on a number of occasions in a number of wards during their admission. The cohort of assessed patients consisted of 86% emergency admissions, 53% female with most coming from the 80-84 age bracket (Table 3). Long term conditions were common with 53% having a Charlson comorbidity index of more than five. The median length of stay was 4.4 days and 4.9% died in hospital.

Table 3: Admission descriptives for modelled dataset

|  |  |  |
| --- | --- | --- |
| **Total N (%)a** |   | 108558 (100.0) |
| **Gender** | Male | 50877 (46.9) |
| **Emergency admission** |  | 92766 (85.5) |
| **Age Grp** | 10-14 | 1 (0.0) |
|   | 15-19 | 1105 (1.0) |
|   | 20-24 | 2765 (2.5) |
|   | 25-29 | 2946 (2.7) |
|   | 30-34 | 3274 (3.0) |
|   | 35-39 | 3275 (3.0) |
|   | 40-44 | 3212 (3.0) |
|   | 45-49 | 4525 (4.2) |
|   | 50-54 | 5775 (5.3) |
|   | 55-59 | 7031 (6.5) |
|   | 60-64 | 7405 (6.8) |
|   | 65-69 | 8663 (8.0) |
|   | 70-74 | 11997 (11.1) |
|   | 75-79 | 11830 (10.9) |
|   | 80-84 | 12886 (11.9) |
|   | 85-89 | 11973 (11.0) |
|   | 90-120 | 9895 (9.1) |
| **Charlson Index** | [0] | 32933 (30.3) |
|  | [1-5] | 17734 (16.3) |
|  | [>5] | 57635 (53.1) |
|  | (Missing) | 256 (0.2) |
| **% SHMI riskb** | Mean (SD) | 7.0 (10.0) |
| **NEWS2c** | Zero 0 | 60179 (55.4) |
|  | Low 1-4 | 36789 (33.9) |
|  | Med 5-6 | 6708 (6.2) |
|  | High 7-20 | 3343 (3.1) |
|  | (Missing) | 1539 (1.4) |
| **Length of Stay** | Median (IQR) | 4.4 (2.0 to 10.5) |
| **Hospital death** |   | 5332 (4.9) |

1. values are number and % unless otherwise stated
2. Summary Hospital Mortality Indicator
3. National Early Warning Score (version 2)

We fitted a GAM to the training set to model the nursing workload measured in WTE per patient associated with the set of patients in each manual ward assessment. When applied to the test set of 11,592 ward assessments from 42 wards and a calibration applied, the mean absolute error of the predictions was 0.078 WTE per patient (a mean absolute percentage error of 4.9% with a 95th centile of 13.9%). The root mean square error was 0.106 WTE per patient, and the percentage of variance explained (PVE) by the model was 65%. The predictors which contributed most to explaining the workload (Figure 2) were: the admission unit indicator, mean patient age, mean Charlson comorbidity index, the proportion of men, the mean NEWS and the proportion of patients with a diagnosed renal condition. The relative contribution of each predictor was computed by calculating the root mean square error if removed from the model.

Figure 2: Relative contribution each predictor makes to the model fit

Univariable regressions of each of the predictors showed them all to be significant, apart from transfer from an admission unit on the day of assessment (Table 4 in the appendix). All predictors were retained in the GAM model.

A calibration plot of predictions in the test set against the recorded WTE per patient with the addition of a loess smoothing line forming a calibration curve [18] (Figure 3) shows that there was some underestimation of workload at high values.

Figure 3: Calibration of predictions in the test set

A Bland-Altman plot of differences between the recorded WTE per patient and model predictions in the test set (Figure 4) had a mean difference of 0.002 WTE per patient. The limits of agreement are at +/- 0.21 WTE per patient, meaning 95% of assessments fall in this range.

Figure 4: Bland-Altman plot of test model assessments and predictions (wpp: WTE per patient; pred: model prediction)

## Discussion

### Main Findings

In this study we used a limited set of routine patient data to estimate the demand for nursing staff on general and high-care hospital wards and admission units. We found that data on patient demographics, routes of admission, ward transfers, diagnoses and admission NEWS which are available in hospital systems could be used to estimate the results of assessments made using a patient classification system with a moderate degree of precision. The predictions of WTE per patient were on average within 5% of the recorded estimates when applied to a test set. Calibration ”in the large” was good in the test set with the mean prediction being only 0.002 WTE per patient less than the recorded mean [18]. However, a calibration curve showed that the model tended to underestimate at higher workloads. This would be the case if some dimensions of patients’ care needs in more acute wards were not being adequately represented in the data set. A “by ward” sub analysis of the model estimates showed that the surgical high care unit and the renal transplant unit were the two wards where the model underestimated most (as proportions of predictions). This might indicate the model is lacking variables which adequately capture the acuity and dependency of patients after major surgery. If the model were to be used as it is, a professional judgement framework such as that published by Saville et al [19], should be employed which might recommend considering adjustment of the estimates where they exceeded 1.75 WTE or where there were high care beds in the ward.

The predictors were not chosen to be orthogonal, they clearly have correlations when considering their potential effect on nurse workload. That means figure 2 explaining their contribution to estimating workload needs to be interpreted with care. It’s not surprising that the admission unit indicator makes the largest contribution since an approximately 10% uplift in nurse staffing for admission units is baked into the PCS which determines the target WTE per patient. It is perhaps expected that the age, comorbidity index and admission NEWS case-mix variables of ward occupancy influence estimates of workload. However, the marked contribution of gender invites comment. It may reflect staffing in single-sex wards dealing with gynaecological or urological conditions.

While overall fit was good the Bland Altman plot shown in Figure 3 revealed a distinct artefact of 45-degree downward sloping lines which arises from ranges of continuous prediction values estimating a given workload derived from the discrete set of multiplier values in the PCS and the numbers of patients in wards. There is a degree of subjectivity in categorising patients in the PCS into one of the four levels and there is a possibility that there was some up-coding of assessments in which patients were put into a higher category, for example level 1b rather than level zero, which might boost a ward’s chances of getting more resources [20]. The high proportion of patient assessments in the level 1b category shown in Table 2 possibly suggests some of these were erroneous although patients with dependent care needs may have longer than average stays with opportunity for more categorisations at that level.

In conversations, safe staffing managers and leads at a number of hospitals have expressed concerns regarding the accuracy and reliability of manual patient categorisations by ward staff. An automated workload prediction system would be objective and eliminate or reduce a source of bias. It would also remove the need for training staff in carrying out the assessments and re-training them when assessment methods are revised.

The OPCq and Perroca patient classification instruments for assessing patient care needs are more structured than the SNCT instrument as they require scoring a patient across several domains before determining an overall category. The drawback to this rigour is that the assessments take longer. One nurse quoted in the publication by Ayan (2024, p. 7) [21] concerning the Perroca instrument said:

Since we have been using it continuously, it does not take me much time now. It only takes five minutes to assess a patient; it couldn’t be easier.

Our observational data (unpublished) from another study [22] suggests that assessing all patients on a thirty-bedded ward three times per day with SNCT categorisations could occupy a senior nurse for up to 30 minutes per day, plus extra time for data entry.

The identification of two wards with outlier predictions is of interest. In the private patient unit the model tended to predict higher staffing requirements than those identified in the assessments. This could arise because of a selection effect, whereby the occurrence of more elective procedures on more medically fit patients was not reflected in the variables used for this population. In contrast, the predicted workload estimates for the surgical high care unit tended to be too low and this might mean the model is not adjusting adequately for these acutely ill patients. Additional predictors to adjust for operational severity might improve the predictions.

The surprising accuracy of the predictions on limited data opens up the possibility of automated near-time assessments and consequently freeing up some nursing time. The feasibility of implementing the prediction algorithm in real-time or near-time is dependent on the data items being recorded and accessible in a timely manner. In the UK, basic patient admission data is usually recorded in near-time in a patient administration system with electronic interfacing, though the clinical coding of diagnoses in the same system is often much later. However, the diagnoses the algorithm employs are largely for long-term conditions or conditions known on presentation in which case they may already be present in the electronic clinical history or recorded in nursing or clinical handover documentation. A National Health Service (NHS) England-sponsored digital maturity assessment of UK hospital trusts reports that 32 of 132 acute trusts have electronic nursing and physician documentation [23] and that number is rising so real-time electronic provision of patient conditions is becoming more widespread.

### Comparison with Prior Work

The approach taken in this study differs from previous studies which built models or made estimates of actual historic staffing levels [24–26]. In contrast, our approach was to build a model of estimated staffing requirements trained on the records of an existing staffing estimation tool [27]. Unlike the classifier model built by Rosa [7] which was trained on values from the Perroca Patient Classification Instrument (PPCI) to make patient categorisations, ours was a prediction model of ward staffing requirements. Whilst prediction of PPCI values was only moderately accurate (72% correct classification), in our study the magnitude of errors for estimating ward staffing requirements was, on average, small with a mean average percentage error of less than 5% and 95% of estimates within 14% of the assessed value. The required precision is an open question but staffing within +/-15% of measured demand is used as a criterion to define acceptable variation in one widely used system [28].

The method we have used to build a model which is trained on the ward workload is flexible and would work for other patient acuity and dependency scoring instruments which assign an ordinal value for the patient’s care needs.

### Limitations

The dataset was limited in terms of longitudinal clinical data regarding patient conditions. Whilst it contained the NEWS on admission, it didn’t have subsequent scores or individual vital signs. Without time-varying variables such as NEWS and other assessments which reflect a patients’ evolving acuity the model is relying on the ward case-mix not varying from the average in terms of proportions of patients medically fit for discharge or proportions deteriorating. These two cases work in opposite directions in terms of workload, those patients who are medically fit probably requiring less nursing care whilst those who deteriorate probably requiring more nursing care. A systematic bias in the model performance is not easy to determine without further information. Also, as noted above in the main findings, the poor performance of the model in the private patient unit and in surgical high care might result from the model lacking information on surgical procedures and operation severity. This limits the settings in which the current model might be used.

Another limitation which might have impacted the accuracy of the predictions was the reliability of the recorded patient classifications. The use of the acuity categorisation tool for estimating staffing requirements in the dataset was new to the hospital at the time, hence reliability of the assessment may have depended for a period on the effectiveness of training given to ward assessors and monitoring of the use of the tool.

The robustness and generalisability of the model needs to be checked by assessing its performance using data from other hospitals. It’s possible that variation in patient case-mix and in how hospitals apply SNCT patient acuity categorisations could affect model performance. It should also be noted that whilst the aim of this study was to estimate workload which matched the existing SNCT-based assessments, the tool in use did not refine workload requirements by skill-mix, for example advising how many registered nurses and how many nurse support staff might meet the requirements, nor about requirements for specialist skills. This would be a useful step in matching requirements to available resources and might be the subject of further work.

### Conclusions

This research demonstrated a way of building a moderately accurate prediction model of nurse staffing requirements from a limited dataset of routine data by training the model on records of an established patient acuity and dependency patient classification tool. The results demonstrate the potential for automating assessments of nurse staffing requirements from routine data, reducing time spent on this non-clinical overhead and improving monitoring of real-time staffing pressures.

## Acknowledgements

This study was funded by the National Institute of Health Research (NIHR) Applied Research Collaboration Wessex. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

## Conflicts of Interest

None declared.

## Abbreviations

AIC: Akaike information criterion

GAM: generalised additive regression model

NIHR: National Institute of Health Research

NEWS: National Early Warning Score

OPCq: Oulu Patient Classification qualisan

PCS: patient classification system

PPCI: Perroca patient classification instrument

SHMI: Summary Hospital Mortality Indicator

SNCT: Safer Nursing Care Tool

WTE: whole-time equivalent

## Data Availability

The nature of the data (individual patient data) and the data sharing agreements with the data providers means we are unable to share data.

## References

1. American Nurses Association. Defining staffing:  workforce management patient classification and acuity  systems. The request for proposal process. 2017. Accessed September 3, 2024. https://www.nursingworld.org/practice-policy/nurse-staffing/workforce-management-pcas-and-the-rfp-process/

2. Shin S, Park JH, Bae SH. Nurse staffing and nurse outcomes: A systematic review and meta-analysis. Nurs Outlook. 2018;66(3):273-282. PMID:29685321

3. Dall’Ora C, Saville C, Rubbo B, Turner L, Jones J, Griffiths P. Nurse staffing levels and patient outcomes: A systematic review of longitudinal studies. Int J Nurs Stud. 2022;134:104311. PMID:35780608

4. Shelford TG. Safer Nursing Care Tool. Published online 2013:1-20.

5. Fagerström L, Lønning K, Helen M. The RAFAELA system: a workforce planning tool for nurse staffing and human resource management. Nurs Manag (Harrow). 2014;21(2):30-36. PMID:24779764

6. Perroca MG. The new version of a patient classification instrument: assessment of psychometric properties. J Adv Nurs. 2013;69(8):1862-1868. PMID:23157307

7. Rosa NG da, Vaz TA, Lucena A de F. Nursing workload: use of artificial intelligence to develop a classifier model. Rev Lat Am Enfermagem. 2024;32. http://www.scielo.br/scielo.php?script=sci\_arttext&pid=S0104-11692024000100335&tlng=en

8. Griffiths P, Saville C, Ball JE, et al. The Safer Nursing Care Tool as a guide to nurse staffing requirements on hospital wards: observational and modelling study. Health Services and Delivery Research. 2020;8(16):1-162.

9. NIHR. Consequences, costs and cost-effectiveness of different workforce configurations in English acute hospitals: a longitudinal retrospective study using routinely collected data - NIHR Funding and Awards. 2020. Accessed September 13, 2024. https://www.fundingawards.nihr.ac.uk/award/NIHR128056

10. RCP. Royal College of Physicians: National Early Warning Score (NEWS) 2. 2022. Accessed August 17, 2023. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2

11. Smith S, Casey A, Hurst K, Fenton K, Scholefield H. Developing, testing and applying instruments for measuring rising dependency-acuity’s impact on ward staffing and quality. Int J Health Care Qual Assur. 2009;22(1):30-39. PMID:19284169

12. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol. 2004;57(12):1288-1294. PMID:15617955

13. Campbell MJ, Jacques RM, Fotheringham J, Maheswaran R, Nicholl J. Developing a summary hospital mortality index: retrospective analysis in English hospitals over five years. BMJ. 2012;344(7848). PMID:22381521

14. Soong JTY, Kaubryte J, Liew D, et al. Dr Foster global frailty score: an international retrospective observational study developing and validating a risk prediction model for hospitalised older persons from administrative data sets. BMJ Open. 2019;9(6):e026759. PMID:31230009

15. Wood SN, Pya N, Säfken B. Smoothing Parameter and Model Selection for General Smooth Models. J Am Stat Assoc. 2016;111(516):1548-1563. Accessed September 15, 2024. https://www.tandfonline.com/doi/abs/10.1080/01621459.2016.1180986

16. Ojeda FM, Jansen ML, Thiéry A, et al. Calibrating machine learning approaches for probability estimation: A comprehensive comparison. Stat Med. 2023;42(29):5451-5478. PMID:37849356

17. Bland MJ, Altman DG. STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. The Lancet. 1986;327(8476):307-310. PMID:2868172

18. Van Calster B, McLernon DJ, Van Smeden M, et al. Calibration: The Achilles heel of predictive analytics. BMC Med. 2019;17(1):1-7. PMID:31842878

19. Saville C, Griffiths P, Casey A, et al. Professional judgement framework: a guide to applying professional judgement in nurse staffing reviews - ePrints Soton. November 4, 2022. Accessed June 3, 2025. https://eprints.soton.ac.uk/478317/

20. Hurst K. Gaming and up-coding. Nurs Manag (Harrow). 2009;15(9):19-23. PMID:19248324

21. Ayan G, Türkmen E, Uğur E, et al. Determination of Nursing Care Times Based on the Perroca Patient Classification Instrument in the Inpatient Oncology Unit: A Mixed Method Study. Semin Oncol Nurs. 2024;40(2):151608. PMID:38402019

22. Griffiths P, Saville C, Ball J, Culliford D, Pattison N, Monks T. Performance of the Safer Nursing Care Tool to measure nurse staffing requirements in acute hospitals: a multicentre observational study. BMJ Open. 2020;10(5):e035828. PMID:32414828

23. NHS England. Digital Health Intelligence Digital maturity of NHS acute trusts. July 2023. Accessed November 10, 2024. https://digitalhealthintelligence.net/digital-maturity-acute-nhs-snapshot-report/

24. Meyer KR, Fraser PB, Emeny RT. Development of a Nursing Assignment Tool Using Workload Acuity Scores. J Nurs Adm. 2020;50(6):322. PMID:32427663

25. Musy SN, Endrich O, Leichtle AB, Griffiths P, Nakas CT, Simon M. The association between nurse staffing and inpatient mortality: A shift-level retrospective longitudinal study. Int J Nurs Stud. 2021;120:103950. Accessed October 11, 2023. https://linkinghub.elsevier.com/retrieve/pii/S0020748921000936

26. Ostberg N, Ling J, Winter SG, et al. Quantifying paediatric intensive care unit staffing levels at a paediatric academic medical centre: A mixed-methods approach. J Nurs Manag. 2021;29(7):2278-2287. PMID:33894027

27. Smith J, Forde V, Goodman M, Cannaby AM, Radford M. How to keep score of acuity and dependency. Nurs Manag (Harrow). 2009;16(8). PMID:20050478

28. Rauhala A, Fagerström L. Determining optimal nursing intensity: the RAFAELA method. J Adv Nurs. 2004;45(4):351-359. PMID:14756829

29. HCUP-US Tools & Software Page. Accessed November 10, 2024. https://hcup-us.ahrq.gov/toolssoftware/ccs/ccsfactsheet.jsp

## Appendix: Diagnosis Groupings

### Frailty Syndromes

The 7 Frailty Syndromes were defined in the paper by Soong [14]: Dementia and Delirium, Mobility Problems, Falls and Fractures, Pressure Ulcers and Weight Loss, Incontinence, Dependence and Care, Anxiety and Depression

### Respiratory conditions

ICD-10 Codes: J41-J44, J60-J68

### History of Poisoning or Self-harm

ICD-10 Codes: Z915

### Self-harm

ICD-10 Codes: X60-X69, X70-X79, X80-X84

### Renal, Cancer and Cardiac conditions

These were defined using the SHMI diagnostic groups [13] which were in turn defined by the clinical condition groups specified by the Agency for Healthcare Research and Quality [29]

#### Renal

|  |  |  |
| --- | --- | --- |
| SHMI grp | CCS Conditions |  |
| 99 | 157 | Acute and unspecified renal failure |
| 100 | 156, 158 | Nephritis; nephrosis; renal sclerosis, Chronic renal failure |
| 102 | 160, 161, 162 | Diseases of kidneys and ureters, bladder and urethra |

#### Cancer

|  |  |  |
| --- | --- | --- |
| SHMI grp | CCS Conditions |  |
| 7 | 11 | Cancer of head and neck |
| 8 | 12 | Cancer of oesophagus |
| 9 | 13 | Cancer of stomach |
| 10 | 14 | Cancer of colon |
| 11 | 15 | Cancer of rectum and anus |
| 12 | 16 | Cancer of liver and intrahepatic bile duct |
| 13 | 17 | Cancer of pancreas |
| 14 | 18 | Cancer of other GI organs; peritoneum |
| 15 | 19 | Cancer of bronchus; lung |
| 16 | 20 | Cancer; other respiratory and intrathoracic |
| 17 | 22, 23 | Melanomas, other cancer of skin |
| 18 | 24 | Cancer of breast |
| 19 | 25 | Cancer of uterus |
| 20 | 26, 28 | Cancer of female genital organs |
| 21 | 27 | Cancer of ovary |
| 22 | 29, 30, 31 | Cancer of male reproductive organs |
| 23 | 32 | Cancer of bladder |
| 24 | 33, 34 | Cancer of urinary organs |
| 25 | 35 | Cancer of brain and nervous system |
| 26 | 37 | Hodgkin's disease |
| 27 | 39 | Leukemias |
| 28 | 40 | Multiple myeloma |
| 29 | 41, 45 | Other cancer (primary) |
| 30 | 42 | Secondary malignancies |
| 31 | 21, 36, 43 | Cancer of bone, thyroid and malignant neoplasm  |
| 32 | 44, 167 | Neoplasms (unspecified), Nonmalignant breast conditions |
| 33 | 46, 47 | Benign neoplasm |

#### Cardiac

|  |  |  |
| --- | --- | --- |
| SHMI grp | CCS Conditions |  |
| 57 | 100 | Acute myocardial infarction  |
| 58 | 101 | Coronary atherosclerosis and other heart disease |
| 59 | 102 | Nonspecific chest pain |
| 60 | 103 | Pulmonary heart disease |
| 61 | 104 | Other and ill-defined heart disease |
| 62 | 105 | Conduction disorders |
| 63 | 106 | Cardiac dysrhythmias |
| 64 | 107 | Cardiac arrest and ventricular fibrillation |
| 65 | 108 | Congestive heart failure; nonhypertensive |

## Appendix Results

Table 4: Results of univariable regressions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Predictor** | **Coeff** | **95% Conf. Int** | **Sign.** |
| **Demographics** | Age (md-point of 5 yr age group) | 0.001 | (0.001 to 0.002) |  p<0.001 |
|  | male | 0.03 | (0.03 to 0.04) |  p<0.001 |
| **Clinical** | NEWS | 0.04 | (0.03 to 0.04) |  p<0.001 |
| **Pathway** | Elective (scheduled) admission | -0.33 | (-0.34 to -0.32) |  p<0.001 |
|  | prior length of stay | 0.001 | (0.001 to 0.002) |  p<0.001 |
|  | Prior no. of ward\_transfers | -0.07 | (-0.07 to -0.06) |  p<0.001 |
|  | Prior no. of consultant\_transfers | -0.04 | (-0.04 to -0.03) |  p<0.001 |
|  | from\_high\_care (that day) | -1.04 | (-1.14 to -0.94) |  p<0.001 |
|  | from\_theatres (ditto) | -1.22 | (-1.38 to -1.06) |  p<0.001 |
|  | from\_adm\_unit (ditto) | -0.02 | (-0.06 to 0.02) |  p=0.260 |
|  | new\_hospital\_admission (ditto) | 0.35 | (0.34 to 0.37) |  p<0.001 |
| **Diagnostic** | renal | 0.48 | (0.45 to 0.50) |  p<0.001 |
|  | cancer | -0.05 | (-0.06 to -0.04) |  p<0.001 |
|  | cardiac | -0.12 | (-0.13 to -0.11) |  p<0.001 |
|  | respiratory | 0.23 | (0.21 to 0.24) |  p<0.001 |
|  | self\_harm | 0.72 | (0.64 to 0.80) |  p<0.001 |
|  | history\_of\_poisoning | 0.48 | (0.41 to 0.54) |  p<0.001 |
|  | dementia, delirium | 0.12 | (0.11 to 0.12) |  p<0.001 |
|  | mobility\_problems | 0.15 | (0.13 to 0.16) |  p<0.001 |
|  | falls, fractures | 0.07 | (0.06 to 0.08) |  p<0.001 |
|  | Pressure\_Ulcers, weight\_loss | 0.21 | (0.19 to 0.24) |  p<0.001 |
|  | incontinence | 0.31 | (0.29 to 0.32) |  p<0.001 |
|  | dependence\_and\_care | -0.10 | (-0.11 to -0.08) |  p<0.001 |
|  | anxiety, depression | 0.03 | (0.01 to 0.05) |  p=0.003 |
| **Organisational** | admission\_unit | 0.19 | (0.18 to 0.19) |  p<0.001 |
| **Derived** | comorbidity\_index | 0.01 | (0.01 to 0.01) |  p<0.001 |
|  | mortality\_risk | 0.86 | (0.82 to 0.90) |  p<0.001 |

## Appendix Software

All analyses were performed using R Statistical Software (v4.4.0)1 in the RStudio integrated development environment (v2024.12.0.467)2. The tidyverse R package (v2.0.0)3 was used for data wrangling, mgcv (v1.9-1)4 for GAM regression and yardstick (v1.31)5 for prediction metrics. Descriptive statistics used the finalfit R package (v1.0.8)6. The rmarkdown R package (v2.28)7 was used to format output in the preparation of this paper.

1 R Core Team (2024). \_R: A Language and Environment for Statistical Computing\_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

2 Posit team (2024). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

 3 Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” \_Journal of Open Source Software\_, \*4\*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686

4 Wood, S.N. (2017) Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

5 Kuhn M, Vaughan D, Hvitfeldt E (2024). \_yardstick: Tidy Characterizations of Model Performance\_. R package version 1.3.1, <https://CRAN.R-project.org/package=yardstick>.

6 Harrison E, Drake T, Pius R (2023). \_finalfit: Quickly Create Elegant Regression Results Tables and Plots when Modelling\_. R package version 1.0.7, <https://CRAN.R-project.org/package=finalfit>.

7 Allaire J, Xie Y, Dervieux C, McPherson J, Luraschi J, Ushey K, Atkins A, Wickham H, Cheng J, Chang W, Iannone R (2023). \_rmarkdown: Dynamic Documents for R\_. R package version 2.25, <https://github.com/rstudio/rmarkdown>.