# **Ward staffing guided by a patient classification system: multi-criteria analysis of "fit” in 3 acute hospitals**

# Abstract

## Aims

To assess how well the Safer Nursing Care Tool (SNCT) predicts staffing requirements on hospital wards.

To use professional judgement to generate hypotheses about factors associated with a ‘poor fit’.

## Background

The SNCT is widely used in the UK but there is scant evidence about factors that influence the quality of staffing decisions based upon such patient classification systems.

## Methods

Secondary analysis of data from 69 wards in three acute hospitals to assess the precision of the estimated staffing requirement, variation of estimates, correspondence with professional judgement and achieved staffing levels. Nursing workforce leads suggested factors associated with poor fit, based on the wards that rated worst.

## Results

39% of wards were frequently understaffed, while frequent overstaffing was less common (12%). 24% of wards needed a sample of over 182 days to estimate the establishment precisely. Potential reasons identified for poor fit included high turnover, older patients, high levels of 1-to-1 specialing, cancer care, small ward size and high within-day variation in demand.

## Conclusions

Using a staffing tool without applying professional judgement or triangulating against other methods can lead to inaccurate estimates of staffing requirements and unsafe staffing levels.

## Implications for Nursing Management

Despite the availability of software to calculate staffing requirements, application of professional judgement remains essential.

Keywords

Staffing levels; patient classification system; Safer Nursing Care Tool; health services research; quantitative methods

# Introduction

How many nursing staff should we employ on hospital inpatient units? A variety of tools exist to help guide this decision, but it has been recognised that no single tool is likely to capture the full extent of nursing work (Arthur & James, 1994; Fasoli & Haddock, 2010). A safe number of nursing staff depends on patient factors - the severity (acuity) of their condition and their dependence on nursing care - but also on contextual factors relating to the unit, as well as factors relating to the staff, including their experience and education (Saville et al., 2019, Hurst, 2003; Twigg & Duffield, 2009; Ball, 2010). Nursing involves doing multiple things at once, often carrying out unobservable work such as critical thinking, as well as responding to emerging demands and prioritising (Hughes, 1999). Add to this the unpredictability and shifting needs and numbers of patients (Edwardson & Giovannetti, 1994; Young, Lee et al., 2015), as well as restrictions on staff schedules, and it is clear why there is no simple, generally applicable method for determining a safe number of nursing staff to employ on a ward (Arthur & James, 1994; Fasoli & Haddock, 2010). Guidance says that professional judgement, (using knowledge and experience to help reach a decision) and triangulation (comparing the results from multiple staffing tools/systems) are both important (Ball, 2010; NICE, 2014; The Shelford group, 2014). This means that numbers from a measurement tool may be revised up or down based on other information.

However, in the context of increasing reliance on data-driven algorithms and evidence-based practice depending on empirical data, there is a risk that the apparent objectivity of a measured quantity is afforded undue weight in decision making (Chin-Yee & Upshur, 2018). This is particularly the case when the measurement itself is subject to statistical error and, potentially, systematic bias. However, evidence about how nurse staffing tools work in practice is scant (Edwardson & Giovannetti, 1994; NICE, 2014; Griffiths, Saville, Ball, Jones, et al., 2020) and generally does little to establish the accuracy of measurement beyond estimating inter-rater agreement and demonstrating correlation with criterion measures, including measures of dependency (Brennan & Daly, 2015; Larson, Cohen et al., 2017; Liljamo, Kinnunen et al., 2017).

Our recent study addressed this evidence gap by considering the Safer Nursing Care Tool (The Shelford group, 2014), the most-used tool in England (Ball et al., 2019). Our study found that the SNCT patient acuity/dependency measure was correlated with professional judgements of staffing adequacy, but for some wards this measure appeared to be biased (Griffiths, Saville, Ball, Chable, et al., 2020; Griffiths, Saville, Ball, Culliford, et al., 2020). Factors other than patient acuity/dependency measured by the SNCT were also correlated with professional judgements of staffing adequacy and for some wards the measured staffing requirement varied substantially by time of day and day of the week, which calls into question whether sampling patients at fixed times or on weekdays leads to an adequate estimate of typical requirements.

When the SNCT is used to estimate the required ward establishment, the measure is also subject to considerable error, even when based on the recommended minimum sample of all patients over 20 days. While the average precision in percentage terms appeared acceptable, the average width of the 95% confidence interval of the estimate is more than one whole-time-equivalent staff member either side of the mean. Thus, on average, even if the tools validity is accepted, the estimated establishment may provide one (or more) too many or one (or more) too few staff members to adequately meet patient need. Even where a tool is able to estimate typical daily workload with more precision, a separate issue is the robustness of a staffing establishment given varying demand. We found that scheduling staff to meet average demand (the basis the SNCT uses to set establishments) was frequently associated with understaffing in the face of variable demand from patients, short notice sickness absence and limited availability of temporary staff (Saville et al., 2021).

However, while we observed these patterns at hospital level, the performance of the tool varied across wards and so an unresolved issue is how well staffing tools work on wards with different characteristics and why. Exploring these issues can help to elaborate the role of professional judgement and factors that need to be considered that are not currently incorporated in the tool.

Using data from our previous study, the present study set out to explore ward-level variation in how the SNCT works for estimating staffing requirements according to multiple measures of fit and to generate hypotheses about wards where the tool, when used without applying professional judgement and triangulation, works least well.

# Methods

## Setting, data sources and study design

This was a secondary analysis of ward data from the parent study (Griffiths, Saville, Ball, Chable, et al., 2020). The setting was 69 acute adult inpatient wards from three hospital Trusts across the Wessex area of England. Wards were excluded if they were out of scope of the SNCT, e.g. those providing maternity or palliative care, and day units. The data consisted of linked data on actual staffing levels, the staffing requirement (using the SNCT) and a report of whether or not there were enough staff for quality. We had access to valid linked data from one year (2017) on 18,927 ward-days across the three hospital Trusts. Where wards underwent significant changes to the size or function during the study they were treated as multiple wards in the analysis, with part-year data available for each. The original study included one specialist hospital, which we did not include in the present analysis.

## a) Assessing how well the SNCT works in practice on different wards

The Safer Nursing Care Tool is a patient classification tool that works by categorising patients into levels (0, 1a, 1b, 2, or 3) according to their acuity and dependency on nursing care (The Shelford group 2014). By multiplying the number of patients in each level with a workload “multiplier”, managers can obtain an estimate for the number of nursing staff (both nurses and assistants) to employ, measured in full-time-equivalents. Since the number of patients and their acuity/dependency vary from day to day, the developers recommend collecting data for at least 20 days and setting staffing at the average estimate.

We considered a range of measures indicating how well the SNCT works in practice for both estimating staffing requirements and recommending baseline staffing levels. These were

1. the direction of the relationship between shortfall relative to the SNCT-requirement and the nurse in charge reporting enough staff for quality (yes or no),
2. the probability of the nurse in charge reporting enough staff for quality when staffing is at the SNCT-recommended level,
3. the number of days of SNCT ratings required for a precise estimate of the staffing establishment,
4. the difference between morning and evening SNCT-calculated staffing requirements,
5. the difference between weekday and weekend SNCT-calculated staffing requirements,
6. observed understaffing and overstaffing based on the nursing hours actually worked (recorded in the electronic roster) relative to the staffing requirement calculated using the SNCT,
7. understaffing and overstaffing in a simulation model of variable patient demand and flexible staffing response (see Saville et al 2020).

Details of these measures are given in Appendix Table 4. The association between a staffing shortfall relative to the SNCT requirement and nurses reporting that they do not have enough staff for quality is expected if the SNCT is measuring demand (criterion i) and the SNCT recommended level is able to meet demand (criterion ii). If the SNCT is to provide a reliable estimate of the staffing establishment needed, the estimates derived from it must be reasonably precise. If the number of days’ data required to provide a precise estimate is high (criterion iii), variability in demand is high which means that the recommended approach to using the tool is less likely to give an accurate estimate. Similarly if there is high variability within or between days it is possible that the tool does not properly estimate demand based on observations taken at fixed times and days (criteria iv and v). Finally, when understaffing or overstaffing (relative to the requirement estimated by the tool) is observed frequently, it is possible that unmeasured factors may be influencing staffing decisions (criterion vi). Alternatively, the calculated staffing requirement (the average of staffing requirements from 20 days) may not be able to meet variable demand from patients and real world constraints caused by (for example) short-notice sickness in the face of limited staff availability (criteria vi and vii).

We ranked wards according to each measure from best (rank 1) to worst fit (highest rank). Ties were given a joint rank (at the average position). We decided on cut-off values for each measure to represent unusual values or extreme poor fit, and flagged wards whose values fell outside this. The cut-off values were chosen based on practical reasons suggesting that the SNCT does not fit well or divisions/jumps in the empirical distributions, which indicated that some wards were unusual.

b) Generating hypotheses about wards where the tool works least well

In order to stimulate discussion and elicit professional judgements about factors that might contribute to a lack of fit, we presented a list of the most unusual wards to nurses with oversight of workforce at each hospital Trust. This list consisted of those wards that were flagged as unusual according to at least three of the nine measures, as well as those whose average rank across measures was in the top quarter of wards. In order to test the sensitivity of our process for deciding which wards to include in this list, we tested several alternatives (Supplementary material Table 5). In each case, the order of ranked wards was slightly different, but there was no change to the list identified. We asked the nursing workforce leads to consider potential reasons why the SNCT appeared to fit least well for these wards, without showing them the reasons why the wards had been flagged. They provided their professional opinion (in writing) of characteristics of those wards that might explain why the SNCT fits less well. We converted these ideas into the implied “hypotheses”, which we consider in the context of existing evidence of factors affecting staffing requirements.

## Ethical considerations

This study received ethics approval through the University of Southampton (18809.A1)and permission to undertake the study was given by the Health Research Authority (IRAS project ID: 190548).

# Results

Ward characteristics

Of the 69 wards included in the study, five are acute admissions units (Table 1), which have higher SNCT multipliers to allow for higher patient turnover (The Shelford group, 2014). Wards cover a wide range of specialties, including both surgical and medical, as well as mixed medical/surgical (e.g. oncology) wards. Wards have different sizes, layouts, staff skill mix and volumes of patients requiring one-to-one care. Wards’ staffing requirements according to the SNCT range from 5.9 to 10.2 care hours per patient day and on average, the actual staffing is lower than this at all Trusts. Nurses in charge reported enough staff for quality most of the time but on some wards, this was rare.

[Insert Table 1 here]

## a) Investigating how the SNCT works in practice on different wards

The number of wards where the establishment calculated from SNCT ratings is flagged as a poor fit according to each measure, and each combination of measures, is displayed in Table 2. For eleven of the 74 wards, no criteria were flagged.

The relationship between shortfalls in staffing relative to the SNCT requirement and nurse reports of having “enough staff for quality” differed between wards (Figure 1a-c). For six wards there was no evidence of the expected relationship of higher shortfalls associated with fewer reports of enough staff for quality. The probability of nurses reporting enough staff for quality when there is zero shortfall according to the SNCT ranged from 27% to 99% (81% on average). The probability of reporting enough staff for quality when staffing is at the SNCT-recommended level is expected to be high – if the SNCT provides a ‘perfect’ measure it would be 100%, but for thirteen wards this was less than 70%.

Although the current SNCT guidelines recommend collecting 20 days’ of data, for most wards this was insufficient for a precise estimate (defined as a 95% confidence interval 1 whole-time-equivalent wide). The number of days required ranged from 20 to more than 365 days’ data (the maximum available) with 165 days required on average. For 18 wards, more than half a year’s data (182 days) was required for this level of precision. Additionally, six wards had less than half a year’s data and could not provide a reliable estimate with the available data (sample sizes were 80-175 days, 112.5 on average).

Wards that had large differences between weekday and weekend requirements, or between mornings and evenings, were rare, but all of these were judged a poor fit according to other criteria too. The difference between weekdays and weekends was 3% on average, but ranged from -1% to 20% across wards. For four wards weekday assessments were substantially (more than 10%) higher than at weekends. Only one ward had higher requirements at the weekend. Four of the wards that were outliers according to the weekday-weekend criterion flagged as poor fit on three measures in total and one ward flagged four. On average, there was no difference between morning and evening assessments of staffing requirements but differences ranged from 19% bigger estimates based on morning observations compared to evenings, to evening assessments yielding 12% bigger estimates than mornings. Differences of more than 5% between morning and evening SNCT-calculated requirements were rare (4 wards). Three of these wards flagged three measures and the other one flagged two measures and ranked in the top (worst) quarter of wards overall.

Actual understaffing ranged between 0% to 85% of days (34% on average), while simulated understaffing of what is likely to happen if wards follow SNCT guidance ranged from 7% to 80% of shifts (31% on average). We considered more than 40% days with understaffing as a sign of poor fit and this was the most commonly flagged measure (29 wards). For 12 of these wards this was the only flagged measure. Simulated understaffing was also common (18 wards). Overstaffing was much less common (although more common in reality than in the simulation). Actual overstaffing ranged from 0% to 59% (10% on average), while simulated overstaffing ranged from 0% to 35% (5% on average).

[Insert Figure 1 here]

[Insert Table 2 here]

b) Generating hypotheses about wards where the tool works least well

We identified the wards that flagged at least three measures (thirteen wards), as well as those that ranked in the top quarter of wards (a further nine wards), and presented a list of these wards to the respective workforce leads for each hospital. Their expert judgements of potential reasons why the SNCT may fit less well in some wards are summarised in Table 3, and verbatim comments and flagged measures for each ward are provided in Supplementary Material Table 7. Some common themes emerged, leading us to formulate a number of implied hypotheses. Characteristics of some wards mean the SNCT multipliers, as they are applied, may underestimate the true workload. These factors include high patient turnover, an older patient population, cancer infusion/device activity and high levels of one-to-one specialing requirements. Although we did not identify which wards flagged high within-day variation, for several wards our experts suggested that large morning-evening differences in staffing requirements may result in the tool performing less well to estimate staffing requirements. For wards that are particularly small or large, it is possible that the number generated by the SNCT is less likely to be sufficient to maintain minimum registered nurse levels when rostering staff to shifts. Other factors that were suggested for a small numbers of wards are presented in the table, these included aspects of ward layout and systematic down coding of patients where ward staff felt the SNCT staffing levels were too generous.

[Insert Table 3 here]

# Discussion

## Implications of findings

This study adds to the scant evidence about how nurse staffing tools work in practice and offers insights into how a widely-used acuity/dependency tool fits in a range of wards. On the one hand, for some wards the tool appears to fit well; we found 11 wards flagged no measures and 30 flagged only one measure, which in some cases may not be due to a problem with the tool and in others could be addressed by minor adjustments to how the tool is implemented. On the other hand, for some wards multiple measures indicating poor fit were flagged indicating potentially serious problems with error and/or bias. In other words, if the SNCT is used to set staffing levels without applying professional judgement or triangulating results against other methods, as is recommended in the accompanying guidance (The Shelford group, 2014), there is a risk that resulting staffing levels will be unsafe.

These findings highlight the importance of training and expertise in setting staffing levels, and gives weight to repeated warnings against over-reliance on staffing tools and software (Edwardson & Giovannetti, 1994; Griffiths, Saville, Ball, Jones, et al., 2020; Twigg & Duffield, 2009). The results of this study suggest that a staffing estimate generated by a tool should be a starting point to question, rather than the indisputable correct answer.

Understaffing relative to the measured requirement was observed in many wards. Understaffing may be due to staffing shortages, but can also indicate that setting staffing at the average requirement does not provide enough staff on the day in the face of variable demand. It could also occur if local decision makers think the tool provides too many staff so adjust the establishment downwards from the recommended level. The fact that understaffing was common in the simulation is consistent with poor fit resulting from mean staffing being inadequate to meet variable demand. Overstaffing was rare, both in the observed data and the simulation.

For many wards a large sample size was needed to estimate the staffing establishment precisely. In some cases this could be dealt with by collecting more data but for wards with very variable requirements, setting the staffing establishment higher than the average is also likely to help cover more of the peaks in demand (Saville et al., 2021) and would reduce the risk of setting the establishment too low due to measurement error. Although shortfalls from the SNCT staffing requirement are associated with professional judgements of staffing adequacy overall (Griffiths, Saville, Ball, Chable, et al., 2020; Griffiths, Saville, Ball, Culliford, et al., 2020), this did not translate into the expected relationship on every ward. For thirteen wards, setting staffing levels each day at the level recommended by the tool, was associated with a low chance of saying there were enough staff, suggesting it is underestimating requirements here. Again this highlights the need for triangulation against other methods as there is no evidence to suggest that the measures are superior to professional judgement (Griffiths, Saville, Ball, Jones, et al., 2020).

The hypotheses suggested by our expert workforce leads, based on scrutinising a list of poor fit wards without knowing the specific flags that led them to be identified are, on the whole, congruent with evidence. Some of these suggested reasons for poor fit can be considered as patient factors: patients requiring one-to-one care, older patients and cancer patients were all put forward as potentially needing higher levels of care than the tool recommends. In our previous study we found that “specialing” requirements was a major source of variation in workload that is not directly accounted for by the SNCT (Griffiths, Saville, Ball, Chable, et al., 2020). However, evidence for older patients requiring adjustment to the multipliers is less convincing: a review of studies in 2014 found no clear evidence of specific differences in staffing requirements between ward types such as older people wards and others (Griffiths et al., 2014) although equally there is no evidence to support lower staffing or skill mix for older patients. More recently, Yoshida et al. (2019) found that age affected nurses’ perceived burden from a patient, although it is unclear to what extent this is already captured in SNCT acuity/dependency ratings.

Some cancer wards, along with some other speciality units, provide specialist care such as managing multiple infusions requiring close observations in patients who are not acutely unwell (Colombo et al., 2005). However it is again unclear to what extent this is captured in SNCT ratings, or whether nurses reliably apply the correct rating. There is a potential problem with nurses misinterpreting the concept of “normal ward care” when “normal” for a particular ward is more demanding than the implied definition in the tool: normal for general medical surgical wards.

High patient turnover was also suggested as a reason for poor fit. Although the SNCT multipliers incorporate an allowance for turnover, this may be insufficient for some wards; our measure of turnover showed a near 30-fold variation between the highest and lowest turnover wards. It seems unlikely that provision for a mean level can accommodate such variation. In the parent study the associations between patient turnover and reports of a variety of measures of staffing adequacy were not statistically significant (Griffiths, Saville, Ball, Chable, et al., 2020), but the direction of estimates was consistent with more turnover reducing perceived staffing adequacy for the same SNCT-shortfall (Griffiths, Saville, Ball, Chable, et al., 2020). Higher patient turnover per registered nurse has been associated with increased risk of mortality in other studies (Griffiths et al., 2018a; Needleman et al., 2011), suggesting it is a significant driver of nurse workload.

Large differences between ratings collected at different times of day or weekdays versus weekends were rare but were an indicator of poor fit overall and our experts highlighted large morning/evening differences as an issue for some wards (they did not see which wards flagged this measure). Wards with high within-day variation in patients per acuity/dependency category may also have high variation between days, making them more likely to flag criteria such as frequent understaffing and a large sample size for a precise establishment.

Nurses highlighted that for wards that are particularly small or large, the number generated by the SNCT is less likely to be sufficient to maintain minimum registered nurse levels when rostering staff to shifts. In the parent study we did not consider ward size in modelling the relationship with staffing adequacy although in our simulation we adjusted to account for the need to deploy whole people. Consequently, establishments were adjusted upwards by 2% on average but such upward adjustments are likely to be of greater magnitude on smaller wards (Griffiths, Saville, Ball, Chable, et al., 2020).

The identification of these factors affecting staffing requirements leads us to question whether they could be incorporated into the SNCT. However some of them are not patient factors so are not compatible with the core patient classification system, although they could potentially be addressed by creating ‘specialist’ versions of the tool for particular ward types. Others factors would add to the complexity of the data collection e.g. considering the age of patients. For each additional version there is a corresponding increase in data to be collected to develop the multipliers and the judgement to be applied in selecting the tool to fit. Given that it is already the case that multipliers need to be re-evaluated every few years as patients and care procedures change, this additional burden seems unwarranted and potentially unfeasible, undermining one of the core virtues of the SNCT – its simplicity. Publishing the variation in staff time in different quality-assured wards, rather than solely the average (the multiplier), might help to give some bounds on what are reasonable adjustments for managers to make to the multipliers for their wards.

## Limitations of this study

The sample in this study was large in terms of ward days, but came from only 69 wards at three hospitals so we cannot know whether these results generalise more widely. The simulation model was subject to assumptions about how staffing is managed on wards. The measures we considered, when taken individually, may be attributable to other causes rather than the performance of the SNCT, which is why we considered multiple measures. In particular, the “actual understaffing” measure could be flagged due to problems recruiting staff to a ward, or because of problems with the tool. We selected wards that exhibited a lack of fit based on multiple criteria to provide a stimulus for senior nurses to offer explanations. We did not directly test whether these explanations were correct and it may be that other factors, not identified here, are also relevant.

# Conclusions

Using a staffing tool without applying professional judgement or triangulating against other methods can lead to unsafe staffing levels; tools give a starting point to be questioned. Empirically, ‘poor fit’ criteria that were flagged most commonly included frequent understaffing and needing a large sample for a precise estimate. Workforce leads were able to identify reasons for poor fit of the tool that are on the whole consistent with research evidence, but more evidence is needed to test the hypotheses about the wards where the SNCT fits least well.

# Implications for Nursing Management

Training in applying professional judgement and triangulation remains important despite the availability of nurse staffing software and tools. In particular, nurses with responsibility for staffing decisions need to be aware of factors aside from acuity/dependency that may affect staffing requirements.

# References

Arthur, T., & James, N. (1994). Determining nurse staffing levels: a critical review of the literature. *Journal of Advanced Nursing*, *19*(3), 558-565. <http://dx.doi.org/10.1111/j.1365-2648.1994.tb01121.x>

Ball, J. (2010). *Guidance on safe nurse staffing levels in the UK*. London, Royal College of Nursing.

Ball, J., Barker, H., Griffiths, P., Jones, J., Lawless, J. A., Burton, C., Couch, R., & Rycroft-Malone, J. (2019). *Implementation, Impact and Costs of Policies for Safe Staffing in Acute Trusts*.

Brennan, C. W., & Daly, B. J. (2015). Methodological challenges of validating a clinical decision-making tool in the practice environment. *Western Journal Of Nursing Research*, *37*(4), 536-545. <https://doi.org/10.1177/0193945914539738>

Chin-Yee, B., & Upshur, R. (2018). Clinical judgement in the era of big data and predictive analytics. *J Eval Clin Pract*, *24*(3), 638-645. https://doi.org/10.1111/jep.12852

Colombo, A., Solberg, B., Vanderhoeft, E., Ramsay, G., & Schouten, H. C. (2005). Measurement of Nursing Care Time of Specific Interventions on a HematologyOncology Unit Related to Diagnostic Categories. *Cancer Nursing*, *28*(6), 476-480.

Edwardson, S. R., & Giovannetti, P. B. (1994). Nursing Workload Measurement Systems. *Annual Review of Nursing Research*, *12*(1), 95-123. http://www.ingentaconnect.com/content/springer/arnr/1994/00000012/00000001/art00006

Fasoli, D. R., & Haddock, K. S. (2010). Results of an integrative review of patient classification systems. *Annual Review of Nursing Research*, *28*, 295-316. <https://doi.org/10.1891/0739-6686.28.295>

Griffiths, P., Saville, C., Ball, J., Jones, J., Pattison, N., Monks, T., & Safer Nursing Care Study Group. (2020). Nursing workload, Nurse Staffing Methodologies & Tools: a systematic scoping review & discussion International Journal of Nursing Studies, 103(March), 103487, Article 103487. https://doi.org/10.1016/j.ijnurstu.2019.103487

Griffiths, P., Saville, C., Ball, J. E., Chable, R., Dimech, A., Jones, J., Jeffrey, Y., Pattison, N., Recio Saucedo, A., Sinden, N., & Monks, T. (2020). The Safer Nursing Care Tool as a guide to nurse staffing requirements on hospital wards: observational and modelling study. Health Services and Delivery Research 8(16), 1-161. https://doi.org/10.3310/hsdr08160

Griffiths, P., Saville, C., Ball, J. E., Culliford, D., Pattison, N., & Monks, T. (2020). Performance of the Safer Nursing Care Tool to measure nurse staffing requirements in acute hospitals: a multi-centre observational study BMJ Open, 10(e035828). https://doi.org/10.1136/bmjopen-2019-035828

Hughes, M. (1999). Nursing workload: an unquantifiable entity. *Journal of Nursing Management*, *7*, 317–322.

Hurst, K. (2003). *Selecting and Applying Methods for Estimating the Size and Mix of Nursing Teams: A Systematic Review of the Literature Commissioned by the Department of Health*. Nuffield Institute for Health.

Larson, E. L., Cohen, B., Liu, J., Zachariah, P., Yao, D., & Shang, J. (2017). Assessing Intensity of Nursing Care Needs Using Electronically Available Data [Article in Press]. *CIN - Computers Informatics Nursing*. <https://doi.org/10.1097/CIN.0000000000000375>

Liljamo, P., Kinnunen, U. M., Ohtonen, P., & Saranto, K. (2017). Quality of nursing intensity data: inter-rater reliability of the patient classification after two decades in clinical use. *J Adv Nurs*, *73*(9), 2248-2259. <https://doi.org/10.1111/jan.13288>

NICE. (2014). *Safe staffing for nursing in adult inpatient wards in acute hospitals*. https://www.nice.org.uk/guidance/sg1/resources/endorsed-resource-the-safer-nursing-care-tool-snct-11943902

Rauhala, A., & Fagerström, L. (2004). Determining optimal nursing intensity: the RAFAELA method. *Journal of Advanced Nursing*, *45*(4), 351-359. https://doi.org/10.1046/j.1365-2648.2003.02918.x

Saville, C., Monks, T., Griffiths, P., & Ball, J. E. (2021). Costs and consequences of using average demand to plan baseline nurse staffing levels: A computer simulation study BMJ Qual Saf, 30, 7-16. https://doi.org/10.1136/bmjqs-2019-010569

Saville, C. E., Griffiths, P., Ball, J. E., & Monks, T. (2019). How many nurses do we need? A review and discussion of operational research techniques applied to nurse staffing. International Journal of Nursing Studies, 97(September), 7-13. https://doi.org/10.1016/j.ijnurstu.2019.04.015

The Shelford group. (2014). *Safer Nursing Care Tool Implementation Resource Pack*. <http://shelfordgroup.org/library/documents/Shelford_Safer_Nursing_23May14a.pdf>

Twigg, D., & Duffield, C. (2009). A review of workload measures: A context for a new staffing methodology in Western Australia. *International Journal of Nursing Studies*, *46*(1), 132-140. <https://doi.org/http://dx.doi.org/10.1016/j.ijnurstu.2008.08.005>

Young, J., Lee, M., Prouty Sands, L., & McComb, S. (2015). Nursing activities and factors influential to nurse staffing decision-making. *Journal of Hospital Administration*, *4*(4). <https://doi.org/10.5430/jha.v4n4p24>

Table Ward characteristics. Supplementary material Table 2 gives the characteristics of wards at each hospital Trust

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | n | Mean | Minimum | Maximum |
| Wards | 69 |  |  |  |
| Medical | 32 |  |  |  |
| Surgical | 22 |  |  |  |
| Mixed medical/surgical | 15 |  |  |  |
| Acute admission units | 5 |  |  |  |
| Days with “enough staff for quality”, percentage |  | 79% | 24% | 100% |
| SNCT-estimated staffing requirement, hours per patient day |  | 7.24 | 5.9 | 10.2 |
| Actual staffing, hours per patient day |  | 6.77 | 5.0 | 10.4 |
| Bedsa |  | 27 | 8 | 63 |
| Single rooms, percentage |  | 23% | 0% | 75% |
| Patient turnover, ward average daily admissions and discharges per care hour |  | 0.06 | 0.01 | 0.29 |
| Skill mix, percentage registered nurses |  | 53% | 39% | 79% |
| Patients requiring one-to-one care (“specialing”), average daily percentage |  | 3.2% | 0% | 14.6% |

Table Number of wards where the establishment calculated from SNCT ratings is flagged as a poor fit according to different criteria

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 11 wards flag no measures | 30 wards flag 1 measure | | | | | 20 wards flag 2 measures | | | | | | | | | | 12 wards flag 3 measures | | | | | | | | | | | 1 ward flags 4 measures | **Number of wards flagging each measure** | **Percentage of wards flagging each measure** |
| **Number of wards with each combination of flagged measures** | **11** | **12** | **10** | **4** | **3** | **1** | **6** | **3** | **3** | **2** | **1** | **1** | **1** | **1** | **1** | **1** | **2** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **1** | **74** |  |
| High understaffing (actual)a |  | **x** |  |  |  |  | **x** | **x** |  |  | **x** | **x** | **x** |  |  |  |  | **x** | **x** | **x** | **x** |  |  |  |  |  |  | **x** | **29** | **39%** |
| Large sample size needed for a precise estimate of the establishment a |  |  |  | **x** |  |  | **x** |  | **x** | **x** |  |  |  |  |  |  | **x** | **x** | **x** |  |  | **x** | **x** | **x** | **x** |  |  | **x** | **24** | **32%** |
| High understaffing (simulation)b |  |  | **x** |  |  |  |  |  |  |  | **x** |  |  | **x** |  |  |  | **x** |  | **x** |  | **x** | **x** | **x** |  | **x** |  |  | **18** | **26%** |
| Low probability of reporting enough staff for quality at SNCT-recommended level a |  |  |  |  |  |  |  | **x** | **x** |  |  |  |  |  | **x** |  |  |  | **x** |  | **x** | **x** |  |  | **x** | **x** | **x** |  | **13** | **18%** |
| High overstaffing (actual) a |  |  |  |  |  | **x** |  |  |  | **x** |  |  |  | **x** | **x** | **x** |  |  |  |  |  |  | **x** |  | **x** |  | **x** |  | **9** | **12%** |
| Non-negative relationship between SNCT-measured shortfall and reports of enough staff for quality a |  |  |  |  | **x** |  |  |  |  |  |  | **x** |  |  |  | **x** |  |  |  | **x** |  |  |  |  |  |  |  |  | **6** | **8%** |
| Outlier for weekend-weekday differences in SNCT-calculated requirement a |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **x** |  |  |  | **x** |  |  |  |  | **x** |  | **x** | **5** | **7%** |
| Large morning-evening differences in SNCT-calculated requirement a |  |  |  |  |  |  |  |  |  |  |  |  | **x** |  |  |  | **x** |  |  |  |  |  |  | **x** |  |  |  |  | **4** | **5%** |
| High overstaffing (simulation)b |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | **x** | **x** | **2** | **3%** |

*Footnotes:* Crosses represent the measures flagged, e.g. Of the 30 wards that flag 1 measure, 12 wards flag high understaffing (actual) and of the 20 wards that flag two measures, 6 flag both high understaffing and large sample size needed for a precise estimate of the establishment.

*a – for these criteria, total possible is 74 because where there were ward changes, we considered as multiple wards,*

*b – total possible is 69 wards to avoid duplication of wards in simulation.*

Table Nursing workforce leads’ suggestions of reasons why the SNCT may fit less well in some wards

|  |  |  |  |
| --- | --- | --- | --- |
| **Ward factor** | **Expert judgement (of why wards in “poor fit” list)** | **Number of wards** | **Implied hypothesis** |
| **Throughput** | “As an assessment area with acutely unwell patients being admitted and transferred throughout the 24 hour period the acuity and dependency of all patients is not necessarily captured using the SNCT census checks am and pm.” | 3 | High throughput/  turnover means some work is not captured by the SNCT. |
| **Older patients** | “The Matron felt that the patients required the care hours allotted to a level 2 patient but could not be scored at this level due to the criteria used within the model.” | 5 | Wards with older patients require more care than what is captured by the SNCT. |
| **Specialing requirements** | “Many of the patients require two nurses to administer care for safety and dependency reasons or 1-1 care” | 5 | Wards with high specialing requirements have higher requirements than what is captured by the SNCT. |
| **Cancer infusion/device activity** | “The area has a high level of activity associated with infusions and devices.”  “There may also be an issue between the SNCT definitions of what is ‘normal ward care’ e.g. level 0 and the actual acuity/dependency of the patients who are having normal ward care… a cancer care ward expects high interventional IV therapies as normal care.” | 4 | Cancer infusion/device activity may not be captured by the SNCT multipliers or  patients may be underscored |
| **Morning-evening differences** | “Acuity AM will be lower than PM due to post op patients returning PM and being more acute.” | 3 | SNCT is a poor fit for  wards with large morning-evening differences in staffing requirements (based on acuity-dependency ratings) |
| **Ward size** | “…. anything less than 16 beds starts to make rostering difficult with the number of staff generated by SNCT and to also maintain things like minimum registered nurse etc. The same seems to be for those above 26.” | 4 | For wards that are particularly small or large, the number generated by the SNCT is less likely to be sufficient to maintain minimum registered nurse levels when rostering staff to shifts. |
| **Escalation ward** | “The area was also used for escalation during times of bed pressures – this means that as well as the ED/GP/’Hot’ clinic patients the area would have patients being admitted into beds for longer stays.” | 1 | Extra patients at some times of year means that there is high variation and may not plan staffing for this. |
| **Waiting patients** | “What’s not captured are patients in waiting area who are not ambulatory or in patients.” | 1 | Additional workload from waiting patients not accounted for |
| **Ward layout** | “Have to have 2 x RN’s in each area whether acuity requires it or not” | 1 | Geography of ward - separate areas mean higher requirements |
| **Change in types of patients** | “they also moved from taking trauma to elective patients so will have had a very mixed acuity/dependency of patients over the time” | 1 | Large variation in SNCT-assessed need over the time period means that SNCT does not fit well |
| **Local adjustment** | “As discussed. Can’t understand why they would look any different to [ward not in study, D\_12 or D\_13]” | 1 | Some nurses-in-charge may be down-coding patients in SNCT because they find that too many staff are provided when they follow SNCT |
| **No reason given** |  | 4 |  |

*Figure 1 Log odds of enough staff for quality dependent on shortfall (random slopes models)*

# Appendix

Table Criteria for judging the “fit” of the staffing establishment calculated from SNCT ratings

|  |  |  |
| --- | --- | --- |
| Measure | Details | What indicates a poor fit? |
| i) Direction of relationship between shortfall relative to SNCT-requirement and nurse-reports of having enough staff for quality | We looked at the association between the SNCT measure of required staffing and professional judgements of staffing adequacy. For this, we compared overall staffing shortfalls as assessed by the SNCT against daily nurse reports of having “enough staff for quality” (yes/no response). We fitted multi-level logistic regression models (nesting wards within hospitals) linking total shortfall according to the SNCT with reports of “enough staff for quality”. We allowed the intercept and slope to vary across wards and hospitals, so the overall slope for each ward gives the direction of the relationship. | For most wards, as would be expected, the slope was negative, i.e. as shortfall increases there are fewer reports of enough staff for quality. Therefore, we flagged those wards which did not exhibit this pattern as unusual (Figure 1a-c). |
| ii) Probability of reporting enough staff for quality when staffing is at SNCT-recommended level | Relatedly, we measured the chance of nurses reporting “enough staff for quality” when staffing is at the level recommended by the tool. This was also derived from the multi-level logistic regression model linking SNCT-shortfall and reports of “enough staff for quality”. It was calculated as the overall intercept for each ward evaluated at zero shortfall with the log odds converted to a probability using the formula exp(log odds)/(1 +exp(log odds)). | If the SNCT provides a valid measure of work on a ward, one would expect a high probability of reporting enough staff for quality when there is zero shortfall according to the SNCT. We flagged wards as unusual if this probability is less than 0.7 because this is where the number of wards drops off quickly (Supplementary material Figure 2). |
| iii) Number of days of SNCT ratings required for a precise estimate of the staffing establishment | Since the SNCT was designed to give an estimate of the staffing establishment required on a ward by taking the average workload across at least 20 days (The Shelford group 2014), we calculated how many days of acuity/dependency ratings were needed for a precise estimate. In particular, we considered the minimum sample size needed for a stable estimate of the average requirement for 95% confidence interval 1 whole-time-equivalent wide. Full details of the calculations, which used a bootstrapping approach, can be found in the report and paper from the previous project (Griffiths, Saville et al. 2020, Griffiths, Saville et al. 2020). | Since the SNCT guidelines state that establishment-setting should take place twice a year, we flagged wards that needed more than half a year’s worth of data (182 days) for a reliable estimate of the average requirement (Supplementary material Figure 3). |
| iv) Difference between morning and evening SNCT-calculated staffing requirements | Given that the guidelines recommend collecting SNCT ratings per day at a fixed time, we also considered the stability of patient numbers per acuity/dependency category over the day. For this, we considered the average absolute difference (i.e. the direction of the difference did not matter)between SNCT staffing requirements (in hours) calculated using data gathered in the morning compared to data gathered in the evening, expressed as a percentage difference. | Large values of this measure highlight wards with large fluctuations in SNCT-calculated staffing requirements between mornings and evenings. For these wards, the time of day at which ratings are collected could make a big difference to the staffing establishments calculated. Differences of more than 5% between morning and evening SNCT-calculated requirements were flagged as unusual because wards outside this range appear to be outliers (Supplementary material Figure 4). |
| v) Difference between weekday and weekend SNCT-calculated staffing requirements | Since the SNCT guidelines recommend collecting ratings on weekdays, we also looked into differences between weekday and weekend staffing requirements (as measured by the tool). We recorded the average percentage difference between weekday and weekend SNCT staffing requirements (in hours). Positive differences indicate that weekdays have higher requirements than weekends. | For most wards, assessments of staffing requirements were higher using weekday data than weekend data (Supplementary material Figure 5), so we flagged wards as unusual if weekend assessments were higher than on weekdays (negative difference) or if weekday assessments were substantially higher than at weekends (more than 10% higher) |
| vi) Observed (actual) understaffing and overstaffing | We compared the daily requirement measured using the SNCT with the staff actually deployed. We considered any day “understaffed” if staffing levels were more than 15% below the required hours according to the SNCT. A 15% threshold was chosen consistent with the RAFAELA tool (Rauhala and Fagerström 2004), which recognises that there is some flexibility to cope with slightly different workloads with the same number of staff.  Analogously to understaffing, we also looked at the frequency with which wards were overstaffed relative to the SNCT requirement. | Wards were flagged as unusual if more than 40% of days were understaffed because there is a dip in the data there (Supplementary material Figure 6). While such understaffing may not result from inaccuracies in the measured staffing requirement, high levels of understaffing could arise if the SNCT calculated establishment is unable to meet fluctuating demand or if there are perceptions that the measured requirement is an over-estimate.  Frequent overstaffing could signify that following the SNCT guidelines to set a staffing establishment leads to this ward having a high chance of being overstaffed. Another explanation is that the ward manager adjusted the staffing level generated by the SNCT upwards because of some work on this ward not being captured by SNCT acuity/dependency ratings. Either way, this is an indication of poor fit. Wards were flagged as unusual if more than 30% of days were overstaffed because on visual inspection there is a big break in the data there (Supplementary material Figure 7). |
| vii) Understaffing and overstaffing in a simulation model of variable patient demand and flexible staffing response (see Saville et al 2020). | Since “actual understaffing” could be a sign of nursing shortages rather than a problem with the tool, as a complementary measure, we also considered the frequency with which wards are likely to experience understaffing if SNCT guidelines for setting staffing establishments are followed. For this we used the outputs of a simulation documented in full elsewhere (Griffiths, Saville et al. 2020, Saville, Monks et al. 2021). In brief, this computer model simulates how often wards are understaffed, given that there is likely to be some attempt to adjust for shortfalls, by limited use of temporary staff and moving staff from apparently overstaffed to understaffed wards. We assumed that staffing establishments are based on morning SNCT acuity/dependency ratings collected on 20 days in June (or next month if missing). The ratings were combined by averaging workload across this 20-day period, as recommended by the SNCT guidelines (The Shelford group 2014), and then adjusting to make sure that there are whole people and at least one registered nurse working per shift. We considered the percentage of patient shifts that were understaffed in the simulation, which gives simulates how often patients are exposed to understaffing more than 15% below the level required according to the SNCT ratings for that shift.  As well as simulated understaffing, we also considered simulated overstaffing. This shows how often it is likely there will be “surplus” staff on a ward if SNCT guidelines for setting staffing establishments are followed. We calculated this analogously to simulated understaffing, considering how often patients were exposed to staffing more than 15% above the level required according to SNCT ratings for that shift in the simulation. | As for actual understaffing, wards were flagged as unusual if more than 40% of shifts were understaffed.  As for actual overstaffing, wards were flagged as unusual if more than 30% of shifts were overstaffed. |