Predicting nursing staff requirements: How can we improve decision support tools?

Rich C. McIlroy[1], Talia Emmanuel[2], Chiara Dall'Ora[2] & Christina Saville[2]

[1]Transportation Research Group, University of Southampton, UK, [2]Health Workforce and Systems Research Group, University of Southampton, UK

SUMMARY

Most hospital trusts across England's National Health Service use acuity/dependency tools to estimate nurse staffing needs, which require trained nurses to input patient and other administrative data at multiple timepoints. It is possible that information already recorded about patients could be used instead to guide staffing requirement predictions. This research asks how best to incorporate such predictions into software products, doing so through a user-centred design process involving focus groups with clinical nurses and nursing managers.

KEYWORDS

User-centred design, applied cognitive task analysis, nursing staff requirements

Introduction

Having enough nurses caring for patients on hospital wards is vital for patient safety but planning for varying numbers of patients with unknown deterioration/recovery trajectories is difficult (Saville et al. 2019; Musy et al. 2020). Extensive research shows that care quality and safety are compromised when nurse staffing is low (Blume et al. 2021; Griffiths et al. 2018; Dall'Ora et al. 2022; Shin et al. 2018). but deciding how to allocate nursing staff is challenging because demand for care fluctuates (Griffiths et al. 2020a); both patient numbers and their needs vary over time. Decisions arise at multiple points including determining how many staff to employ (the establishment) and how many to deploy on each shift (Ernst et al. 2004).

To help guide decisions, a range of tools attempt to measure demand for nursing care and determine the required number of staff (Griffiths et al. 2020b). In England, the National Institute for Health and Care Excellence (NICE) endorsed the Safer Nursing Care Tool (SNCT), now used in almost all acute NHS Trusts (Griffiths et al. 2020a). Designed for establishment planning and also incorporated into SafeCare, a digital tool for short term decisions around on-shift staff deployment, it requires that patients are classified into one of seven categories based on an assessment of acuity and dependency on nursing care. When used for establishment planning, trained observers assess all patients on a ward daily for 30 days and the establishment is set based on an average (InnovaHealthTec, 2023). Increasingly the acuity/dependency classification is carried out by ward nurses, sometimes multiple times per day. This is a labour-intensive process that necessitates the collection of data specifically to measure staffing requirements, not informed by data gathered for administration or care management. Because acuity/dependency assessments are intermittent, there is limited ability to match demand to staff in real-time. Even when used for establishment setting, prescribed methods for sampling do not reflect typical demand on some types of wards (Griffiths et al. 2020a; Griffiths et al. 2020c). Furthermore, the use of a limited number of categories does not reflect the continuous variability of need.

This project therefore seeks to develop and validate machine learning methods to estimate nurse staffing requirements using routinely collected data. If successful, nurses' time can be saved, because bespoke assessments can be avoided. Moreover, training nurses to use such tools requires time and resources. However, for such a system to prove effective, it must be deemed both useful and useable by those who will ultimately interact with it, namely clinical nurses and nursing managers. For this to be achieved, those end-users must be involved in the development process.

This research is guided by user-centred design principles, specifically the human-centred design of artificial intelligence (Margetis et al. 2021). The main point of user-centred design is "to ask what the goals and needs of the users are, what tools they need, what kind of tasks they wish to perform, and what methods they would prefer to use." (Norman & Draper, 1986). This will increase the chance that any systems developed will be fit for purpose and serve nurses' and managers' needs (van Oostveen et al. 2016). This is relevant for the design of both the underlying algorithms and of the information system with which nurses and managers interact.

Study aims and objectives

The aim of this research is to understand what does/does not work for nurses and managers when using staffing tools, and incorporate this into the design of a new, machine-learning based tool. The research will capture nurses' and managers' needs and views around staffing decision tools, adapting the Applied Cognitive Task Analysis interview method (Militello & Hutton, 1998). It will extract information about the cognitive demands and skills required in making nursing staffing decisions, and how existing tools are currently used to support that decision-making process.

Method

Two focus groups were convened to understand end-user decision support needs. One focus group was held with clinical nurse ward managers that record acuity/dependency ratings and make operational staffing decisions. The other was held with nursing staff leaders responsible for staffing decisions at the strategic level. Each focus group was hosted using Microsoft Teams, lasted approximately one hour, and was facilitated by two researchers (with one taking the lead).

The first, held in January 2026, had four participants, all of whom were clinically active registered nurses at the ward manager (or sister) level, with responsibility for the daily operational management of a ward or unit. Three were hospital based, one was based in a community day unit. Two NHS trusts were represented (two nurses from each), both of which were in the north of England. The second, held in February 2026, had six participants, all of whom were senior staffing leaders at or above the matron level. Five were nurses, one was administrative. All were responsible for strategic decision making and workforce planning, sitting at one or more organisational levels above the ward managers that took part in the first focus group. Five NHS trusts were represented, from across England. All participants were female. The study was approved by the University of Southampton and the Health Research Authority (UoS ID 100780; IRAS ID 346148).

The focus groups were structured using probes adapted from those discussed by Militello and Hutton (1998) in their Applied Cognitive Task Analysis framework. Some probes were worded differently for the two focus groups, reflecting the different responsibilities of the participants, and the ways in which they use and interact with staffing decision support tools. The probes were developed in consultation with the project's advisory group, comprising health workforce organisation research experts and practice-based experts. Three broad topics were covered in each focus group. In advance of probing the participants for input, they were asked to think about times when they were making staffing decisions and using acuity/dependency tools. The investigator then facilitated the discussion by providing several prompts. Some examples are summarised in Table 1, below, alongside the three main topics covered in each focus group.

Table 1. Summary of topics and example probes used in the two focus groups.

Sister / ward manager focus group	Matron / staffing leader focus group
Current staffing levels: To map how staffing decisions are made and help participants recall experiences. Example probes: • Think about a time when decisions around nurse staffing numbers were being made on your ward. What sorts of data and tools do you believe were needed to make those decisions? • When reviewing your list of data and tools, how was this information collected? • Have there been times when a decision around nurse staffing has been made that doesn't seem to reflect the reality on wards?	*Current staffing levels* To map how staffing decisions are made and help participants recall experiences. Example probes: • Think about a time when decisions around nurse staffing numbers were being made on your ward(s). What were the most important sources of information you used to help make your decisions around nurse staffing? • Which colleagues did you reach out to or rely on to help you make these decisions? • In what ways have any difficulties in recruiting staff impacted your ability/procedure for making decisions around nurse staffing?
Recognising ward cues: To understand the cues and signals nurses rely on to escalate risks relevant to nurse staffing levels • What are the signals or cues that would lead you to believe that your ward is at risk of being understaffed? • When reviewing your list of signals and cues, how do you go about escalating any risks? • What might someone with your professional expertise and experience notice that someone else might miss?	*Use and perception of existing tools:* To understand how tools/data systems are used, trusted, or adapted by participants in practice. • What is the most useful information source/tool? The least useful? • When looking at your list, do you generally trust what these systems or data tell you? Why or why not? • Have there been times when data/tools pointed in one direction, but your own professional judgement told you to do something else?
Decision support needs: To understand which decision-support features are most helpful for making staffing decisions • If you could design the ideal decision-support tool or data dashboard, what would it include? • How could this ideal tool be designed so that it fits into your daily workflow without adding additional workload or burden? • How might this tool be customised for different types of wards and specialties?	*Decision support needs:* To identify unmet needs and determine which decision-support features or tools are most helpful for making decisions around staffing. • What type of staffing decisions do you most want support with? E.g., would you prefer tools to help you make decisions in real-time or tools to help with long-term planning? • What type of support would be most useful? • What would make you trust a recommendation or alert that was automatically generated?

Analysis

The focus group recordings were transcribed, de-identified, and uploaded to an offline large-language model (LLM, Microsoft Copilot) for textual content analysis. This approach to the analysis of qualitative data is growing in popularity for its efficiency and reliability (Liu & Sun, 2025; Randerson et al., 2025) and is particularly well suited to identifying broad concepts in data (Acciai et al., 2025). The approach was thus deemed appropriate for the context of this conference paper, which emphasises global conceptual discussion (due to limitations on article length that preclude detailed data reporting). In line with advice from the field (e.g., Castellanos et al., 2025), the outputs of the LLM were reviewed by the human research team alongside reviewing of the

focus group recordings. The LLM outputs were interpreted and revisions made where considered appropriate. Further reviewing is currently underway and the results presented below can be considered a preliminary analysis that will be refined by discussions with the wider research team.

Results and discussion

The thematic coding schemes resulting from the analysis of the sister / charge nurse focus group and the matron / staffing manager focus group transcripts are presented in Tables 2 and 3, respectively, with short definitions of each theme.

Table 2. Themes (and descriptors) identified in the sister / ward manager focus group.

Theme	Description
1. Tool - reality mismatch: Acuity	
1.1 *Judgement overrides tool*	Instances where clinical judgement explicitly supersedes tools / acuity classification in deciding staffing.
1.2 *Category doesn't fit patient risk*	Where the tools risk categories do not match actual patient risk as judged by clinical staff.
2. Tool - reality mismatch: Staffing	
2.1 *"Hours short" while fully staffed*	Tool shows hours deficits even when ward perceived as fully staffed.
2.2 *Medico-legal anxiety (incidents/Coroner)*	Concern about scrutiny of tool shortfall in the event of an incident.
3. Hidden workload	
3.1 *Estate constraints (lift/heating/call systems)*	Building failures (lifts, heating, call systems) increasing workload.
3.2 *Security/medical cover gaps*	Insufficient on-site security/medical cover affecting staffing need.
3.3 *Ward layout/geography workload*	Side rooms and large ward footprints increasing time needed.
3.4 *Patient information*	Patient characteristics increase workload, e.g., bariatric, cognitive.
4. On-shift availability	
4.1 *Off-site appointments*	Out-of-ward appointments removing staff for multiple hours.
4.2 *RN escort requirement*	Requirement for RN (not HCA) to escort (e.g., in case of sedation).
4.3 *Variation in "clock-out" capability*	Some sites can redeploy/clock out escorts in the tool; others cannot.
5. Skill and experience mix	
5.1 *Band distribution matters*	Impact of Band 2/3/4/5 mix on safe coverage.
5.2 *Senior presence (Band 6/7) at risk times*	Need for Band 6/7 during volatile periods; leadership availability.
5.3 *Task/competency constraints (sedation)*	Role-based tasks (e.g., sedation recovery requires RN).
6. Temporal variability	
6.1 *Mealtimes*	Feeding needs temporarily occupy multiple staff members.
6.2 *Night assumptions ("everyone asleep")*	Tool assumptions that nights are lower intensity; contested by lived reality.
6.3 *Out of hours thin cover*	Reduced medical cover windows (e.g., 09:00–14:00 weekends); dismissive/slow responses from registrars; documentation delays.
7. Escalation Barriers & Culture	
7.1 *Band 5 confidence*	Reluctance/insecurity to escalate without senior staff present.
7.2 *Distance/logistics to decision-makers*	Physical separation from decision-makers; time to find help.

7.3 *Local workarounds (phone/WhatsApp)*	WhatsApp, nurse-in-charge phone, management support.
8. Local customisation & Prediction	
8.1 *Site-/ward-specific parameters*	Desire to tailor algorithms to site support, layout, service type.
8.2 *Structured professional overrides*	Want an input to capture judgement with reasons/time limits.
8.3 *Event-driven, real-time updates*	Automatic re-calculation when escorts depart/return; new DoLS; transfers.
8.4 *Forward forecasting (appointments/transfers)*	Predictive insight based on booked appointments/transfers/known outages.

Notes: DoLS: Deprivation of Liberty Safeguards; HCA: Health Care Assistant; RN: Registered Nurse.

Table 3. Themes (and descriptors) identified in the matron / staffing leader focus group.

Theme	Description
1. Evidence-informed but judgement-led	
1.1 *Data + judgement on the day*	Data guides decisions, but clinical judgement and interpretation key.
1.2 *Professional validation*	Senior nurse review to ensure accuracy and credibility of scores.
1.3 *Touch-points vs. deep reviews*	Light-touch mid-year vs in-depth annual establishment reviews.
2. Strengths and limits of tools	
2.1 *Good for 'typical' inpatient wards*	Outputs generally align well with stable, routine, predictable wards.
2.2 *Poor fit for nuanced settings*	Struggle to represent changeable, highly acute, or specialist wards.
2.3 *Birthrate Plus/other specialties*	Midwifery and neonatal services lack equivalent tools.
3. Data quality, training & consistency	
3.1 *Structured training & audits*	Formal training required to standardise scoring and prevent drift.
3.2 *Over- and under-scoring patterns*	Variation in how staff score acuity leads to inflation or minimisation.
3.3 *Ward-level validation layers*	Multiple levels of checking (ward leader, matron, divisional review) needed to confirm scoring accuracy.
4. Skill mix matters beyond headcount	
4.1 *Seniority/banding effects*	More experienced and higher-band staff influence safety.
4.2 *Qualified vs. unqualified balance*	The ratio of registration types affects clinical capability.
4.3 *New roles not captured well*	Emerging roles (e.g., Nursing Associates, clinical aides) sit awkwardly in existing staffing categories.
5. Colour codes, red flags, and redeployment	
5.1 *Utility for rapid site-wide decisions*	RAG ratings provide overview to support quick staff redeployment.
5.2 *Manipulation/behavioural risks*	Staff may alter acuity entries to influence the wheel's RAG rating and avoid losing staff.
5.3 *Diverse operating models*	Trusts adapt tools differently to combat potential manipulation/gaps.
6. Specialty standards, layouts, and escalation spaces	
6.1 *Standards (BTS, BAPM)*	External clinical standards not always reflected in tool calculations.
6.2 *Ward layout effects*	Layout variations increase supervision needs beyond tool assumptions.
6.3 *Pop-up escalation areas*	Temporary or overflow spaces create additional staffing pressures.

7. Headroom, sickness, and workforce pipelines	
7.1 *Local headroom policies vary*	Trusts set different uplift, making outputs difficult to compare directly.
7.2 *Seasonality & absence forecasting*	Patterns of sickness, demand surges, and patient acuity vary seasonally.
7.3 *Bank/agency thresholds*	Leads seek clearer guidance on safe levels of temporary staffing use.
8. Triangulating quality and harm with staffing	
8.1 *Harm indicators alongside rosters*	Falls, pressure ulcers, and other harms are analysed alongside staffing.
8.2 *Narrative building*	Multiple datasets scrutinised when tool outputs contradict experience.
8.3 *AI wish-list*	Desire for systems that integrate staffing, acuity, and harm data.
9. Interoperability and benchmarking challenges	
9.1 *Model Hospital mismatch*	Comparative tools use inconsistent definitions, reducing reliability.
9.2 *System interoperability gaps*	Different systems don't seamlessly link, limiting holistic insight.
9.3 *Group variation*	Even within merged trusts, staffing practices differ significantly.
10. Interoperability and benchmarking challenges	
10.1 *Live acuity from EPR/NEWS*	Real-time patient acuity from observations more credible than snapshot.
10.2 *Explainability and context*	Alerts must show why they were generated and include contextual factors such as supervision or mitigations
10.3 *Preserve "authentic intelligence"*	Digital recommendations must support, not replace, human judgement.

Notes: RAG: red, orange, green; BTS: British Thoracis Society; BAPM: British Association of Perinatal Medicine; EPR: electronic patient record; NEWS: National Early Warning Score.

Sister / ward manager focus group summary

The focus group revealed that charge nurses view the tools as useful but fundamentally limited when compared with their day-to-day professional judgement. Across services and sites, participants consistently described situations in which the tool's outputs diverged from the lived reality of providing safe care. As one ward manager put it, *"Sometimes what's written down… just doesn't tally. That does just come from your professional judgement… knowing your patients."* This theme of judgement overriding the tool was especially evident in discussions about risk categories and acuity levels, which managers felt often misclassified patients whose behaviour or needs were more complex than the labels implied.

A central concern was the mismatch between tool results and staffing reality, with many describing incidents where they were fully staffed on paper yet were declared significantly understaffed by the tool. One nurse noted, *"I'm fully staffed… but SafeCare says I'm 30 hours short,"* highlighting the anxiety this creates, particularly where incident investigation or Coroner scrutiny (in the event of an incident) may subsequently rely on recorded metrics. This discrepancy contributed to reduced confidence in the tool as a real-time operational guide.

Managers also emphasised that the tool does not capture crucial contextual factors, including ward layout, building constraints, and site-specific risks. At an off-site rehabilitation unit, the absence of overnight medical cover and security fundamentally shaped staffing needs. As one participant described, *"We have no medical cover overnight whatsoever… and no security,"* emphasising that

the tool's assumptions about comparable conditions across sites did not reflect reality. Others highlighted the impact of estate failures, such as a broken bed-sized lift requiring staff to hoist patients into chairs and transport them via a single remaining lift, activities that consume substantial staffing hours.

Another recurring theme was the hidden workload of escorts and transfers. Routine appointments frequently removed staff from the ward for extended periods. One manager reported, *"The quickest turnaround is two hours… so I'm losing three staff members for between two and four hours,"* yet this was typically invisible to the tool. Some sites had the ability to "clock out" staff temporarily within the tool, but others did not, leading to systematic underestimation of demand and variability in the utility of the tool across settings.

Finally, participants argued strongly for customisation and predictive capability. They expressed a desire for tools that incorporate local factors, skill-mix nuances, and real-time events. A sentiment repeated across the group was that tools should be able to integrate professional insight: *"It needs to balance between what it says in black and white and what the reality is."*

Matron / staffing leader focus group summary

This focus group revealed a workforce leadership community that is data-informed yet firmly judgement-led. Participants consistently described a triangulation process that blends tool outputs, roster efficiency metrics, red-flag escalations, and patient harm and experience data, tempered by senior clinical review. As one lead put it, establishment decisions are like *"adding all the jigsaw pieces together"* to build a defensible narrative for strategic staffing decisions.

While the SNCT system is widely embedded, its limitations are salient in nuanced or high-flow settings. Leaders accept that *"for the majority of our inpatient areas, yes… it has its successes,"* yet it *"is not reflective"* for complex admissions/assessment units, emergency departments, respiratory or oncology specialisms, and single-room layouts that drive supervision intensity. Midwifery and neonatal services were flagged as areas lacking robust acuity tools aligned to specialty standards.

A dominant theme was data quality and scoring culture. Variability across wards, and even across trusts within a group, can be profound. One matron described daily scrutiny, WhatsApp prompts, and weekly review meetings to ensure consistency: *"I train everybody… scrutinise the data every day."* Others report over- and under-scoring, detected by benchmarking similar wards: *"Two same profiles… one area is scoring way higher… it's the culture of the scoring on the ward."* Structured validations (ward, matron, divisional) are used to correct drift and ensure the numbers *"feel right"*.

Crucially, skill mix was highlighted as being more important than headcount. Banding, experience, and role composition (e.g., Nursing Associates, clinical aides) are not adequately captured in binary registered/unregistered splits. *"It's not just looking at the numbers, it's looking at the skills that they possess…"* Moreover, the presence of senior staff can prevent escalation and reduce harms, while cohorts of newly qualified nurses change supervision needs. Supervisory ward leaders are seen as pivotal, yet their contribution is hard to represent in current tools.

On day-to-day coordination, organisations diverge. Some of the participants discussed a wheel with red / amber / green (RAG) ratings for safety, and twice-daily safety huddles to enact redeployment, overlaying RAG status with contextual mitigations. Others discussed having disabled multipliers or removed acuity/dependency ratings from the tool after concerns about predictive accuracy, burden, or gaming to ensure ward staff were not redeployed: *"It was potentially open to manipulation."* Where used, red flags are intended for escalation, but leaders report mixed utilisation and the need for oversight to prevent false positives and to ensure timely updates when mitigations occur.

Participants reported wanting predictive analytics that forecast demand and absence (seasonality, sickness) and safe thresholds for bank/agency utilisation. They also sought interoperability, linking electronic patient record (EPR) observations (e.g., National Early Warning Scores), rosters, incidents, and patient harms data, so that alerts are real-time, contextual, and explainable. As one lead suggested, hour-by-hour acuity from the EPR would be *"more accurate"* than twice-daily snapshots. Another envisioned AI supported systems that *"interlink[s] incidents with staffing"* to guide redeployment and establishment reviews while preserving *"authentic intelligence"*.

Finally, trust in any new automated recommendation hinges on transparent logic, triangulated data, and a clear route for senior validation. Tools must recognise specialty standards, physical layout, escalation areas, and supervisory presence; all the context that experienced leaders already weigh instinctively. As one participant concluded, the goal is not to replace human judgement, but to augment it with timely, integrated insight.

Similarities and differences

Across both workshops, participants emphasised that nurse staffing decisions rely on a blend of routinely collected data (e.g., SafeCare, SNCT scoring, EPR, patient harms, acuity and dependency measures) and professional judgement. In both groups, staff acknowledged the limitations of existing tools, particularly the mismatch between recorded acuity and the lived reality of patient needs, the impact of ward layout, and the importance of skill mix rather than simple headcount. Data from both workshops highlighted the complexity of interpreting measures produced by current tools, inconsistencies in scoring practices, and substantial variation in data quality depending on staff training and ward culture. Participants of both groups described relying heavily on tacit knowledge, cues from staff morale, patient behaviours, and the physical environment when determining risk and deciding whether staffing was safe.

The key differences related to focus and scale. The staffing leads focus group dealt mainly with strategic, trust-wide establishment setting, validating tool data, and ensuring consistency across multiple wards and hospitals. They emphasised benchmarking, governance, and the challenges of applying national tools to specialist areas (e.g., respiratory, oncology, neonatal). As would be expected, the ward managers focused more on operational realities: responding to sudden acuity changes, escorting patients off the ward, infrastructure failures, lack of security, geographical isolation, and barriers to escalation. Their discussion was grounded in hour-to-hour decision making rather than organisation-level workforce planning.

Summary and conclusion

The focus group insights highlight that any AI-supported staffing system must be grounded in genuine user input. Individuals at both the charge nurse and matron levels of the organisational hierarchy emphasised that measurements from acuity/dependency tools are often inaccurate because they do not reflect real-world context, local workflow, or professional judgement. For AI systems to be trusted, they must therefore be co-designed with end-users, incorporate ward-specific nuances (layout, security, patient flow, escorts, estate failures), and provide transparent reasoning so that nurses can understand and challenge predictions rather than feel overruled by them.

The findings also show strong appetite for AI that uses routinely collected data, for example electronic patient record observations, rosters, incident reports, skill mix data, and sickness patterns, to produce real-time and predictive insights. Such data offers richer, continuous signals than twice-daily acuity scoring and could reduce the burden of manual acuity/dependency assessments. However, participants stressed that data quality varies and cultural factors shape scoring practices, meaning AI must include mechanisms for validation, anomaly detection, and user feedback to avoid embedding existing inaccuracies.

Ultimately, the implications point to AI as an augmentative tool. Due to the complexity and unmeasurable nature of some aspects of nursing, predictive models are helpful tools but are unlikely to ever be able to fully replace professional expertise. Ideally these models would automatically incorporate multiple data streams and present context-sensitive recommendations that reflect the complexity and variability of nursing work.

The work presented here marks the beginning of a broader user-centred design process. The next phases of the ongoing project will build on these findings to shape staffing decision-support tools that harness the value of routinely collected data to enhance efficiency and workflow. Equally, these tools will be developed in ways that future users perceive as practical, relevant, and, especially vital for AI-enabled systems, worthy of their trust.

Acknowledgement

This project has been funded by the National Institute of Health and Care Research (NIHR; Award ID: NIHR166784). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. We acknowledge and thank the wider PREDICT-NURSE study team, as well as the workshop participants for their time and insights.

References

Acciai, A., Guerrisi, L., Plebe, A., & Suriano, R. (2025). Neural language models as content analysis tools in psychology. *Philosophical Psychology,* 1-24.

Blume KS, Dietermann K, Kirchner-Heklau U, Winter V, Fleischer S, Kreidl LM, et al. (2021) Staffing levels and nursing-sensitive patient outcomes: Umbrella review and qualitative study. *Health Serv Res*.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3(2),* 77-101

Castellanos, A., Jiang, H., Gomes, P., Vander Meer, D., & Castillo, A. (2025). Large Language Models for Thematic Summarization in Qualitative Health Care Research: Comparative Analysis of Model and Human Performance. *JMIR AI, 4*, e64447.

Dall'Ora C, Saville C, Rubbo B, Turner L, Jones J, Griffiths P. (2022) Nurse staffing levels and patient outcomes: A systematic review of longitudinal studies. *Int J Nurs Stud., 134,* 104311.

Ernst AT, Jiang H, Krishnamoorthy M, Sier D (2004). Staff scheduling and rostering: A review of applications, methods and models. *Eur J Oper Res., 153(1),* 3-27.

Griffiths P, Recio-Saucedo A, Dall'Ora C, Briggs J, Maruotti A, Meredith P, et al. (2018). The association between nurse staffing and omissions in nursing care: A systematic review. *J Adv Nurs, 74(7*), 1474-87.

Griffiths P, Saville C, Ball JE, Chable R, Dimech A, Jones J, et al. (2020a). 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),* 1-162.

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

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

InnovaHealthTec (2023). Adult Inpatient Wards SNCT. Available from https://www.innovahealthtec.com/adult-in-patient-ward-snct.

Liu, A., & Sun, M. (2025). From voices to validity: Leveraging large language models (llms) for textual analysis of policy stakeholder interviews. *AERA Open, 11,* 23328584251374595.

Margetis G, Ntoa S, Antona M, Stephanidis C. (2021). *Human-centered design of artificial intelligence.* Handbook of human factors and ergonomics, pp.1085-106.

Militello, L. G., & Hutton, R. J. (1998). Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41(11),* 1618-1641.

Musy, S. N., Endrich, O., Leichtle, A. B., Griffiths, P., Nakas, C. T., & Simon, M. (2020). Longitudinal study of the variation in patient turnover and patient-to-nurse ratio: descriptive analysis of a Swiss University Hospital. *Journal of medical Internet research*, *22(4),* e15554.

Randerson, S., Graydon-Guy, T., Lin, E.-Y., & Casswell, S. (2025). Exploring the use of a large language model for inductive content analysis in a discourse network analysis study. *Social Science Computer Review*, 08944393251326175.

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*, 7-13.

Shin, S., Park, J. H., & Bae, S. H. (2018). Nurse staffing and nurse outcomes: A systematic review and meta-analysis. *Nursing outlook*, *66*(3), 273-282.

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