Title: Tools & systems to determine nurse staffing requirements: an evidence review

Working Paper: February 2019 (draft 0.1)

Author List: Peter Griffiths, Christina Saville on behalf of the Safer Nursing Care Tool Study Group

Health Work Research Group, University of Southampton

Key words:
- Decision Making, Organizational
- Hospitals
- Humans
- Models, Nursing*
- Models, Organizational
- Nursing Staff, Hospital/supply & distribution*
- Outcome and Process Assessment (Health Care)
- Personnel Staffing and Scheduling/classification*
- Personnel Staffing and Scheduling/organization & administration
- Practice Patterns, Nurses'/classification*
- Practice Patterns, Nurses'/organization & administration
- Quality Indicators, Health Care/organization & administration
- United Kingdom
- United Kingdom National Health Service
- Workforce

Introduction

Multiple reviews of research have established that higher registered nurse (RN) staffing levels in hospitals are associated with better patient outcomes and care quality, including lower risks of in-hospital mortality, shorter lengths of stay and fewer omissions of necessary nursing care (e.g.Griffiths et al., 2016, Griffiths et al., 2018, Kane et al., 2007, Shekelle, 2013). Newer studies exploring patient level longitudinal relationships and causal mechanisms provide increasing confidence that the observed relationships are, at least in part, causal (Ball et al., Griffiths et al., Griffiths et al., 2018, Needleman et al.).

However, beyond providing an injunction to invest in 'more' staff, such studies rarely provide a direct indication of how many staff are required in a given setting. Given the size of the nursing workforce and shortages in supply, the ability to properly determine an 'optimal' staffing level is an imperative from the perspective of both quality and efficiency of care. Identifying low staffing as a significant contributor to *"conditions of appalling care"*, a key recommendation of the Francis Inquiry into the failings of the Mid Staffordshire General Hospital (Mid Staffordshire NHS Foundation Trust) was the development of guidance for nurse staffing including:

"…evidence-based tools for establishing what each service is likely to require as a minimum in terms of staff numbers and skill mix."(p 1678)(Francis, 2013)

There is a long history of methods to determine appropriate nurse staffing levels; an early study published in 1922 looked at how much nursing work was done by students and graduate nurses in New York City hospitals (Lewinski-Corwin, 1922).

However, the determination of the required hours remains problematic. Commissioned to develop national safe staffing guidelines for the National Health Service in the UK following the Francis inquiries, the National Institute for Health and Care Excellence (NICE) concluded that:

"*There is a lack of research that assesses the effectiveness of using defined approaches or toolkits to determine nursing staff requirements and skill mix.*"(National Institute for Health and Care Excellence, 2014)

The evidence review undertaken by NICE was based on existing reviews and focussed on the impact of staffing tools and methodologies on outcomes and quality (Griffiths et al., 2014), but other reviews of broader scope over many years have reached similar conclusions.(Arthur and James, 1994, Aydelotte, 1973, DHSS Operational Research Service, 1982, Edwardson and Giovannetti, 1994, Fasoli and Haddock, 2010, Hurst, 2003) One early review, published in 1973, included a bibliography of over 1000 studies but found no evidence concerning the relative costs or effectiveness of different staffing methods and concluded that there was no substantive basis for choosing any one system over another, little evidence for validity beyond "face" validity and scant evidence even of reliability concluding *"Although the intent of the methodologies is admirable, all are weak"*(p 57) (Aydelotte, 1973) A review undertaken for the then Department of Health and Social Services (DHSS) in the UK in 1982 identified over 400 different systems for determining staffing requirements.(DHSS Operational Research Service, 1982)

Subsequent reviews have had to embrace an ever growing body of research, yet reach remarkably similar conclusions. Writing in 1994, Edwardson and Giovanetti noted the absence of published scientific evidence for a number of commercial systems, such as Grasp or Medicus, which were in widespread use in North America.(Edwardson and Giovannetti, 1994) They also noted that although different systems tended to produce results that were highly correlated they could, nonetheless, produce substantially different estimates of the required level of nursing staff for a given patient or unit. (Edwardson and Giovannetti, 1994) Edwardson & Giovanetti's review concluded by asking the following questions, which remained unanswered by the evidence they reviewed:

- *"Have the implementation of staffing methodologies affectively(sic) altered the level of staffing within facilities or have predetermined staffing levels altered the outcome of the workload measurement systems?*

- *What effect have workload measurement systems had on the quality of care?*

- *Do the results of a workload measurement system depart significantly from the professional judgment of practicing nurses?*

- *Have workload measurement systems resulted in an improvement in the utilization of nursing personnel?"* (Edwardson & Giovanetti p117)(Edwardson and Giovannetti, 1994)

More recently, in 2010 Fasoli and Haddock reviewed 63 sources (primary research, theoretical articles and reviews) and again concluded that there was insufficient evidence for the validity of many current systems for measuring nursing workload and staffing requirements, concluding that systems may not be sufficiently accurate for resource allocation or decision making. (Fasoli et al., 2011, Fasoli and Haddock, 2010) The field is dominated by descriptive reports of locally developed approaches with limited or weak evidence for validity even for widely used and commercially supported approaches. (Edwardson and Giovannetti, 1994, Fasoli and Haddock, 2010)

These reviews, and others, (Arthur and James, 1994, Aydelotte, 1973, Butler et al., 2011, DHSS Operational Research Service, 1982, Edwardson and Giovannetti, 1994, Fasoli et al., 2011, Fasoli and Haddock, 2010, Griffiths et al., 2016, Hurst, 2003, Twigg and Duffield, 2009) give a clear indication of vast literature on nurse staffing methodologies, but indicate that the field that has made little progress over many years with most research reporting on the underlying development of tools or assessing their implementation but little research validating the tools and none evaluating the impact of implementation on outcomes for staff or patients. However, the topic remains important with key questions unanswered. Given that the last review considered in the review of reviews by NICE was published in 2010, it is possible that the field has moved on, prompting the need to reassess the evidence on the what is known about nurse staffing methodologies, their development, validity & reliability, and impact of their use in terms of costs and effects on care quality and patient outcomes.

The makes a comprehensive review of the literature on this topic a somewhat daunting prospect. In this paper we aim to give a summary of the field, starting with a broad overview of evidence linking nurse staffing to patient outcomes. We follow that with an overview of approaches to measuring nurse staffing requirements, drawing primarily on existing reviews and recent primary research.

Review methods and scope

We developed a search strategy that was specific and reasonably sensitive using the results of our comprehensive review for NICE as a test set. We used the terms "Workload"[key word, MESH] or "Patient Classification"[key word] AND "Personnel Staffing and Scheduling" AND "Nurs*"[key word] or "Nursing"[Mesh] and limited using the OVID Medline sensitive limits for reviews, therapy, clinical prediction guides costs or economics. We also searched across Medline, CINAHL (key word only) and The Cochrane Library. We performed additional searches for citations to existing reviews and the works of their authors. We also undertook focussed searches on databases for works by key authors and on databases / the World Wide Web for specific terms associated with widely used tools. Searches were completed in mid-October 2018. To describe the field in general, we drew selectively on authoritative sources and existing reviews. We looked specifically for primary studies published from 2008 onwards and new reviews published after 2014. After removing duplicates, we had 392 sources to consider for this aspect of the review.

Selection

We took a liberal approach to inclusion. Firstly, we included primary studies that described the development, reliability or validity testing of systems or tools for measuring nursing workload or for predicting staffing requirements. Secondly, we included studies that compared the workload as assessed by different measures or which used a tool as part of a wider study in such a way that it might usefully give some insight into the validity of tools or some other aspect of the determination of nurse staffing requirements. Thirdly, we included studies that reported the costs and or consequences of using a tool, including the impact on patient outcomes. We only included studies that were of direct relevance to calculating the staffing requirement on general acute adult inpatient units and this excluded studies focussing on (for example) intensive or maternity care. However, had we identified material that illustrated a significant methodological advance or other insight we were open to including it for illustrative purposes.

Results
Nurse staffing levels & outcomes

Reviews of observational studies have consistently concluded that there is an association between higher registered nurse staffing levels and better patient outcomes for patients in acute hospitals.(Griffiths et al., 2016, Griffiths et al., 2014,

Griffiths et al., 2018, Kane et al., 2007, Lang et al., 2004, Lankshear et al., 2005, Shekelle, 2013) In 2007, a systematic review and meta-analysis identified 96 studies and concluded from a meta-analysis of 28 that increased RN staffing levels are associated with lower risks of in-hospital mortality (for ICU, surgical and medical patients), shorter lengths of stay (for ICU and surgical patients) and reductions in occurrences of various other adverse patient events (Kane et al., 2007). The majority of the papers included in this meta-analysis were from the USA, but the association between nurse-to-patient ratios and in-hospital mortality has subsequently been confirmed in several other countries and regions including many European countries in a large scale cross-sectional study of over 400 hospitals.(Aiken et al., 2014) Subsequent reviews have confirmed associations between nurse staffing and care quality across a range of outcomes and process measures, including omissions of essential nursing care, which is identified as a key mechanism leading to adverse patient outcomes.(Griffiths et al., 2018)

Because of different scope across different reviews it is now hard to judge the full extent of the evidence, but an estimate that the size of the evidence base had more than doubled since Kane's 2007 review, would likely be a highly conservative one. However, the evidence is, without exception, observational and, with few exceptions, cross-sectional, albeit based on large studies often involving hundreds of hospitals and millions of patients.(Griffiths et al., 2016)

Causality

While an overall association seems clear, establishing a causal relationship between increased nurse staffing levels and improved patient outcomes has been more challenging. In the abundant cross-sectional studies, it is conceivable that poor outcomes lead to a decision to increase staff levels, which may explain why associations with some of the most plausibly nurse sensitive outcomes such as pressure ulcers are least clearly supported.(Griffiths et al., 2016) It is also difficult to disentangle nurse-staffing effects from wider hospital resources, including medical staffing levels, although there are now a number of studies that demonstrate a nurse staffing effect controlling for levels of medical or other staff (Bond et al., 1999, Griffiths et al., 2016, Ozdemir et al., Ozdemir et al.) or are otherwise largely immune from such confounding.(Griffiths et al., 2018, Griffiths et al., in press, Needleman et al., 2011)

Kane et al.'s meta-analysis provides evidence for causality assessed against the so called 'Bradford Hill' criteria (Hill, 1965) by firstly showing the consistency of results from multiple studies, secondly showing that results are not sensitive to the study designs used, and thirdly providing evidence of a dose-response relationship (2007) More recent studies provide evidence for plausible causal mechanisms related to omissions of care, which have been show to partially mediate the association between staffing levels and outcomes for both mortality and patient satisfaction. (Ball et al., 2018, Bruyneel et al., 2015, Griffiths et al., 2018, Griffiths et al., 2018, Recio-Saucedo et al., 2018) Recent studies have used a longitudinal rather than cross-sectional design (Griffiths et al., 2018, Griffiths et al., in press, Needleman et al., 2011), thus showing that the association is operating at the level of individual patient

exposures to low nurse staffing on wards and establishing the temporal order of events.

Cost-effectiveness & efficiency

Despite the plentiful evidence of the link between increased nurse staffing levels and improved patient outcomes, the NHS advisory body NICE found it insufficient to translate into specific staffing recommendations. Since NICE's framework for evaluating service changes is to measure the potential cost-benefit, for example cost per life saved, reliable estimates of these measures are needed before evidence-based recommendations can be made. Furthermore, while the existing evidence might inform questions about the cost-effectiveness of change, it does little to identify either baseline staffing or desired endpoints. Existing international economic modelling studies have produced widely disparate estimates of cost effectiveness. (Griffiths et al., Twigg et al.) Some studies suggest that increases in registered nurse staffing might fall within accepted cost effectiveness thresholds and increases in skill mix could prove cost neutral, but there is substantial uncertainty associated with the finding.(Griffiths et al., 2018)

However, the problem of how best to determine the required staffing levels remains. Policies of mandatory staffing minimums have been much discussed and implemented in a number of jurisdictions, most notably California, USA (Donaldson and Shapiro, 2010, Mark et al., 2013, Royal College of Nursing, 2012). Such policies remain controversial and evidence of the benefits associated with them remains elusive, and is, in any case, difficult to disentangle from the effects of staffing increases (Donaldson and Shapiro, 2010, Griffiths et al.). Furthermore, these policies have been criticised as providing an inflexible system that is potentially inefficient (Buchan, 2005). Even when mandatory minimum policies are implemented, there remains a requirement that patient care needs over and above the minimum are assessed and staffing adjusted accordingly.

Overview of approaches to setting nurse staffing

A wide variety of approaches have been described, although many approaches share similar characteristics. Telford's *professional judgement* method (Telford, 1979), which was first formally described in the UK in the 1970s and documented by Hurst (Hurst, 2003), provides a way of converting the shift level staffing plan, decided using expert opinion, into the number of staff to employ. As described by Hurst, the decision on the required staffing is based solely on professional judgement and so the bulk of the method simply deals with the calculation of the number of nurses to *employ* in order to fill the roster, after making allowance for holidays, study leave and sickness/absence. Conversely, the method described can be used to infer the daily staffing plan from the whole time equivalent staff employed by the ward and Hurst provides an illustration (Hurst, 2003). The full method provides a framework for wider deliberation and consideration but at the core the judgement of required staffing is not supported by any direct calculation or formal use of an objective measure to determine need (Arthur and James, 1994). In recent years this deliberative approach without formal measures appears to be reflected in

the United States Veteran's Administration staffing methodology (Taylor et al., 2015). Other approaches that could be characterised under this method include "benchmarking" approaches where expert judgements are made about suitable comparators and the staffing levels and achieved quality are compared between similar units.

Benchmarking exercises are often used to arrive at a *nurse-patient ratio* or equivalent (such as nursing hours per patient) to assign a minimum or fixed number of nursing staff per occupied bed (Hurst, 2003). While characterised by Hurst as a distinct method, there is little initial formal assessment of patient requirements for nursing care, although consensus methods and expert judgement are often used in determining the staffing requirement for particular types of wards and selecting appropriate benchmarks. Effectively, these approaches assume that all patients in similar wards have similar requirements for care, or that the average is stable across patient groups, so needs can largely be anticipated and met with a set roster. Thus these approaches tend to set a minimum staffing level, with an explicit or implied expectation that any additional staffing requirements be determined when demand increases. While the benchmarking comparison with similar wards gives the appearance of objectivity, much depends on how the initial staffing levels were arrived at and there is ample evidence that perceptions of staffing requirements are largely anchored to historical staffing levels.(RN4CAST survey, Twigg and Duffield, 2009) Because some more recent approaches to monitoring workload (see below) extend this approach to take a wider view of activity, adding in admissions and discharge over and above the patient census, we will term these benchmarking and ratio approaches *'patient volume'* approaches.

Approaches which seek to account for individual patient variation in need can be used as alternatives to, or in conjunction with, minimum staffing levels. Hurst describes three main types: patient classification systems, timed-task approach and ?: (Fasoli and Haddock, 2010) *Patient classification systems* group patients according to their nursing care needs and assign a required staffing level for each (Fasoli and Haddock, 2010, Hurst, 2003). These use either pre-existing categorisations, e.g. diagnosis-related groups (Fasoli and Haddock, 2010), or bespoke categorisations, e.g. classifications based on levels of acuity and / or dependency groups. The Safer Nursing Care Tool, which is the most widely used method for determining staffing requirements in the UK is one such system, allocating patients to one of five acuity / dependency categories with a weighting (described as a 'multiplier') to indicate the required staff to employ associated with patients in each category (The Shelford group, 2014). In the *timed-task approach*, a detailed care plan, consisting of tasks and activities that are assigned a recommended length, is constructed for each new patient and used to determine the required staffing (Hurst, 2003). The commercial GRASP system, widely used in the United States is an example of such a system (Edwardson and Giovannetti, 1994). In all cases, the total staffing required is based on the sum of individual patient requirements across all patients.

The approach to determining the required times for patient groups, tasks and activities varies. The literature describes the use of both empirical observations and

expert opinion to determine the average time associated with activities or patient classifications.(De Cordova et al., 2010, Myny et al., 2014, Myny et al., 2010) In some cases there is an explicit attempt to make workload / time allocations based on reaching some threshold of quality or optimum care.(Fagerstrom and Rainio, 1999, Fagerström and Rauhala, 2007) Approaches to non-patient contact time and activities vary but all approaches take this into account, often assigning a fixed percentage of time over and above that measured for patient contact.

Hurst also identified *regression*-based approaches, which model the relationship between patient, ward and hospital-related variables, and the establishment, in adequately-staffed wards.(Hurst, 2003) To obtain the recommended establishment for a particular ward, coefficients derived from the regression models are used to estimate the required staffing based on a number of factors. Perhaps due to the obvious complexity, there are relatively few examples, although Hoi and colleagues provide one recent example, the Workload Intensity Measurement System. (Hoi et al., 2010) In some respects the regression based models simply represent another approach to allocating time across a number of factors rather than directly observing time linked to specific activities. The Rafaela system, widely used in the Nordic Countries, although based on a relatively simple patient classification, also uses a regression based approach to determine the staffing required to deliver an acceptable intensity of nursing work for a given set of patients in a given setting (FagerstrÖM and Rainio, 1999, Fagerstrom and Rauhala, 2007, Rauhala and Fagerström, 2004).

Edwardson and Giovannetti offer a slightly different typology, differentiating '*prototype approaches*' from '*indicator approaches*' and *'task approaches'* (Edwardson and Giovannetti, 1994). Prototype and task approaches broadly resemble the patient classification and task approaches (respectively) previously described. Indicator approaches ultimately assign patients to categories but it is done based upon ratings across a number of factors that are related to the time required to deliver patient care. These can include broad assessments of condition (e.g. 'unstable'), states (e.g. 'non ambulatory'), specific activities (e.g. complex dressings) or needs (e.g. for emotional support or education).(Edwardson and Giovannetti, 1994) The Oulu Patient Classification, part of the Rafaela system, is one such example, where patients are assigned to one of 4 classifications, representing different amounts of care required, based upon a weighted rating of care needs across six dimensions.(FagerstrÖM and Rainio, 1999) However, the inclusion of some specific activities in Edwardson and Giovennetti's definition of indicator approaches makes it clear that the distinction from task / activity based systems is not an absolute one. Typically, though, task based systems take many more elements into account: over 200 in some cases. (Edwardson and Giovannetti, 1994) We will adopt this terminology when characterising systems described in recent literature below)

While these classifications are useful to distinguish broad approaches, caution needs to be exercised. As an example, professional judgement based approaches might involve benchmarking to set a fixed establishment for a ward based on an underlying staffing model that aims for a given patient to nurse ratio on each shift (and so might

be classified as a volume based approach). However, the original determination of that requirement might have involved a detailed appraisal of patient need on a given ward involving many factors, but no formal measurement. On the other hand, while a patient classification system may set an establishment based on a measurement of individual patient needs, without the application of further professional judgement the establishment, once set, implies that care needs are then met by a fixed nurse to bed ratio regardless of daily variation. Indeed, a simple prototype classification system, such as the Safer Nursing Care Tool, resembles a volume based mandatory minimum staffing policy supplemented by assessment of variation above the base requirement, such as that implemented in California, because there is an implied absolute minimum staffing level per patient, associated with the lowest weighting.

Staffing decisions and the use of tools

There are a number of different decisions that can be made using staffing systems. Nursing managers must decide how many nursing staff to *employ* (often referred to as the *nursing establishment*) and how many nursing staff to *deploy* each shift. These are separate but inter-related decisions, which rely on being able to quantify nursing *workload.* For example, the Safer Nursing Care Tool (SNCT), was designed to support decisions about the total nursing establishment required on a ward (The Shelford group, 2014). More recently, its core acuity-dependency scoring system has been used to plan and review daily staffing levels, supporting deployment and real-time redeployment decisions, for example using the SafeCare system from the commercial rostering system provider Allocate (Allocate software, 2017). There are also examples of tools specifically to balance workload within a unit, and are thus focussed primarily on immediate assignments for staff members (Brennan and Daly, 2015, Brennan et al., 2012). Finally, tools can be used retrospectively to review the success of staffing plans – how well the plan met needs - or as a measure of resource use for pricing, budgeting or billing purposes (Kolakowski, 2016). (Table 1) The distinction is often unclear in much of the published literature and the relationship between these uses is implicit rather than explicit.

Table 1 Uses of staffing systems and tools

Prospective	Concurrent	Retrospective
• Establishment setting: employment and base deployment decisions (long term). • Predict immediate future demand (e.g. next shift)	• Determine current staffing adequacy and guide deployment / redeployment • Prioritise and allocate work to a team	• Review success of staffing plans • Billing & resource use

Evidence

The numerous reviews cited at the beginning of this paper made it clear that there was little basis to prefer any one tool or approach over another based on the available evidence. Professional judgement based approaches, despite being open to accusations of subjectivity, cannot be readily dismissed, since there is no substantial evidence that moving from a judgment based staffing model to one informed by a tool has improved any outcomes or made more efficient staffing allocations (Arthur and James, 1994, Aydelotte, 1973, DHSS Operational Research Service, 1982, Fasoli and Haddock, 2010, Griffiths et al., 2016, Hurst, 2003, Twigg and Duffield, 2009). One of the most comprehensively researched systems determines the staffing requirement by titration against a subjective report of work intensity, (FagerstrÖM and Rainio, 1999, Rauhala and Fagerström, 2004). This might not matter if different approaches gave similar results. While direct comparisons are few and far between it is however clear that different systems can give vastly different estimates of required staffing (Jenkins-Clarke, 1992, O'Brien-Pallas et al., 1991, O'Brien-Pallas et al., 1992, O'Brien-Pallas et al., 1989). In one study, the 5 systems tested provided estimates that correlated highly. However, they offered a wide range of average staffing requirements for the same sample of 256 patients, with the lowest average less than 60% of the highest (6.65hours per patient per day vs 11.18).(O'Brien-Pallas et al., 1992)

From our searches for more recent reviews (from 2013 on) and primary studies (from 2008 onwards) we found no new reviews and 36 primary sources to consider (see supplemental material for fuller descriptions). We classified the sources according to the main purposes of the articles. Some articles did not clearly sit in one or another category and were given a dual classification.

Six sources simply described the use of a staffing system or tool (classified as 'description') although they also reported some data, generally exemplar graphs or charts of varying workload (Fagerström et al., Fenton and Casey, Kolakowski, Smith et al., Taylor et al., The Shelford group). Additionally, three studies were classified as 'descriptive', where the focus of the study was to measure nursing workload or demand associated with different ward designs (Hurst), diagnostic groups (Hurst) or to determine variability in patient need prior to developing a new workload

management system (Gabbay and Bukchin). These studies provide ample reminder that however measured there are many sources of variation in the requirement for nursing care on a ward. These arise from the number of patients and the numbers of admissions and discharges, individual patient characteristics and their specific needs, and contextual factors such as the physical arrangement of the ward. These needs vary considerably between units and from day to day, although none of these studies provided a measure that allowed the variation to be directly quantified in terms of staff required from day to day.

Four sources compared workload as assessed by different approaches ('comparison') although of these 2/4 simply compared patient volume based approaches (different approaches to calculating hours per patient day, inclusion of turnover) (Beswick et al., Simon et al.) while one compared an acuity dependency measure with a purely volume based approach (Rivera) and one compared a newly developed indicator system with the results of a pre-existing system (Hoi et al.). Again these studies emphasise the lessons determined from previous research – different methods arrive at different results and methods that take into account more factors appear to arrive at higher workloads.

Although differences between alternative approaches to counting patients in methodologies using hours per patient day seem to be of marginal overall significance (Beswick et al., Simon et al.), other factors can make a great difference to estimated workload. An unquantified statistically significant increase to workload from including patient turnover in a volume based measure was noted in one study,(Beswick et al., 2010) an acuity dependency system identified an additional 6 hours of care per day compared to a standard hours per patient day method in another, (Hoi et al.) and a multifactorial indicator system led to an estimated patient to nurse ratio that was half of that determined by an existing system (Hoi et al., 2010).

Perhaps most striking is that the second largest group of studies (thirteen) were those described as 'development', reporting on the full or partial development of a new measure or adaptation of an existing measure, often for local use only, although three papers related to a single programme of work (Baernholdt et al., Brennan et al., de Cordova et al., Ferguson-Paré and Bandurchin, 2010, Gabbay and Bukchin, Hoi et al., Hurst et al., 2008, Larson et al., Morales-Asencio et al., 2015, Myny et al., Myny et al., Myny et al., 2012, Perroca). Rather than going into detail on the specifics of each of these tools, it is perhaps more instructive to note that all the system types noted earlier, from professional judgement and volume based approaches through to timed task, have appeared in these reports on developments since 2008.

A number of general observations can be made. These reports typically identify time or some weighting associated with aspects of care or particular groups of patients 'on average'. However, they rarely report or consider variability in the underlying estimates. A rare exception to this is the work of Mynwy and colleagues in Belgium. (Myny et al., 2014, Myny et al., 2010). Although the reports were focussed on demonstrating the precision of the standard time estimates they derived, the degree

of variation associated with a particular task is well illustrated. The estimated standard time for "partial help with hygienic care in bed" had a 95% confidence interval from 7.6 to 21.2 minutes. That for "settling a bed ridden patient" was from 5 to 25.75 minutes.(Myny et al., 2010) The underlying sample of observations could not be easily determined, but given that the study as a whole observed over 13,000 care activities, suggesting the wide confidence intervals may arise from intrinsic variability rather than simply small samples.

It may be that prototype approaches, where measures are based on typical care needs of patients fitting a particular profile, are less subject to variation between individuals with the same classification but we found no equivalent estimates of variation for such systems. One reason that such measures of variability rarely appear may be that despite the external appearance of 'objectivity' the times or weights assigned within the system are often wholly or partly arrived at through an expert consensus exercise, for example (Brennan et al., 2012, Hurst et al., 2008, Myny et al., 2010). It is clear that professional judgement remains an important source of information and validation for any system.

Evaluation

Many of the sources classified as 'development' were also classified as 'evaluation'. In total 16 sources were classified as evaluation based upon reporting attempts to assess the reliability or validity of a measure or assessment of usability or user experience of the system. (See *supplemental material* for a full listing of studies).

Three studies demonstrated that staffing below the level recommended by staffing tools was associated with adverse patient outcomes. A US study using an unspecified commercial patient classification system found that the hazard of death was increased by 2% on every occasion a patient was exposed to a shift with 8 or more hours below the target defined by the system. (Needleman et al., 2011) Mortality was also increased by exposure to shifts with unusually high patient turnover, suggesting that this might be generating additional workload that was not measured by the system. Two studies used the 'optimal' staffing level defined by the Oulu Patient Classification (OPC), part of the Rafaela System, as a reference (Fagerstrom et al., 2018, Junttila et al., 2016). These studies both found that nursing workload above the defined 'optimal' level was associated with adverse patient outcomes, including increased mortality. However, nursing workload *below* the optimal level (higher staffing) were associated with further improvements in outcomes (Fagerstrom et al., 2018, Junttila et al., 2016). Furthermore, the OPC workload measure was not clearly superior to a simple patient per nurse measure based on analysis of decision curves. (Fagerstrom et al., 2018)

More recently, a UK study has shown that Registered Nurse staffing below the planned level, based on establishments determined using the Safer Nursing Care Tool was associated with a 9% increase in the hazard of death in one English Hospital Trust, although low assistant staffing according to this criterion was not associated with mortality increases (Griffiths et al., 2018). This study also explored

staffing level as a continuous variable and found that the relationship between mortality and registered nurse staffing levels appeared to be linear.

We found no evidence of the impact of implementing a staffing tool or system on costs, patient outcomes, quality of care or staff, other than the implementation of a 'Nursing Hours per Patient Day' methodology in Australia. This methodology assigns a minimum staffing level (measured in hours per patient day) for six different ward types, based on the patient case mix and complexity. There were improvements over time with several patient outcomes, including mortality, although it is unclear if this was specifically associated with the system used or the resulting increases in staffing levels (Twigg et al., 2011).

Consequently, while these studies give some indication of the validity of the tools used as workload measures they do not provide any clear evidence of the effectiveness of staffing using the tool as opposed to a simpler volume based system, nor do they show that the tools give an "optimal" staffing level. In effect the results show that increases in staffing result in improved outcomes. The tools may motivate increases in staffing although results from using the Rafaela system suggest that they could also restrict staffing at a level, which is not clearly 'optimal'.

Other studies demonstrate that staffing or workload measured by the Rafaela system correspond with some other measure of workload or measure of acuity / dependency (for example, Brennan et al., 2012, Hurst et al., 2008, Larson et al., 2017, Morales-Asencio et al., 2015, Smith et al., 2009). In all but one of these examples, the criterion is, in effect, a professional judgement of some measure of demand for nursing care. Additionally the determination of the required staffing level in the Rafaela system is based on identifying a level of OPC weighting that is associated with nurses' reports that staffing is 'optimal' (Fagerström et al., 2014).

Reliability studies show that nurses can make reliable assessments using different classification systems (Brennan et al., 2012, Liljamo et al., 2017, Perroca, 2013) although achieving inter-rater agreement is not always straightforward and the reliability of ratings in a new setting should not be assumed. (van Oostveen et al.) Van Oostveen's pre-implementation study for the Rafaela system demonstrates that successful implementation of any system requires significant investment to engage and train staff. Even in the face of broad support there was insufficient engagement with the measures of staffing adequacy required by the system and satisfactory reliability also proved hard to achieve (van Oostveen et al.). Brennan also identified challenges in achieving reliable assessment and consensus on the validity of some important items of her indicator measure. (Brennan and Daly, 2015)

Taylor and colleagues describe the substantial challenges faced in implementing a professional judgement based system for the US Veteran's administration.(Taylor et al., 2015) While concluding that their system can be successfully implemented they highlighted nursing leadership and front line staff buy-in as essential. They also emphasised the importance of staff training and the risk of cynicism if staff invest effort in a new system but see little tangible outcome.

Operational research

We identified four studies that emanated from the tradition of operational research methods (Davis et al., Harper et al., Kortbeek et al., Maenhout and Vanhoucke). While these papers are only examples of a much larger body of literature that focuses on nurse rostering as opposed to developing new workload measurement tools, they do serve to highlight one important issue. Determining the average staffing requirement does not necessarily provide a solution to the base staffing required in order to meet varying patient need. Two studies, one using an acuity dependency measure 'informed' by the work of the developers of the Safer Nursing Care Tool determined that optimal staffing in the face of varying patient demand was higher than a level determined simply by staffing to meet the mean demand (Davis et al., Harper et al.). In one case the apparent 'overstaffing' was associated with net cost savings in modelling / simulations, in part because of the potential value of 'excess' staff who were available for redeployment to understaffed units (Davis et al.). Other studies modelled the effects of the use of varying configurations of 'float' pools to meet fluctuation in demand (Kortbeek et al., Maenhout and Vanhoucke). While the purpose of these studies seems to be primarily to generate 'local' solutions, rather than provide generalizable answers, they do indicate that even where a tool correctly determines average staffing requirements that does not, in itself, necessarily provide the optimal answer to the question of how many staff should be employed or routinely deployed on a ward.

Discussion

While recent years see a continued interest and a significant number of publications relating to systems and tools to determine nurse staffing requirements, the unanswered questions identified by Edwardson & Giovanetti in 1994 (Edwardson and Giovannetti, 1994) remain largely unanswered. The finding that outcomes improve as staffing levels rise above the 'optimal' determined by the Rafaela system (Fagerstrom et al., 2018) is consistent with perceptions of adequate staffing being anchored by expectations set by current staffing levels, a phenomenon that has been widely noted and in turn is cited as a rationale for the use of formal systems.(Taylor et al., 2015, Telford, 1979) It may well be that predetermined or historical staffing levels do alter the outcome of the workload measurement systems, despite the apparent objectivity. Although there are perceptions of benefits, the effect of workload measurement systems on the quality of care remains unclear. In general though, the results of a workload measurement system seem to correspond closely with professional judgement although the magnitude and significance of departures is hard to quantify in terms of staffing levels since it is generally not reported. However, different systems can give dramatically different results and so it is clear that there can be no one answer to the questions of whether workload measurement systems result in an improvement in the utilization of nursing personnel.

Sources of variation

The methods described generally match staffing levels to the average (mean) demand associated with a particular patient group, factor or aspect of care when attempting to estimate current or future staffing requirements. The operational research studies make it clear that in the face of variable demand, simplistic responses based on the average may not be the best way to use the results of measurement systems. Indeed, it is ironic that while much of the literature is concerned with measurement and identification of sources of variation, it is poor at quantifying such variation in a way that allows assessment of decisions based on the mean. Reports rarely provide estimates of standard deviations around the mean. One of the few recent studies that gave any estimate of variability gave confidence intervals with up to five fold variation from upper to lower bounds.(Myny et al., 2014) Since confidence intervals are based on the sampling distribution, time taken to deliver care to an individual patient will vary more than this.

In Table 2 we summarise some major sources of variation. While different systems account for these factors to a varying degrees they rarely consider variation around the averages they use to determine staffing requirements. Averaging can be compounded at multiple levels, for example average patient census, average need per patient and average time taken to deliver care for a patient with a given set of needs.

Table 2 Sources of variation

Individual patient	Ward	Staff
Differing care needs • *Different patients have different need, even within the same prototype* • *Variability unknown*	Patient census / occupancy • *Variation between and within days, known to be right skewed*	Sickness / absence • *Relatively rare occurrence with non-random clustering and seasonal variation*
Varying time to deliver care • *Different lengths of time to undertake the same aspect of care* • *Limited evidence of variability but can be considerable*	Turnover (admission / discharge) • *Considerable variation, likely to be left skewed.*	Staff leave (holiday and study) • *Predictable but with predictable / controllable seasonal variation*
	Ward layout • *Ward layout and other factors potentially systematically alter the time required for certain aspects of care*	Vacancies • *Unpredictable with non-random clustering*

When workload distributions are approximately normal and have small standard deviations, using the mean may be an appropriate basis for planning, as the workload will vary from the mean by a relatively small and known amount. For a normal distribution, 95% of values are less that two standard deviations of the mean. Assuming some degree of flexibility in the work capacity of a given group of staff, most patients' needs might be safely accommodated most of the time. However, while some systems are explicit about an acceptable degree of variation from the mean (e.g. the RAFAELA system)(Fagerström et al., 2014), this is rare, and the impact on safety of even small deviations has not been widely researched.

However, while little data is available to assess the extent of variation, substantial variability and highly skewed distributions seem plausible. Certainly right-skewed ward occupancy distributions have been reported (Davis et al., 2014) which would tend to generate mean staffing requirements that are lower than the median, leading to relative understaffing more than 50% of the time. Additionally, when planning workload in advance it is clear that some sources of variation do not occur at random. As an example, in order to ensure that there are sufficient staff available to provide cover on wards the literature describes the need to add an "uplift" to establishments to allow for staff sickness.(Hurst, 2003, Telford, 1979) However, staff sickness does not occur uniformly and tends to occur in clusters, with clear seasonal patterns and variation by day of the week. (Barham and Begum, 2005)

It appears unlikely that any multifactorial system can fully encompass all aspects of the work associated with patients or the variation in time taken to complete tasks. Prototype systems may bypass this limitation because they are not reliant on estimates of specific care activities, but this remains untested. However, such systems cannot account for variation associated in activities that are not directly linked to the patient prototype and must instead rely on allowing some form of 'overhead' to account for such activities. In systems based on observing care delivered directly to patients a large proportion of nursing time is unaccounted for (typically 50% or more), associated with indirect as opposed to direct care. This time is simply added as an overhead (Gabbay and Bukchin, 2009). Factors such as patient turnover, which clearly has significant potential to generate nursing work, as demonstrated by studies showing that volume measures of nursing workload incorporating this factor are strongly associated with patient outcomes. (Griffiths et al., 2018, Griffiths et al., in press, Needleman et al., 2011) are not easy to account for in a patient prototype because they occur at a fixed point in time for a patient.

Patient turnover is highly variable between and within wards, with some predictable sources of variation (such as day of the week) (Griffiths et al., 2018). More complex indicator based systems are better able to take this into account. For example, the RAFAELA system assesses the "planning and coordination of nursing care " required for each patient (Fagerström et al., 2014), but again a mean may not properly reflect varying requirements. In this case, it seems likely that the distribution for most wards is left skewed with the mean higher than the median.

Only systems which are based on local assessments of time required can take into account additional variation that arises from factors such as ward layout. Evidence that ward layout may alter staff requirements is limited (Hurst, 2008) although simple variation in travel distances and opportunity for patient surveillance are recognised as having the potential to generate considerable variation in workload associated with different ward configurations (Maben et al., 2016, Maben et al., 2015). While this source of variation can be accommodated in the means estimated for the ward, this does raise a final issue associated with the lack of quantification of variation in the literature. Even where a mean adequately allows staff to meet variable demand, it is unclear how much care needs to be observed to establish a reliable mean.

Skill mix

The proportion of the total nursing workforce that are registered nurses is referred to as *skill mix*. Given the significant evidence base that emphasises the specific association between registered nurse staffing levels or skill mix and outcomes (e.g. Aiken et al., 2017) it is perhaps surprising that skill mix is rarely addressed directly by traditional staffing methods (Buchan et al., 2001). This may be because many systems and approaches have their origins in settings where the contribution of support staff to direct patient care was relatively slight (e.g. the USA Aiken et al., 2017). The issue of determining skill mix is compounded by the fact that unregistered nurses' roles can vary widely (Kessler et al., 2010). Some tools consider only registered or licensed nurses while others, such as the SNCT (The Shelford group, 2014), plan the total nursing team size and defer the skill mix decision to professional judgement.

"Optimal" staffing

Each staffing method makes an underlying assumption about what constitutes "adequate", "safe" or "quality" staffing although these are often implicit. The staffing to deliver the "right" frequency and length of nursing tasks in the timed-task approach, and the "right" amount of care per patient in the nurse-patient ratio approach must be decided upon. These parameters are obtained either from expert judgement or from observations of care provided or from existing establishments, ideally in settings deemed to meet some quality criteria (Hurst, 2003). The quality threshold is not always made explicit.

There is evidence that staffing to the 'optimal' level defined by the RAFAELA tool is associated with reduced mortality when compared to lower staffing (Junttila et al., 2016). However, mortality is further reduced by staffing at higher levels than the tool suggests, which raises questions about how 'optimal' staffing is defined here. Another study using the same tool suggests that there is a marginal improvement in the ability to predict outcomes when using mismatch between actual staffing levels and the assessed staffing requirement compared to staffing measures based on simple nurse to patient ratios but the improved fit is small and of unclear significance in terms of the impact on decision-making. (Fagerstrom et al., 2018)

Conclusions

The volume of literature on staffing methodologies is vast and growing, suggesting continued interest. However, the conclusions of previous reviews, that there is no substantial evidence base on which to base staffing decisions or to select a decision support tool remains true. It also seems clear that there has been a repeated pattern whereby new tools continue to be developed with little programmatic research addressed at existing tools, even when they are widely used. The extensive research reporting the development of the Rafaela system stands out as an honourable exception in this regard although neither costs nor effects of using the tool have been reported either compared to another tool or no tool at all.

Despite the lack of evidence, an appetite for formal systems and tools exists. While professional judgement remains the nearest to a gold standard, the desire to use a tool or other formal system to support and indeed justify such a judgement has remained a constant theme which can be traced back to Telford's work in the 1970s in the UK, and no doubt beyond. The lack of evidence for existing tools seems to have propelled each generation (and indeed in some case each service) to develop a new tool. On the other hand the widespread adoption of particular tools, such as the Safer Nursing Care Tool in the UK, might be indicative of a perceived validity.

Rather than continually invent new tools it seems time to take a closer look at those already in use. Controlled trials may be problematic but there are sufficient unanswered questions that much progress can be made before determining whether such a trial is strictly necessary. There is little evidence about many of the assumptions underlying the use of tools in general. Simple questions that remain unanswered include: the extent to which the tools truly identified a level of staffing sufficient to meet the needs of a ward of patients, and the number of observations required to get an accurate baseline to estimate average need. The apparently simple assumption, that staffing to meet average need is the optimal response to varying demand is also untested empirically, although research reviewed here suggests it is likely to be incorrect.

References

Aiken, L.H., Sloane, D., Griffiths, P., Rafferty, A.M., Bruyneel, L., McHugh, M., Maier, C.B., Moreno-Casbas, T., Ball, J.E., Ausserhofer, D., Sermeus, W., Consortium, R.C., 2017. Nursing skill mix in European hospitals: cross-sectional study of the association with mortality, patient ratings, and quality of care. BMJ Qual Saf 26 (7), 559-568.

Aiken, L.H., Sloane, D.M., Bruyneel, L., Van den Heede, K., Griffiths, P., Busse, R., Diomidous, M., Kinnunen, J., Kózka, M., Lesaffre, E., McHugh, M.D., Moreno-Casbas, M.T., Rafferty, A.M., Schwendimann, R., Scott, P.A., Tishelman, C., van Achterberg, T., Sermeus, W., 2014. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. The Lancet 383 (9931), 1824-1830.

Allocate software, 2017. SafeCare.

Arthur, T., James, N., 1994. Determining nurse staffing levels: a critical review of the literature. Journal of Advanced Nursing 19 (3), 558-565.

Aydelotte, M.K., 1973. Nurse Staffing Methodology, A Review and Critique of Selected Literature. Public Health Service (DHEW) , Washington, D.C. Div. of Nursing., Washington.

Baernholdt, M., Cox, K., Scully, K., 2010. Using clinical data to capture nurse workload: implications for staffing and safety. CIN: Computers, Informatics, Nursing 28 (4), 229-234.

Ball, J.E., Bruyneel, L., Aiken, L.H., Sermeus, W., Sloane, D.M., Rafferty, A.M., Lindqvist, R., Tishelman, C., Griffiths, P., Consortium, R.N.C., 2018. Post-operative mortality, missed care and nurse staffing in nine countries: A cross-sectional study. Int J Nurs Stud 78, 10-15.

Barham, C., Begum, N., 2005. Sickness absence from work in the UK. Labour market trends 113 (4), 149-158.

Beswick, S., Hill, P.D., Anderson, M.A., 2010. Comparison of nurse workload approaches. J Nurs Manag 18 (5), 592-598.

Bond, C.A., Raehl, C.L., Pitterle, M.E., Franke, T., 1999. Health care professional staffing, hospital characteristics, and hospital mortality rates. Pharmacotherapy 19 (2), 130-138.

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.

Brennan, C.W., Daly, B.J., Dawson, N.V., Higgins, P.A., Jones, K.R., Madigan, E., Van Der Meulen, J., 2012. The oncology acuity tool: a reliable, valid method for measuring patient acuity for nurse assignment decisions. Journal of nursing measurement 20 (3), 155.

Bruyneel, L., Li, B., Ausserhofer, D., Lesaffre, E., Dumitrescu, I., Smith, H.L., Sloane, D.M., Aiken, L.H., Sermeus, W., 2015. Organization of hospital nursing, provision of nursing care, and patient experiences with care in Europe. Medical Care Research and Review 72 (6), 643-664.

Buchan, J., 2005. A certain ratio? The policy implications of minimum staffing ratios in nursing. J Health Serv Res Policy 10 (4), 239-244.

Buchan, J., Ball, J., O'May, F., 2001. If changing skill mix is the answer, what is the question? Journal of Health Services Research & Policy 6 (4), 233-238.

Butler, M., Collins, R., Drennan, J., Halligan, P., O'Mathúna, D.P., Schultz, T.J., Sheridan, A., Vilis, E., 2011. Hospital nurse staffing models and patient and staff-related outcomes. Cochrane Database Syst Rev 7.

Davis, A., Mehrotra, S., Holl, J., Daskin, M.S., 2014. Nurse Staffing under Demand Uncertainty to Reduce Costs and Enhance Patient Safety. Asia-Pacific Journal of Operational Research 31 (01), 1450005.

De Cordova, P.B., Lucero, R.J., Hyun, S., Quinlan, P., Price, K., Stone, P.W., 2010. Using the nursing interventions classification as a potential measure of nurse workload. Journal of nursing care quality 25 (1), 39.

de Cordova, P.B., Lucero, R.J., Hyun, S., Quinlan, P., Price, K., Stone, P.W., 2010. Using the Nursing Interventions Classification as a Potential Measure of Nurse Workload. Journal of nursing care quality 25 (1), 39-45.

DHSS Operational Research Service, 1982. Nurse Manpower Planning: Approaches and Techniques. DHSS, London:.

Donaldson, N., Shapiro, S., 2010. Impact of California mandated acute care hospital nurse staffing ratios: A literature synthesis. Policy, Politics, & Nursing Practice 11 (3), 184-201.

Edwardson, S.R., Giovannetti, P.B., 1994. Nursing Workload Measurement Systems. Annual Review of Nursing Research 12 (1), 95-123.

Fagerstrom, L., Kinnunen, M., Saarela, J., 2018. Nursing workload, patient safety incidents and mortality: an observational study from Finland. BMJ Open 8 (4), e016367.

Fagerström, L., Lønning, K., Andersen, M.H., 2014. The RAFAELA system: a workforce planning tool for nurse staffing and human resource management: Lisbeth Fagerström and colleagues describe a method pioneered in Finnish hospitals that aims to uphold staffing levels in accordance with patients' care needs. Nursing Management 21 (2), 30-36.

FagerstrÖM, L., Rainio, A.-K., 1999. Professional assessment of optimal nursing care intensity level: a new method of assessing personnel resources for nursing care. Journal of Clinical Nursing 8 (4), 369-379.

Fagerstrom, L., Rainio, A.K., 1999. Professional assessment of optimal nursing care intensity level: a new method of assessing personnel resources for nursing care. Journal of clinical nursing 8 (4), 369-379.

Fagerstrom, L., Rauhala, A., 2007. Benchmarking in nursing care by the RAFAELA patient classification system - a possibility for nurse managers. J Nurs Manag 15 (7), 683-692.

Fagerstrom, L., Rauhala, A., 2007. Benchmarking in nursing care by the RAFAELA patient classification system - a possibility for nurse managers. Journal of nursing management 15 (7), 683-692.

Fasoli, D.R., Fincke, B.G., Haddock, K.S., 2011. Going Beyond Patient Classification Systems to Create an Evidence-Based Staffing Methodology. Journal of Nursing Administration 41 (10), 434-439.

Fasoli, D.R., Haddock, K.S., 2010. Results of an integrative review of patient classification systems. Annu Rev Nurs Res 28, 295-316.

Fenton, K., Casey, A., 2015. A tool to calculate safe nurse staffing levels. Nurs Times 111 (3), 12-14.

Ferguson-Paré, M., Bandurchin, A., 2010. The Ontario nursing workload demonstration projects: Rethinking how we measure, cost and plan the work of nurses. Nursing Leadership 23 (Special Issue).

Francis, R., 2013. Report of the Mid Staffordshire NHS Foundation Inquiry. The Stationary Office, London.

Gabbay, U., Bukchin, M., 2009. Does daily nurse staffing match ward workload variability? Three hospitals' experiences. Int J Health Care Qual Assur 22 (6), 625-641.

Griffiths, P., Ball, J., Bloor, K., Böhning, D., Briggs, J., Dall'Ora, C., Iongh, A.D., Jones, J., Kovacs, C., Maruotti, A., Meredith, P., Prytherch, D., Saucedo, A.R., Redfern, O., Schmidt, P., Sinden, N., Smith, G., 2018. Nurse staffing levels, missed vital signs observations and mortality in hospital wards: retrospective longitudinal observational study using routinely collected data. Health Services and Delivery Research Journal 6 (38), (whole issue).

Griffiths, P., Ball, J., Drennan, J., Dall'Ora, C., Jones, J., Maruotti, A., Pope, C., Recio Saucedo, A., Simon, M., 2016. Nurse staffing and patient outcomes: Strengths and limitations of the evidence to inform policy and practice. A review and discussion paper based on evidence reviewed for the National Institute for Health and Care Excellence Safe Staffing guideline development. Int J Nurs Stud 63, 213-225.

Griffiths, P., Ball, J., Drennan, J., James, L., Jones, J., Recio-Saucedo, A., Simon, M., 2014. The association between patient safety outcomes and nurse/healthcare assistant skill mix and staffing levels and factors that may influence staffing requirements (NICE evidence review). University of Southampton Centre for innovation and Leadership in Health Sciences.

Griffiths, P., Ball, J., Murrells, T., Jones, S., Rafferty, A.M., 2016. Registered nurse, healthcare support worker, medical staffing levels and mortality in English hospital trusts: a cross-sectional study. BMJ Open 6 (2), e008751.

Griffiths, P., Maruotti, A., Recio Saucedo, A., Redfern, O., Ball, J., Briggs, J., Dall'ora, C., Schmidt, P., Smith, G., On Behalf of the Missed Care Study Group, in press. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal observational study BMJ Quality and Safety.

Griffiths, P., Maruotti, A., Recio Saucedo, A., Redfern, O.C., Ball, J.E., Briggs, J., Dall'Ora, C., Schmidt, P.E., Smith, G.B., 2018. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. BMJ Quality & Safety.

Griffiths, P., Recio-Saucedo, A., Dall'Ora, C., Briggs, J., Maruotti, A., Meredith, P., Smith, G.B., Ball, J., Missed Care Study Group, 2018. The association between nurse staffing and omissions in nursing care: A systematic review. J Adv Nurs.

Harper, P.R., Powell, N.H., Williams, J.E., 2017. Modelling the size and skill-mix of hospital nursing teams. Journal of the Operational Research Society 61 (5), 768-779.

Hill, A.B., 1965. The Environment and Disease: Association or Causation? Proc R Soc Med 58 (5), 295-300.

Hoi, S.Y., Ismail, N., Ong, L.C., Kang, J., 2010. Determining nurse staffing needs: the workload intensity measurement system. Journal of Nursing Management 18 (1), 44-53.

Hoi, S.Y., Ismail, N., Ong, L.C., Kang, J., 2010. Determining nurse staffing needs: the workload intensity measurement system. J Nurs Manag 18 (1), 44-53.

Hurst, K., 2009. Nursing and payment By Results: Understanding the Cost of Care (RCN Policy Briefing: 11/2009). Royal College of Nursing, London.

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.

Hurst, K., 2008. UK ward design: Patient dependency, nursing workload, staffing and quality: An observational study. International Journal of Nursing Studies 45 (3), 370-381.

Hurst, K., Smith, A., Casey, A., Fenton, K., Scholefield, H., Smith, S., 2008. Calculating staffing requirements. Nurs Manag (Harrow) 15 (4), 26-34.

Jenkins-Clarke, S., 1992. Measuring nursing workload: a cautionary tale. Centre for Health Economics, University of York.

Junttila, J.K., Koivu, A., Fagerstrom, L., Haatainen, K., Nykanen, P., 2016. Hospital mortality and optimality of nursing workload: A study on the predictive validity of the RAFAELA Nursing Intensity and Staffing system. Int J Nurs Stud 60, 46-53.

Kane, R.L., Shamliyan, T.A., Mueller, C., Duval, S., Wilt, T.J., 2007. The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. Med Care 45 (12), 1195-1204.

Kessler, I., Heron, P., Dopson, S., Magee, H., Swain, D., Askham, J., 2010. The nature and consequences of support workers in a hospital setting. NIHR Service Delivery and Organisation programme.

Kolakowski, D., 2016. Constructing a nursing budget using a patient classification system. Nurs Manage 47 (2), 14-16.

Kortbeek, N., Braaksma, A., Burger, C.A.J., Bakker, P.J.M., Boucherie, R.J., 2015. Flexible nurse staffing based on hourly bed census predictions. International Journal of Production Economics 161 (167–180), 167-180.

Lang, T.A., Hodge, M., Olson, V., Romano, P.S., Kravitz, R.L., 2004. Nurse-patient ratios: a systematic review on the effects of nurse staffing on patient, nurse employee, and hospital outcomes. Journal of Nursing Administration 34 (7-8), 326-337.

Lankshear, A.J., Sheldon, T.A., Maynard, A., 2005. Nurse staffing and healthcare outcomes: a systematic review of the international research evidence. ANS Adv Nurs Sci 28 (2), 163-174.

Larson, E.L., Cohen, B., Liu, J., Zachariah, P., Yao, D., Shang, J., 2017. Assessing Intensity of Nursing Care Needs Using Electronically Available Data. Comput Inform Nurs 35 (12), 617-623.

Lewinski-Corwin, E.H., 1922. The hospital nursing situation. American Journal of Nursing 22 (8), 603-606.

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. Journal of advanced nursing 73 (9), 2248-2259.

Maben, J., Griffiths, P., Penfold, C., Simon, M., Anderson, J.E., Robert, G., Pizzo, E., Hughes, J., Murrells, T., Barlow, J., 2016. One size fits all? Mixed methods evaluation of the impact of 100% single-room accommodation on staff and patient experience, safety and costs. BMJ Qual Saf 25 (4), 241-256.

Maben, J., Griffiths, P., Penfold, C., Simon, M., Pizzo, E., Anderson, J., Robert, G., Hughes, J., Murrells, T., Brearley, S., Barlow, J., 2015. Evaluating a major innovation in hospital design: workforce implications and impact on patient and staff experiences of all single room hospital accommodation. Health Serv Deliv Res 3 (3), 254.

Maenhout, B., Vanhoucke, M., 2013. An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. Omega 41 (2), 485-499.

Mark, B.A., Harless, D.W., Spetz, J., Reiter, K.L., Pink, G.H., 2013. California's minimum nurse staffing legislation: results from a natural experiment. Health Serv Res 48 (2 Pt 1), 435-454.

Morales-Asencio, J.M., Porcel-Galvez, A.M., Oliveros-Valenzuela, R., Rodriguez-Gomez, S., Sanchez-Extremera, L., Serrano-Lopez, F.A., Aranda-Gallardo, M., Canca-Sanchez, J.C., Barrientos-Trigo, S., 2015. Design and validation of the INICIARE instrument, for the assessment of dependency level in acutely ill hospitalised patients. J Clin Nurs 24 (5-6), 761-777.

Myny, D., De Bacquer, D., Van Hecke, A., Beeckman, D., Verhaeghe, S., Van Goubergen, D., 2014. Validation of standard times and influencing factors during the development of the Workload Indicator for Nursing. J Adv Nurs 70 (3), 674-686.

Myny, D., Van Goubergen, D., Limere, V., Gobert, M., Verhaeghe, S., Defloor, T., 2010. Determination of standard times of nursing activities based on a Nursing Minimum Dataset. J Adv Nurs 66 (1), 92-102.

Myny, D., Van Hecke, A., De Bacquer, D., Verhaeghe, S., Gobert, M., Defloor, T., Van Goubergen, D., 2012. Determining a set of measurable and relevant factors affecting nursing workload in the acute care hospital setting: a cross-sectional study. Int J Nurs Stud 49 (4), 427-436.

National Institute for Health and Care Excellence, 2014. Safe staffing for nursing in adult inpatient wards in acute hospitals (Guidance).

Needleman, J., Buerhaus, P., Pankratz, V.S., Leibson, C.L., Stevens, S.R., Harris, M., 2011. Nurse staffing and inpatient hospital mortality. New England Journal of Medicine 364 (11), 1037-1045.

Needleman, J., Buerhaus, P., Pankratz, V.S., Leibson, C.L., Stevens, S.R., Harris, M., 2011. Nurse staffing and inpatient hospital mortality. N Engl J Med 364 (11), 1037-1045.

O'Brien-Pallas, L., Cockerill, R., Leatt, P., 1991. A comparison of workload estimates of five patient classification systems in nursing. Final Report. National Health Research and Development. National Health and Welfare Government of Canada: Ottawa, CN.

O'Brien-Pallas, L., Cockerill, R., Leatt, P., 1992. Different systems, different costs? An examination of the comparability of workload measurement systems. The Journal of nursing administration 22 (12), 17-22.

O'Brien-Pallas, L., Leatt, P., Deber, R., Till, J., 1989. A comparison of workload estimates using three methods of patient classification. Canadian journal of nursing administration 2 (3), 16-23.

Ozdemir, B.A., Karthikesalingam, A., Sinha, S., Poloniecki, J.D., Vidal-Diez, A., Hinchliffe, R.J., Thompson, M.M., Holt, P.J., 2015. Association of hospital structures with mortality from ruptured abdominal aortic aneurysm. Br J Surg 102 (5), 516-524.

Ozdemir, B.A., Sinha, S., Karthikesalingam, A., Poloniecki, J.D., Pearse, R.M., Grocott, M.P., Thompson, M.M., Holt, P.J., 2016. Mortality of emergency general surgical patients and associations with hospital structures and processes. Br J Anaesth 116 (1), 54-62.

Perroca, M.G., 2013. The new version of a patient classification instrument: assessment of psychometric properties. J Adv Nurs 69 (8), 1862-1868.

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

Recio-Saucedo, A., Dall'Ora, C., Maruotti, A., Ball, J., Briggs, J., Meredith, P., Redfern, O.C., Kovacs, C., Prytherch, D., Smith, G.B., Griffiths, P., 2018. What impact does nursing care left undone have on patient outcomes? Review of the literature. J Clin Nurs 27 (11-12), 2248-2259.

Rivera, K., 2017. A Comparison of a Standardized Method of Identifying Nursing Staff Needs with Measures of Nursing Staff Needs by Patient Acuity. The William Paterson University of New Jersey, Ann Arbor, pp. 58.

RN4CAST survey, unpublished data.

Royal College of Nursing, 2012. Policy Briefing: Mandatory Nurse Staffing levels. RCN, London.

Shekelle, P.G., 2013. Nurse-patient ratios as a patient safety strategy: a systematic review. Ann Intern Med 158 (5 Pt 2), 404-409.

Simon, M., Yankovskyy, E., Klaus, S., Gajewski, B., Dunton, N., 2011. Midnight census revisited: Reliability of patient day measurements in US hospital units. International Journal of Nursing Studies 48 (1), 56-61.

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

Taylor, B., Yankey, N., Robinson, C., Annis, A., Haddock, K.S., Alt-White, A., Krein, S.L., Sales, A., 2015. Evaluating the Veterans Health Administration's Staffing Methodology Model: A Reliable Approach. Nurs Econ 33 (1), 36-40, 66.

Telford, W., 1979. Determining nursing establishments. Health services manpower review 5 (4), 11-17.

The Shelford group, 2014. Safer Nursing Care Tool Implementation Resource Pack. The Shelford Group.

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.

Twigg, D., Duffield, C., Bremner, A., Rapley, P., Finn, J., 2011. The impact of the nursing hours per patient day (NHPPD) staffing method on patient outcomes: a retrospective analysis of patient and staffing data. Int J Nurs Stud 48 (5), 540-548.

Twigg, D.E., Geelhoed, E.A., Bremner, A.P., Duffield, C.M., 2013. The economic benefits of increased levels of nursing care in the hospital setting. J Adv Nurs 69 (10), 2253-2261.

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. J Nurs Manag 24 (2), 184-191.