Title

# Workforce thresholds and the non-linear association between registered nurse staffing and care quality in long term residential care: a retrospective longitudinal study of English care homes with nursing


## Authors

Andy **Charlwood**[1*], Danat **Valizade**[1], Louise **Winton Schreuders**[2], Carl **Thompson**[3], Matthew **Glover**[4], Heather **Gage**[4], David **Alldred**[3], Chris **Pearson**[5], Julie **Kerry**[5] and Karen **Spilsbury**[3]

## Affiliations

[1]Leeds University Business School, University of Leeds

[2]Independent researcher

[3]School of Healthcare, University of Leeds

[4]University of Surrey

[5]HC-One

## Corresponding author

*Andy Charlwood, Leeds University Business School, University of Leeds, Woodhouse Lane, Leeds, LS2 9JT

Twitter: @ProfAndyC

Email: a.charlwood@leeds.ac.uk



## Abstract

**Background:** Care needs amongst 425,000 dependent older residents in English care homes are becoming more complex. The quality of care in these homes is influenced by staffing levels, especially the presence of registered nurses (RNs). Existing research on this topic, often U.S.-focused and relying on linear assumptions, has limitations. This study aims to investigate the non-linear relationship between RN staffing and care quality in English care homes using machine learning and administrative data from two major care home providers.

**Methods:** A retrospective observational study was conducted using data from two English care home providers. Each were analysed separately due to variations in data reporting and care processes. Various care quality indicators and staffing metrics were collected for a 3.5-year period. Regression analysis and machine learning (Random Forest) were employed to identify non-linear relationships. Ethical approval was obtained for the study.

**Results:** Using linear methods, higher skill mix - more care provided by RNs - was associated with lower incidence of adverse outcomes, such as urinary tract infections and hospitalisations. However, non-linear skill mix-outcome relationship modelling revealed both low and high skill mix levels were linked to higher risks. The effects of agency RN usage varied between providers, increasing risks in one but not the other.

**Discussion:** The study highlights the cost implications of increasing RN staffing establishments to improve care quality, suggesting a non-linear relationship and an optimal staffing threshold of around one-quarter of care provided by nurses. Alternative roles, such as care practitioners, merit exploration for meeting care demands whilst maintaining quality. This research underscores the need for a workforce plan for social care in England. It advocates for the incorporation of machine learning models alongside traditional regression-based methods. Our results may have limited generalisability to smaller providers and experimental research to redesign care processes effectively may be needed.

**Conclusion:** RNs are crucial for quality in care homes. Contrary to the assumption that higher nurse staffing necessarily leads to better care quality, this study reveals a nuanced, non-linear relationship between RN staffing and care quality in English care homes. It suggests that identifying an optimal staffing threshold, beyond which increasing nursing inputs may not significantly enhance care quality may necessitate reconsidering care system design and (human) resource allocation. Further experimental research is required to elucidate resource-specific thresholds and further strengthen evidence for care home staffing.

### Tweetable abstract

How much nursing care is needed to assure quality in care homes? Evidence from 2 English care home providers shows that nurse sensitive outcomes (an indicator of quality) are better when ~25% of care is provided by nurses. Nurse shortages increase risks for residents

### Key words

Clinical indicators; Older people; Long-term care; Nurse staffing; Quality of care

### What is already known

- Care quality in long term residential care facilities (also known as care homes) tends to be higher when more care is provided by registered nurses (RNs).
- It is not clear whether this finding generalises to countries beyond North America.
- Previous studies' results may be biased as a result of methodological limitations including, data quality, omitted confounders, and methods that are unlikely to identify non-linear relationships between RN staffing and quality.

### What this paper adds

- Using machine learning methods, we found evidence of non-linear relationships between RN staffing and nurse sensitive indicators of care quality in long-term residential care facilities run by two large care providers in England.
- RN staffing below the planned level was associated with deteriorating care quality, RN staffing above the planned level did not improve care quality. Staffing levels were optimal, from a risk reduction perspective, when around 25% of care hours were provided by RN.
- There is mixed evidence that measured care quality suffers when agency RNs are used instead of permanent RNs.

## Workforce thresholds and care quality in care homes

### Background

Care homes (long-term care facilities) are an important part of societal provision of long-term residential care for increasing numbers of dependent older people. In England, the setting for this study, an estimated 425,000 older people live in ~18,000 care homes (Buisson, 2014). The acuity and complexity of resident care needs has risen significantly in recent decades (BGS, 2011; Clemens et al., 2021; CMA, 2018). This has important implications for the direct care workforce in care homes – a mix of registered nurses (RNs) and non-registered care staff. In care homes with nursing (or nursing homes), registered nurses are employed 24/7 to supervise care delivered mainly by a large workforce of non-registered care staff. In care homes without nursing (or residential homes), the workforce is comprised solely of care staff; the National Health Service (NHS) provides health-care input (including registered nursing care) on an 'as required' basis. RNs employed by the NHS may be involved in supporting specialist care for residents in both types of care homes (e.g. palliative care). Care staff in either of these settings (with and without nursing or dual registered) are employed at different levels – for example as care assistant of senior care assistant – and while not registered with any professional body (e.g. the Nursing and Midwifery Council), many of these social care staff possess vocational qualifications or have completed the Care Certificate (Skills for Care, 2021). In recent years (from 2019), the nursing associate role has been introduced into the sector (Care Quality Commission, 2019). Nursing Associates work alongside RNs, taking on some clinical skills previously undertaken solely by RNs. This is important context for understanding our sample of homes.

The literature on the associations between staffing and quality in long-term residential care environments (Spilsbury et al., 2011; Backhaus et al., 2014; Clemens at al., 2021; Blatter et al., 2023) is extensive. The balance of evidence suggests positive linear relationships between the quantity and quality of workforce inputs and nurse-sensitive indicators of care quality (for example lower incidence of medication errors, falls or pressure ulcers), i.e. quality tends to be higher when more care is provided by RNs. Reviews have identified a number of limitations with the existing evidence base. First, all previous studies use methods that assume linear relationships between nursing inputs and care quality when theoretically, we might expect non-linear relationships (Backhaus et al., 2014; Spilsbury et al. 2011). Second, most high quality longitudinal prospective or retrospective studies in this area were conducted in the United States, with one or two studies from Canada, Norway, Italy, Belgium and the Netherlands. There is a dearth of recent studies with only seven published since 2014 (Clemens et al., 2021). Third, studies rely on measures of quality that lack precision and do not have sufficient controls for resident-specific risks. Consequently, results are likely to be biased by measurement error and confounding. To address these points, it has been suggested that future studies should be based on retrospective chart-review of individual residents (Clemens et al., 2021; Blatter et al., 2024). To date, only two studies have followed this approach (Konetza et al., 2008; Feng et al., 2023).

In this context, our paper's key contribution is to apply machine-learning methods that are better able to detect non-linear relationships between measures of care quality and staffing than methods used in previous studies, using data and measures that are broadly comparable to those used previously. This is important, there are sound theoretical reasons to expect such relationships may be non-linear: a certain "threshold" level of RN staffing is likely to be necessary to achieve and sustain care quality, but tipping points may exist, beyond which additional RN staffing does not increase quality (Yakusheva et al., 2022l Griffiths, et al., 2019; Spilsbury et al., 2011; Donabedian, 2003; Hendrix, 2003; Jelinek, 1967). Understanding if such thresholds exist is important for policy makers and care providers making decisions about RN staffing. A secondary contribution is to investigate this issue in a context (England) with no previous studies. We draw on analysis of routinely collected administrative data from two large care home provider organisations, one a private company (provider one) the other a company limited by guarantee (i.e. a limited company that reinvests any profits in the business; provider two). These providers operated circa five per cent of English care homes providing nursing care at the time of the study.

In light of the previous literature, we were particularly interested in research questions that examine relationships between nurse sensitive indicators of care quality and differences in levels and variations in RN staffing over time:

- Skill mix: is there evidence of an optimal skill mix?
- Proportion of planned nursing hours actually worked: what happens to quality in periods where RN staffing is below the planned level (i.e. below protocol staffing)?
- Proportion of RN hours worked by agency nurses: does quality suffer when sickness absence and unfilled vacancies mean temporary agency nurses need to be employed?

## Methods

This was a retrospective longitudinal observational study. Data were provided by two large English care home provider organisations: *provider one,* operated 262 care homes at the time of the study; p*rovider two* operated 134 homes. We analysed data from each provider separately because the data and information we had did not lend itself to legitimately inferring the extent to which any between provider differences might represent unobserved differences in: 1) recording conventions and accuracy or record keeping; 2) resident specific risk; and 3) structures and processes of care.

### Data and variables

#### Provider One

Provider one data covered a period of 3.5 years or 182 weeks (September 2016 to February 2020). Care homes with nursing (nursing homes) or dual registered homes (with a mix of nursing and residential care) were included in our sample (n=186). Care homes without nursing (residential homes) were excluded from the sample, along with recently acquired care homes with significant data irregularities during the period of the study (n=76). The unit of observation used for analysis was per care home per week. Therefore, there were 33,852 care home week observations in total in the dataset (186 homes multiplied by 182 weeks).

Our outcome variables were nurse-sensitive indicators of care quality, i.e. resident outcomes influenced, but not necessarily determined, by nursing care. The key outcome variables in the provider one data set, measured as rate per occupied bed per week, were: falls, urinary tract infections (UTIs), chest infections, medication errors, pressure ulcers developed in the home, and resident hospitalisations. The measure of medication errors did not indicate the seriousness of the error or consequences for residents. We used it as a broad measure of whether differing levels and variations in staffing were associated with an error in care processes. Note that the use of hospitalisations as an indicator of care quality is controversial. While it has been used in a number of previous studies, review articles have criticised its use because not all hospitalisations are the result of failures or omissions in nursing care (Backhaus et al., 2014; Clemens et al., 2021). We have included hospitalisations as there are reasonable grounds for seeing it as a nurse-sensitive indicator of care quality (i.e. influenced but not necessarily determined by nursing care), and discussions with care home managers, staff and carers suggested many saw a link between nurse shortages, inexperience, and risk of hospitalisation. All data were recorded contemporaneously by care home staff using their incident reporting IT system. Descriptive statistics for these outcomes are reported in the results section below.

Weekly staffing measures were calculated from the provider's care planning tool and payroll IT systems. Specific workforce measures, all used in the study were: total care hours (carer + nursing assistant + registered nurse) per occupied bed; skill mix, i.e. the proportion of total care hours provided by RN; the proportion of planned RN hours actually worked, measuring above or below protocol staffing; the proportion of RN hours worked by agency RN; the proportion of carer hours worked by agency carers; and whether a manager was in post.

We included controls that measured aspects of the structures and processes of care: total beds (occupied and unoccupied); weekly occupancy rate; proportion of beds where residents were receiving nursing care per week; average daily admissions per occupied bed per week; if an embargo by a regulatory agency that prevented the home accepting new residents was in place for a given week; and planned care hours per occupied bed per week. Whilst strictly speaking a workforce measure, the provider's care planning tool meant carer hours were flexed according to regular assessments of residents' care needs. In the absence of more detailed information on residents' health this measure functions as a proxy for the acuity of residents' healthcare needs.

Provider two

Provider two data covered the 3.5 year period from December 2014 to May 2018. Differences in information systems and management processes between the two providers meant that, although recorded contemporaneously, data on nurse sensitive indicators of care quality were reported on a monthly (rather than weekly) basis. Therefore, the unit of analysis was the care home month rather than the care home week. As we have 134 homes covering a 42-month period (up to May 2018), there were 5,628 care home month observations.

Five of the outcome measures, chest infections, UTIs, falls, pressure ulcers developed in the home and medication errors, were the same as for provider one. There was no measure of hospitalisations, although falls that resulted in a bone fracture (which would likely require a hospital stay) were recorded as a separate category. All outcomes were measured per occupied bed per month.

Monthly staffing measures, again generated from payroll and rota systems, were: total care hours per occupied bed per month; skill mix (the proportion of total care hours provided by RN); the share of RN and carer hours provided by employment agencies; the proportion of planned RN hours worked; and the proportion of carer hours worked. Apart from a differing unit of analysis these measures were the same as for provider one. Provider two did not have data on whether a manager was in post.

Control variables for the provider two study were the total number of beds, monthly occupancy rate and average weekly admissions per month. Additionally, there were more detailed measures of resident characteristics: proportion of beds occupied by residents with nursing needs; proportion of residents with nursing needs; proportion of residents in dedicated dementia units, and those who exhibited 'challenging' behaviour; young disabled residents; residents with learning difficulties; residents with Parkinson's disease; residents with Huntingdon's disease; residents receiving end of life care; and residents with other specific care needs (see supplementary material for descriptive statistics of these additional variables). However, these were only recorded at a single point in time (at end of study period). These provide broad measures of home caseload and resident acuity, assuming that resident characteristics changed little over the time of the study.

Descriptive statistics for workforce and control variables enable the identification of similarities and differences between the two providers (Table 1). The main difference between providers was the proportion of residents receiving nursing care. In provider two at the end of the period of study around 66 per cent or residents were receiving nursing care, whilst around 40 per cent of residents in provider one's homes were receiving nursing care. Despite this difference, staffing arrangements were remarkably similar (mean total weekly care hours per occupied bed in provider two were 28.7 (SD 7.2 when analysed per week rather than month) compared to 27.8 (SD 5.46) for provider one. Skill mix was almost identical, with an average of 20 per cent of care hours delivered by RN in both providers (SD 5.7 for provider one and 9.3 for provider two). On average, provider two was slightly more likely to experience below protocol RN staffing. Provider one made much more extensive use of agency RN to cover unfilled vacancies and RN absence (22.5 per cent of RN hours provided by agency RN in provider one (SD 20.4) compared to 4.7 per cent (SD 3.1) in provider two). Both providers had similarly high occupancy rates (88 per cent (SD 20) for provider one, 86.5 per cent (SD 17.2) in provider two). Provider two's homes were both slightly larger on average and had more variation in home size.

It is important to understand the data generating processes that determined staffing levels and variations in staffing levels over time, which was broadly similar in both providers. RNs assess, plan, provide and supervise (largely non-registered) care staff in the delivery of care for residents with nursing needs, as well as administering medicines. Additional (unobserved) nursing care may be provided by nurses employed by primary or community care services (predominantly the National Health Service) who visit residents with specific nursing needs (for example, specialist nurses for dementia, palliative or wound care). For these providers, the amount of 'in-house' nursing care provided was a relatively fixed property of each care home: the number of RNs employed reflected the number of beds available for residents with nursing needs. This was typically one nurse per shift for 30 nursing beds, although homes where residents care needs were more acute might have larger nursing establishments. During the study period (final 18 months), provider one introduced nursing assistants into the staffing establishment in a small subset of homes in the study. These were experienced care staff trained by the organisation to carry out specific tasks usually carried out by registered nurses, for example medicines administration, but were another level of non-registered care staff employed by the organisation. If a RN was on sick leave, or if there were unfilled RN vacancies, both providers attempted to cover nursing shifts by employing agency RNs (if possible).

Both providers used broadly comparable acuity assessment tools to assess the care needs of residents. Planned care worker staffing reflected the aggregates of the regular acuity assessments. Therefore, if occupancy rates increased, or new residents with more acute care needs entered the care home, or the acuity of care needs of existing residents increased, care worker shifts were increased (when possible). This meant skill mix was 'diluted' when occupancy was higher or residents had greater care needs, and more 'concentrated' when occupancy was lower, or residents had less complex care needs. Care home managers made judgements about the capacity of the home to care for prospective new residents. If they did not think the home had the RN or care staff capacity to look after a potential resident, then they did not offer a bed to the individual. Managers sought to optimise the care provided, given available resources by limiting demands for care within the home when necessary. In both providers, managers told us that the financial performance of a home had no bearing on workforce planning for care and nursing staff. The same approach to workforce planning was used in homes where most of their residents were paying higher fees (self-funders) as in homes with more residents where state health and social care funding paid lower fees.

We conducted regression analysis to produce results comparable with previous international observational longitudinal studies. We initially experimented with several different regression methods, using data from provider two. First, pooled cross-sectional ordinary least squares (OLS) regression, then home fixed-effects, and then conditional growth models with random intercepts. The latter approach accounted for time invariant home characteristics (fixed-effects) and different home specific trajectories in outcome variables over time. The Intra Class Correlation (ICC) was used to illustrate the proportion of total variation outcomes that were due to differences in home-specific trajectories over time. The difference between conditional (variance explained by fixed effects only) and marginal (variance explained by the entire model) R-squared ($R^2$) showed that the conditional growth model with random intercepts fitted the data better than the fixed effects model (except for the models with falls with fracture where the ICC score was low). To test for non-linearity in relationships between staffing measures and care quality outcomes, we employed sensitivity analysis by including squared and quadratic terms for key staffing measures to see if we could detect non-linear relationships. We did the same with lagged variables of staffing measures to see whether staffing in one week was sensitive to changes in outcomes in subsequent weeks. We omitted lagged variables and squared and quadratic terms from the final analysis because the coefficients had confidence intervals that suggested a true value was close to zero.

Brieman (2001) argues that traditional regression based statistical analysis has serious limitations. The most pertinent being that goodness of fit tests do not lead to rejection of a linear relationship unless the non-linearity is extreme. For linear regression to identify non-linear relationships (e.g. through the use of quadratic terms), the non-linearity needs to be clear-cut, with as little variation as possible around the inflection point. In other words, regression analysis over-identifies linear relationships in non-extreme data. This is particularly important for our study given the theoretical prediction that there is likely to be an inflection point, or threshold, beyond which more staffing does not improve quality.

We used random forest (RF) analysis, a widely used machine learning algorithm. RF is an ensemble algorithm of simple predictive models (decision trees) built by bagging process (also termed 'bootstrapped aggregation') where explanatory variables ("features" in machine learning parlance) with best predictive performance are selected from a random combination of features at every splitting point. The number of random samples (equivalent to the number of trees) is itself a learning process, balancing between computational complexity and predictive accuracy of an algorithm. In this analysis, a model with 500 decision trees was deemed appropriate in that marginal improvements with the addition of further trees did not offset increasing computational demands. The RF algorithm arrives at the final solution by rejecting individual trees with high mean squared error (MSE), building the model from averaged results of the trees with lower MSE.

To guard against "over-fitting" - where a model is fitted to the data so well that results do not generalise (because the model is fitted to idiosyncratic aspects of the data) - the data were randomly split into test and training datasets. The model from the training data (75 per cent of the data) was used to predict the outcomes in the test dataset (25 per cent of the data); a process known as cross-validation. In the final provider two analysis, we did not include measures of resident characteristics (which were only measured at a single point in time). This did not affect model performance.

Another integral element of a machine learning approach is parameter tuning to further guard against model over-fitting and improve predictive accuracy (Valizade et al., 2024).We used an Exhaustive Grid Search approach where all possible combinations of RF parameters are tried to find the best set of random forest hyper-parameters, the types of parameters that cannot be learned through training process and are therefore specified manually (Choudhury et al., 2021; Svetnik et al., 2003). In random forest, such parameters include maximum number of features considered for splitting a node, maximum number of decision tree levels, minimum number of samples considered within a tree and whether data points are further bootstrapped at each split (through hyper-tuning we fitted an algorithm with 1000 random splits at each node).

Theoretically, we might expect there to be a time-delay between staffing shortages and the consequences of these shortages. Technically, if this was the case, data would be non-stationary, so failure to account for this by including lagged measures of staffing variables would lead to spurious results. Consequently we tested (using augmented Dickey Fuller tests) whether data were stationary/non-stationary. The results of these tests are reported in figure S4 and visualised in figure S5, included in the online supplementary material. They suggest that data are stationary so lagged staffing measures are not necessary. In these circumstances including lags would add to the complexity of the analysis without improving model performance or changing our main results, so we report the results of analyses without lagged staffing measures.

To unpack the RF results we first generated feature importance scores for each of our independent variables in each model. Feature importance scores tell us the extent to which the predictive accuracy of a model (measured by MSE) would change if the variable was excluded from the model. Then, to visualise relationships between dependent and key independent variables, we plotted accumulated local effects (ALE) plots (Biekek, 2018; Apley and Zhu, 2020). ALE plots visualise the relationship between two features (variables) in a model while controlling for the effects of other features. Additionally, individual conditional expectation (ICE) plots were inspected to examine the extent to which patterns of results for individual homes corresponded to the average relationships revealed by the ALE plots. Space precludes including these ICE plots but we have included and reported on ALE plots where inspection of ICE plots suggests the average relationships revealed are typical of patterns in individual homes.

Based on the RF results, an exploratory cost analysis was conducted. For a hypothetical “average” care home (provider 1 and 2), the cost of increasing the proportion of care provided by RN to optimal levels, and the potential associated treatment cost savings from reduced events to the NHS were estimated (2019/20 prices). Details of costing methods, including staffing and outcome unit costs, are available in the supplemental material.

We were granted ethical approval for our work by the Social Care Research Ethics Committee (17/WM/0232) and the University of Leeds, Faculty of Medicine and Health, Ethics and Governance Committee (HREC 20-003).

## Results

### Incidence of outcome measures

We report the incidence of different outcome measures for both providers (Table 2). Note that all outcomes are rare events. The most common in both providers is falls, occurring at a mean rate of 0.033 (SD 0.039) per occupied bed per week (or one fall per 30 occupied beds) in provider one and 0.197 (SD 0.009) falls per occupied bed per month in provider two (that equates to 0.045 or one fall per 22 occupied beds per week if monthly results were averaged out over 52 weeks). Rates of pressure ulcers developed in the home, hospitalisations (provider one) and falls with fractures (provider two) were much lower. For example, the rate of pressure ulcers in provider one was 0.001 (SD 0.005) per occupied bed per week, or one per thousand occupied beds. For provider two, the equivalent figure was 0.002 (SD 0.001) or one per 500 occupied beds.

### Regression analysis

Full results for our preferred growth mixture model with random intercepts are reported in tables S1 and S2 of the online supplementary material (further results from pooled OLS and fixed effects models for provider two can be found in Spilsbury et al., in press). The most interesting results (represented in Figure one) are statistically significant associations between skill mix, UTIs (both providers) and hospitalisations (provider one) and falls that resulted in fractures (provider two). Here we see linear relationships between a higher skill mix and lower rates of the outcomes, so that the greater the proportion of care provided by RN, the lower the incidence of the outcomes. Additionally, a higher skill mix is associated with fewer medication errors in provider two only.

Random forest analysis

We fitted separate RF models for each provider, with the six nurse sensitive indicators of care quality (falls, UTIs, chest infections, medication errors, hospital admissions/ falls with fractures and pressure ulcers) as the outcome measures. We do not report results for the models for pressure ulcers because the models we fitted to the training data had extremely low predictive accuracy. We report key model diagnostic information in the accompanying online supplementary material (Tables S4 & S5). Model diagnostics suggest reasonably well-fitting models with respectable levels of predictive accuracy for data of this sort. Feature importance scores are also reported in the supplementary material (Tables S6 & S7). These suggest that skill mix, the proportion of planned nursing hours actually worked and total care hours are all important predictors of our outcome measures in both provider datasets. Contrastingly, agency nurse use has virtually no impact on predictive accuracy in the provider one models, but a large impact on predictive accuracy in the provider two models.

ALE plots (figures 3 & 4) visualise the relationships between our outcome variables, skill mix and actual vs planned nursing hours. Space constraints prevent us from including the ALE plots for total care hours and agency RN use. These are included in the supplementary material. Recall that ALE plots show the relationship between independent and key dependent variables controlling for other variables included in the RF analysis (including measures that control for caseload and resident acuity).

Skills mix

The left panel of figure 3 shows relationships between skills mix and the outcome measures for provider one. A consistent finding across all the plots is that risks to residents increase when less than a fifth of care hours are provided by RN. Rates of hospitalisation and UTIs also increase when more than one third of care hours are provided RN. The equivalent results for provider two (left panel of figure 4) also show rates of falls with fractures and medication errors increase when less than a fifth of care hours are provided by RN. Rates of UTIs start to increase when less than 14 per cent of care hours are provided by RN and rates of falls rise when less than 10 per cent of care hours are provided by RN. Rates of falls with fractures, UTIs, chest infections and medication errors all increase when more than one third of care hours are provided by RN.

Staffing below planned levels

The right panel of figure 3 shows that for provider one, rates of all outcome measures increase when RN staffing is both above and below planned levels. The right panel of figure 4 shows similar results for provider two.

Agency RN use

Agency RN use was associated with increased risks to residents in provider two but not provider one (ALE plots visualising agency nurse use/outcome relationships for provider two are included in figure S1 in the supplementary material). We will consider possible reasons for this in the discussion section below.

Total care hours

Feature importance scores suggested that total care hours were important for determining outcomes. ALE plots (included in figure S2 in the supplementary material) suggest a broadly flat relationship between total care hours and outcomes up to a certain point, with risks increasing when total care hours are higher. In provider one, risks to residents tend to increase when total care hours are higher than 35 hours per occupied bed/week. In provider two, risks of UTIs, chest infections and medication errors increased when total care hours per occupied bed/month rose above 180. The patterns for falls related measures in provider two were different to provider one.

Cost analysis

Exploratory cost analysis is reported in supplemental material (Table S10 & 11). Results of this analysis suggest modest potential savings in annual treatment costs when one fifth of care is provided by RN, relative to the costs of substituting non-registered care staff with RN (additional c. £80-90,000 annually). Highest cost savings were associated with lower hospital admission (£2,891 in provider 1) and falls (£3,691 provider 1; £1,360 provider 2). An element of double counting in these calculations cannot be ruled out and translating these cost savings into cost benefit analysis is not straightforward: the counter-factual costs of ensuring a home is *not* short of nursing staff are unclear.

## Discussion

The starting point for our analysis were linear regression models to compare with previous studies examining staffing – care quality in care homes. These results were broadly comparable with the majority of previous studies in that they suggest that if a greater proportion of care is provided by RN risks to residents are lower. Very large increases in skill mix are associated with modest reductions in the incidence of UTIs and hospitalisations/falls resulting in fractures.

We turned to RF analysis to better identify non-linear relationships between staffing and care quality. In contrast to the regression results, RF analysis suggests that relationships between key staffing measures (skill mix and whether RN staffing was above or below planned levels) and care quality measures tended to be non-linear. For both providers, RN staffing below planned levels was associated with increased risks to residents, with higher rates of all quality measures in both providers when homes were short of RN. This relationship is non-monotonic in that it appears staffing can be slightly below protocol before risks start to rise. However, when RN staffing was above planned levels, risks either increased or were not reduced by the increased staffing levels. Similarly, examining relationships between skill mix and quality related outcomes using RF suggests non-linear relationships where care quality is optimised when around one quarter of care is provided by RN (with some variation around this figure depending on outcome and provider). When more than one quarter of care is provided by RN, there is no further reduction in risks, and risk of some outcomes occurring actually increase. These results are in line with the (theoretical) proposition that there is likely to be a point beyond which more nursing care does not improve outcomes (Donabedian, 2003; Hendrix, 2003; Jelinek, 1967). This finding is significantly different to all previous studies of relationships between staffing and care quality in care homes (Clemens et al., 2021; Backhaus et al., 2014) and likely reflects the differences in methods used. Future studies on this topic should (and could) use methods better able to detect non-linear relationships.

We also explored relationships between total care hours and quality outcomes. Space constraints prevent us from fully reporting these results in the article, but they can be found in the accompanying online supplementary material. In provider one, relationships between total care hours and quality outcomes were broadly flat until total care hours exceeded around 30 hours per occupied bed per week. Similarly, in provider two relationships between outcomes (other than falls) and total care hours were broadly flat until total care hours rose above around 180 care hours per occupied bed per month. A likely explanation for these results is that higher care hours is associated with (unobserved) greater acuity in resident care needs, with higher acuity being a significant risk factor for the outcome measures. This points to the need for studies that are able to adjust for resident specific risk using chart review methods (Clemens et al., 2021; Blatter et al., 2024). Note the key difference in patterns of results between provider one and two here relates to measures of falls. It was evident from our discussions with providers that they had different approaches to managing falls risks, and this may explain the different patterns of results.

Space constraints also prevented the inclusion of full results for relationships between agency nurse use and quality outcomes (full results can be found in the supplementary material). There is an expectation that quality may be worse when agency RN are used, because agency RN will have less home specific experience and will not know residents and their care needs as well as permanent RN but this is an issue that has not been well-explored in previous studies (Clemens et al., 2021; Spilsbury et al. 2011). Our results appear somewhat contradictory. Agency RN use was an important predictor of outcomes in provider two but not provider one. Inspection of the ALE plots from provider two provide some insight into why this might be. Provider two made less extensive use of agency RNs than provider one. Provider two had on average around 5% of RN hours provided by agency RN, compared to provider one where ~20% of RN hours were provided by agencies. The ALE plots for provider two suggest that risks to residents tended to increase when around 10% of care was provided by agency RNs. Risk tended to increase in a linear fashion as agency use increased, but there were few observations with the level of agency use routinely observed in provider one. By contrast, ALE plots for provider one did not suggest any meaningful relationship between agency use and risks to residents. From this we infer that in circumstances where agency RN use is rare, above a certain level (approximately around 10 per cent) increasing agency RN use is associated with worsening quality. However, in circumstances where agency RN use is more common, relationships between agency RN use and quality are less clear-cut. Therefore agency RN/quality relationships may also be moderated by different systems and processes in different providers.

What are the implications of these results for policy and practice? Given the distribution of below protocol staffing within and between homes over time, there is no easy solution to the problem of below protocol staffing. Both providers told us they used agency RNs to plug rota gaps. Routinely higher RN staffing establishments is an expensive way of trying to solve the problem. The typical model of staffing in provider one was one RN per 30 residents per shift. This means homes with 30 residents with nursing needs might have to double RN staffing to ensure they always have one RN per 30 residents (so there was already one RN in place if the other RN left the organisation or was absent due to sickness). Although we found there might be savings to the wider healthcare system in reduced treatment costs, any savings would be wiped out by the high additional costs of employing more RN. Even if significantly greater financial resources were available to meet the costs of hiring more RNs, recruiting the extra would not be easy at a time of widespread nursing shortages. Indeed, care provider organisations are having to consider how to meet the growing demands for care that may previously have been undertaken by RNs, by developing care practitioner roles to plug the gaps. As provider organisations flex to meet demand, there will be a need to research the impact of such roles for people living in care homes, as well as on the work, well-being and satisfaction of staff who make-up the team. In this context, any actions on the part of care home providers that can protect and enhance RN wellbeing so that sickness absence and attrition due to poor morale and burnout are minimised will likely save costs and enhance quality. Our study further highlights the need for a workforce plan for social care in England (National Care Forum, 2023).

Our results have implications for the study of relationships between nurse staffing and care quality beyond our long-term care context. Carefully conducted studies that account for unobserved differences between high and low staffed hospitals by looking at associations between hospital nurse staffing and patient outcomes have largely found linear associations between RN staffing levels and measures of care quality (Griffiths et al., 2019; Needleman & Shekelle, 2019). Machine learning methods might identify non-linear relationships that these studies have missed. A recent study looking at association between RN staffing levels and readmission in US hospitals using machine learning methods found evidence of a previously unidentified non-linear association (Yakusheva et al., 2021). We recommend that in addition to traditional regression-based methods, researchers studying associations between nursing workforce and care quality related patient outcomes consider incorporating machine learning models into their analytical strategies and report models that best fit their data. When combined with theory-based modelling and interpretation the evidence base for staffing and quality could grow more efficiently and reliably.

The non-linear relationships we have identified in this study also point to the limitations and difficulties of trying to infer optimal nurse staffing levels from observational data. One possible inference from this study is that the result that quality is maximised when between one fifth and one quarter of care is provided by RN can be generalised to other contexts and countries. We think this inference is wrong: the providers we studied tried to optimise systems and processes of care *given their available resources*. Increasing staffing levels above protocol may not improve care quality because systems and processes of care are not designed to make effective use of the additional resource. Consequently, we cannot be confident results would generalise to other settings and contexts, with their own unique systems and processes. Therefore, while observational studies are likely effective for identifying the harms that arise from below-protocol RN staffing, they cannot tell us what would happen if systems and processes of care were redesigned to optimise care with routinely higher or lower levels of RN care input. This would suggest a need for more experimental or quasi-experimental studies. Advocates of observational studies have pointed to the ethical and practical difficulties in randomising nurse staffing for experimental analysis (Needleman & Shekelle, 2019). Nevertheless, we are unlikely to be able to come up with a convincing answer to the question of "what level of nursing care is needed to optimise outcomes?" without experiments or quasi-experiments that build on each other in theoretically defensible ways (Ivers et al., 2014).

## Limitations

This study has a number of limitations that need to be kept in mind when interpreting the results. First, data come from two large providers while most care home operators in England operate on a much smaller scale. This may limit the generalisability of results. Second, lack of variation in staffing models observed within this study may mute important results that would be apparent in a more diverse sample. Third, nurse-sensitive indicators of care quality captured in incident reporting systems may measure quality of record keeping not quality of care (Spilsbury et al., 2011). Fourth, we also lack controls for a number of potentially important confounders, particularly resident specific risk and the care inputs provided by the primary care sector. Nevertheless, we think the additional evidence provided by this paper makes a valuable contribution to the literature given the “dearth of recent studies” providing empirical evidence on relationships between nurse staffing and quality in care home, particularly in non-US contexts (Clemens et al., 2021: 11) and the need to better test for nonlinear relationships between care quality and staffing (Backhaus et al., 2014: 392).

The “gold standard” method for future research that would address these limitations is retrospective chart review. In the English context, this might involve analysis of residents’ NHS electronic health records. A broader picture of nursing care received could also come from data on primary and community care workforces from the workforce minimum datasets. Data linkage between care home records and residents’ NHS electronic care records could also allow for more accurate estimation of the additional treatment costs associated with lower nurse-staffing levels and staffing shortages. Finally, the aspect of quality that matters most to residents is the way care makes them feel (Haunch et al., 2021; Spilsbury et al., in press) and our measures of care quality do not capture this. Future research should aim to develop and deploy measures of care quality that capture residents’ experiences of care and quality of life to use alongside measures of health outcomes (Burton et al., 2022).

## Conclusion

RNs are crucial for quality in care homes. Previous studies of relationships between RN staffing and nurse-sensitive indicators of care quality in the care home context have tended to suggest linear relationships between RN staffing and quality where more RN reduces risks to residents. Using ML methods better able to detect non-linear relationships, the results of this study challenge the findings of previous studies. Contrary to the assumption that higher nurse staffing necessarily leads to better care quality, this study reveals a nuanced, non-linear relationship between RN staffing and care quality in English care homes. Care quality was optimised when between a fifth and a quarter of care is provided by RN. Increasing RN staffing above planned levels is associated with increased risks. However, we do not infer that these results de facto generalise to other contexts, or represent a global optima. Systems and processes of care were designed to optimise care *given* available resources. If systems and processes of care were re-designed around routinely higher or lower levels of RN resource, the observed optima might be different. Further experimental or quasi-experimental research is required to elucidate resource-specific thresholds and further strengthen evidence for care home staffing.

**Funding**

The work is funded by the National Institute for Health and Care Research (NIHR) Health and Social Care Delivery Programme (NIHR15/144/29) and Policy Research Programme (NIHR201429). KS is supported by the NIHR Applied Research Collaboration Yorkshire and Humber and is a NIHR Senior Investigator.

**Disclaimer**

The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Data availability statement**

The data sharing agreements we have with care home providers means that we are unable to share data with third parties.

## References

Apley, D.W. and Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 82(4), pp.1059-1086.

Backhaus, R., Verbeek, H., van Rossum, E., Capezuti, E. & Hamers, J.P.H. (2014). Nurse Staffing Impact on Quality of Care in Nursing Homes: A Systematic Review of Longitudinal Studies. Journal of the American Medical Directors Association, 15(6):383-393.

Biecek, P., 2018. DALEX: explainers for complex predictive models in R. The Journal of Machine Learning Research, 19(1), pp.3245-3249

Blatter, C., Osińska, M., Simon, M. & Zúñiga, F. (2024) The relationship between nursing home staffing and resident safety outcomes: A systematic review of reviews. International Journal of Nursing Studies, 150 (104641).

Breiman, L. (2001). Statistical Modelling: The Two Cultures. Statistical Science, 16(3): 199 – 231.

Burton, J., Wolters, A., Towers, AM., Jones, L., Meyer, J., Gordon, A., Irvine, L.. Hanratty, B., Spilsbury, K., Peryer, G., Rand, S., Killett, A., Akdur, G., Allan, S., Biswas, P. & Goodman, C. (2022) Developing a minimum data set for older adult care homes in the UK: exploring the concept and defining early core principles, The Lancet Healthy Longevity, 3(3):e186-e193.

Care Quality Commission. (2019) Briefing for Providers: Nursing Associates. URL: https://www.cqc.org.uk/sites/default/files/20190123_briefing_for_providers_nursing_associates_0.pdf

Choudhury, P, Allen, RT, Endres, MG. (2021). Machine learning for pattern discovery in management research. Strategic Management Journal, 42(1): 30– 57.

Clemens, S., Wodchis, W., McGilton, K., McGrail, K. & McMahon, M. (2021). The relationship between quality and staffing in long-term care: A systematic review of the literature 2008–2020. International Journal of Nursing Studies, Volume 122, https://doi.org/10.1016/j.ijnurstu.2021.104036.

CMA (2018). Care Homes Market Study. https://www.gov.uk/government/publications/care-homes-market-study-summary-of-final-report/care-homes-market-study-summary-of-final-report (downloaded 07-03-22)

Donabedian, A. (2003). An Introduction to Quality Assurance in Healthcare. Oxford: Oxford University Press.

Feng Z, Vadnais A, Huber B, Deutsch A, Li Q, Bercaw L, Ingber MJ, Segelman M, Khatutsky G, Sroczynski N, Xu L. (2023). Hospital Transfer Rates among Long-Stay Nursing Home Residents: Variation by Day of the Week. Journal of the American Medical Directors Association, 24(9):1361-1362.

Griffiths P, Maruotti A, Recio Saucedo A et al. (2019). Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. BMJ Quality & Safety, 28:609-617.

Haunch, K., Thompson, C., Arthur, A., Edwards, P., Goodman, C., Hanratty, B., Meyer, J., Charlwood, A., Valizade, D., Backhaus, R., Verbeek, H., Hamers, J. & Spilsbury, K. (2021) Understanding the staff behaviours that promote quality for older people living in long term care facilities: A realist review. International Journal of Nursing Studies, 117: 103905.Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of statistical software, 28: 1-26.

Hendrix, T. (2003) Optimization techniques: industrial production tools with applications in nurse staffing Efficiency Research. Outcomes Management, 7(4): 194 – 197

Ivers, N.M., Sales, A., Colquhoun, H. et al. (2014) No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. Implementation Science. 9(14) https://doi.org/10.1186/1748-5908-9-14

Jelinek, R. (1967) A structural model for the patient care operation. Health Services Research, 2(3): 226 – 242.

Konetzka, R.T., Stearns, S.C. and Park, J. (2008), The Staffing–Outcomes Relationship in Nursing Homes. Health Services Research, 43(3):1025-1042.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.I., (2020). From local explanations to global understanding with explainable AI for trees. Nature machine intelligence, 2(1): 56-67.

National Care Forum (2023) https://www.nationalcareforum.org.uk/ncf-response-to-nhs-englands-long-term-workforce-plan/

Needleman J., Shekelle P.G. (2019) More ward nursing staff improves inpatient outcomes, but how much is enough? BMJ Quality & Safety 28:603-605.

Skills for Care. (2021) The State of the Adult Social Care Sector and Workforce in England 2021. URL: https://www.skillsforcare.org.uk/adult-social-care-workforce-data-old/Workforce-intelligence/documents/State-of-the-adult-social-care-sector/The-State-of-the-Adult-Social-Care-Sector-and-Workforce-2021.pdf

Spilsbury, K., Hewitt, C., Stirk, L., Bowman, C. (2011). The relationship between nurse staffing and quality of care in nursing homes: A systematic review. International Journal of Nursing Studies, 48(6): 732-750.

Spilsbury K, Charlwood A, Thompson CA, Haunch K, Valizade D, Devi R, Jackson C, Alldred DA,Arthur A, Brown L, Edwards P, Fenton W, Gage H, Glover M, Hanratty B, Meyer J, Waton A (in press) Understanding the Staffing Relationship to Quality in care homes: the StaRQ mixed-methods study. Health and Social Care Delivery Research

Svetnik V, Liam A, Tong C, et al. (2003) Random forest: A classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Science 43(6): 1947–1958.

Valizade, D., Schulz, F. and Nicoara, C. (2024), Towards a Paradigm Shift: How Can Machine Learning Extend the Boundaries of Quantitative Management Scholarship?. British Journal of Management, 35(1):99-114.

Yakusheva O, Bang J, Costa L, Hughes R, Weiss MA. Non-linear association of nurse staffing with readmissions uncovered in machine learning analysis. Health Services Research. 2022 Health Serv Res. 2022 Apr;57(2):311-321

**Table 1 Descriptive statistics for workforce and control variables**

| Variable | Mean | SD | Median | IQR | Mean | SD | Median | IQR |
|---|---|---|---|---|---|---|---|---|
| | PROVIDER ONE (weekly observations) | | | | PROVIDER TWO (monthly observations) | | | |
| **WORKFORCE MEASURES** | | | | | | | | |
| **Total care hours (carer + nursing assistant + registered nurse)/week or month (ratio of total care hours/occupied beds)** | 27.8 | 5.46 | 27.37 | 5.77 | 124.52 | 31.1 | 122.48 | 35.48 |
| **Skill mix: Proportion of total care hours provided by registered nurses (ratio of nursing hours/total care hours)** | 0.202 | 0.057 | 0.201 | 0.079 | 0.203 | 0.093 | 0.225 | 0.101 |
| **Proportion of planned nursing hours actually worked (ratio of weekly or monthly nursing hours worked/weekly or monthly nursing hours planned)** | 1.06 | 0.26 | 1.008 | 0.228 | 0.982 | 0.973 | 0.123 | 0.116 |
| **Proportion of planned carer hours actually worked (ratio of weekly or monthly carer hours worked/weekly or monthly planned carer hours)** | - | - | - | - | 1.001 | 0.134 | 0.99 | 0.136 |
| **Proportion of RN hours worked by agency RN** | 0.225 | 0.204 | 0.187 | 0.31 | 0.047 | 0.07 | 0.019 | 0.067 |
| **Proportion of carer hours worked by agency carers** | 0.0730 | 0.102 | 0.026 | 0.114 | - | - | - | - |
| **Manager in post (Weekly records 0 – no manager in post, 1 – manager in post)** | 0.918 | 0.271 | 1 | - | - | - | - | - |
| | | | | | | | | |

| CONTROLS | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Occupancy rate (ratio of occupied beds/available beds)** | 0.88 | 0.2 | 0.891 | 0.172 | 0.865 | 0.122 | 0.09 | 0.138 |
| **Proportion of total number of beds that are for residents receiving nursing care (ratio of nursing beds/all beds)*** | 0.397 | 0.192 | 0.39 | 0.261 | 0.661 | 0.323 | 0.714 | 0.424 |
| **Total numbers of beds in a care home (occupied and unoccupied)** | 63.31 | 32.723 | 52.0 | 33 | 56.61 | 25.83 | 52 | 29 |
| **Average admissions per occupied bed per week or month** | 0.131 | 0.19 | 0.143 | 0.156 | 0.024 | 0.024 | 0.02 | 0.022 |
| **Home under embargo by regulatory agency (unable to accept new residents)** | 0.155 | 0.255 | 0 | - | - | - | - | - |
| **Planned care hours per occupied bed (ratio of planned weekly care hours/occupied beds)** | 26.141 | 5.344 | 25.65 | 5.943 | - | - | - | - |
| **Home/week where nursing hours have been substituted for nursing assistant hours** | 0.026 | 0.191 | 0 | - | - | - | - | - |

Base: 33,852 care home week observations from 186 care homes with nursing (provider one) and 5,628 care home month observations from 134 care homes (provider two)

* For provider two, the proportion of residents receiving nursing care is measured at a single point in time at the end of the study period.

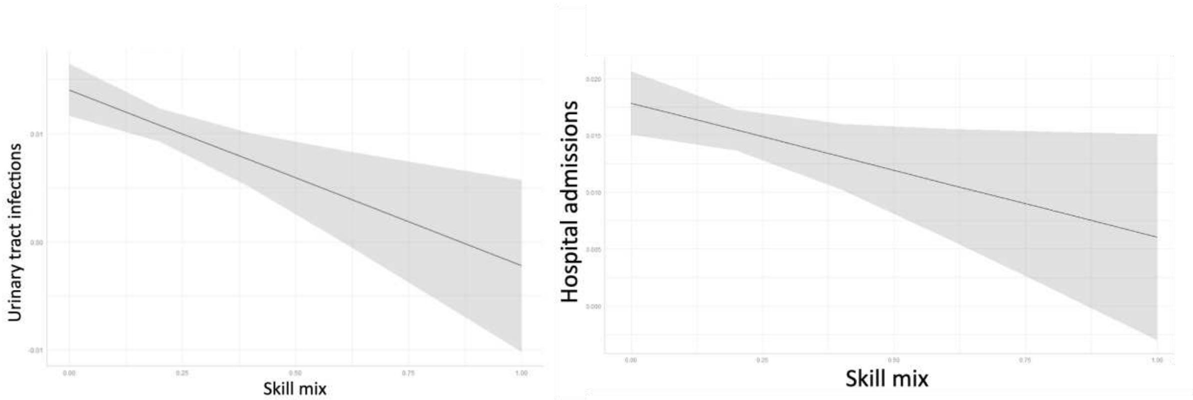
Further descriptive statistics for resident characteristics of provider two at the end of the study period can be found in table S1 in the online supplementary material.

**Table 2 Descriptive statistics for outcome variables**

| | PROVIDER ONE (rate per occupied bed/week) | | | | PROVIDER TWO (rate per occupied bed/month) | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | **Mean** | **SD** | **Median** | **IQR** | **Mean** | **SD** | **Median** | **IQR** |
| **Falls** | 0.033 | 0.039 | 0.026 | 0.05 | 0.197 | 0.009 | 0.151 | 0.201 |
| **Urinary tract infections** | 0.009 | 0.017 | 0.00 | 0.017 | 0.069 | 0.082 | 0.044 | 0.0103 |
| **Chest infections** | 0.011 | 0.021 | 0.00 | 0.02 | 0.051 | 0.07 | 0.029 | 0.072 |
| **Medication errors** | 0.003 | 0.011 | 0.00 | 0.000 | 0.016 | 0.05 | 0 | 0.018 |
| **Pressure ulcers** | 0.001 | 0.005 | 0.00 | 0.000 | 0.009 | 0.021 | 0 | 0.009 |
| **Hospital admissions** | 0.013 | 0.019 | 0.00 | 0.022 | - | - | - | - |
| **Falls resulting in fractures** | - | - | - | - | 0.003 | 0.009 | 0 | 0 |

Base: 33,852 care home week observations from 186 care homes with nursing (provider one) and 5,628 care home month observations from 134 care homes (provider two

**Figure 1 Marginal effects of the relationship between Skill mix and rate per occupied bed/week of UTIs (left panel) and hospitalisations (right panel), provider one**




Calculated from the results of regression analysis (conditional growth models with random intercepts) reported in supplemental material table 1

**Figure 2 Marginal effects of the relationship between Skill mix and rate per occupied bed/month of UTIs (left panel) and falls with fractures (right panel), provider two**

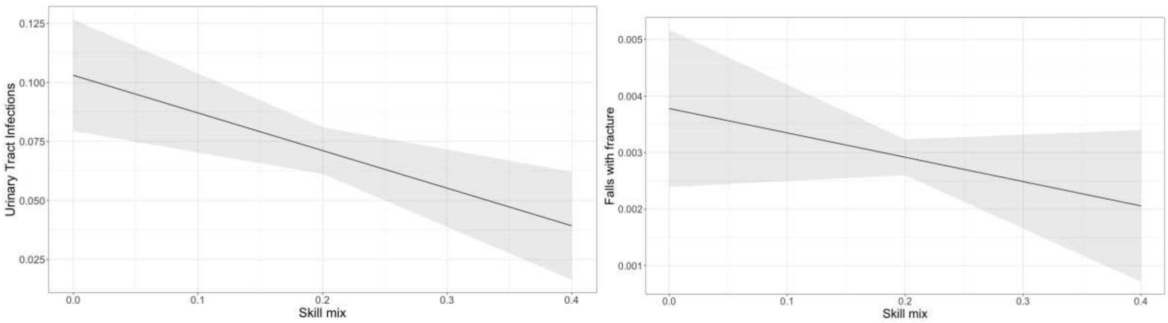


**Figure 3 ALE plots showing the associations between the weekly incidence (per occupied bed) of hospitalisations, UTIs, chest infections, medication errors, falls , and skill mix (left hand panel panel) and the proportion of planned nursing hours actually worked (right hand panel) (provider one)**

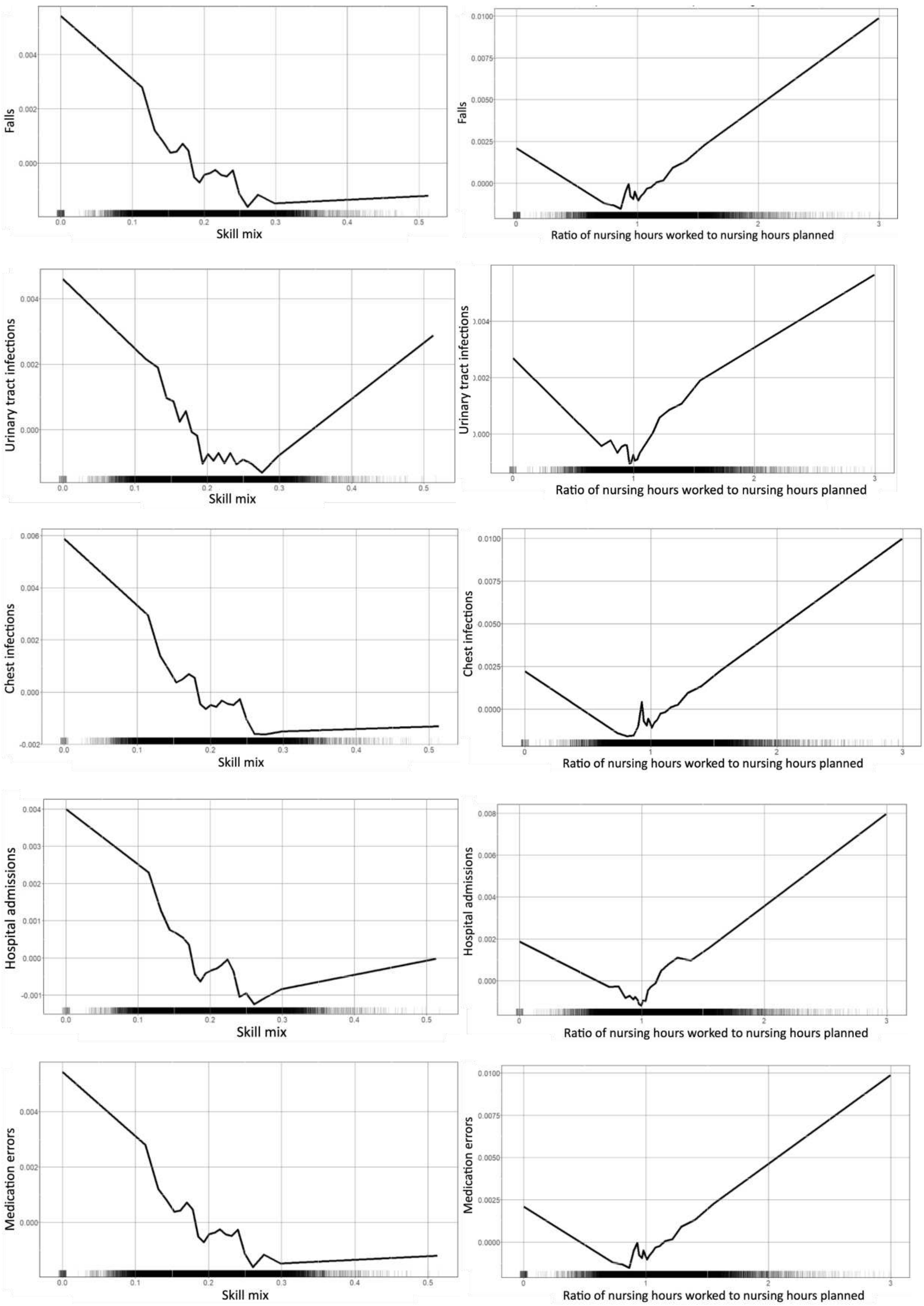


**Figure 4 ALE plots showing the associations between monthly incidence (per occupied bed) of falls with fractures, UTIs, chest infections, medication errors, falls , and skill mix (left hand panel panel) and the proportion of planned nursing hours actually worked (right hand panel) (provider two)**