An observational study on the rate of reporting of adverse event on healthcare staff in a mental health setting: an application of Poisson Expectation Maximisation analysis on nurse staffing data

Key words: nurse staffing, safety, workforce, retention, count regression

Robert M. Cook1, Sarahjane Jones1, Daniel Worsley2, Gemma C. Williams1. Ray Walker3, Mark Radford1&4 and Alison Leary5

1Birmingham City University, 2Mersey Care NHS Trust, 3Health Education England, 4NHS England and NHS Improvement and 5London South Bank University

## Abstract:

Evidence highlights the intrinsic link between nurse staffing and expertise, and outcomes for service users of healthcare, and that workforce retention is linked to the clinical and organisational experiences of employees. However, this understanding is less well established in mental health. This study comprises a retrospective observational study carried out on routinely collected data from a large mental healthcare provider. Two databases comprising nurse staffing levels and adverse events were modelled using latent variable methods to account for the presence of multiple underlying behaviours. The analysis reveals a strong dependence of the rate of adverse events on the location and perceived clinical demand of the wards, and a reduction in adverse events where registered nurses exceed ‘clinically-required levels’. In the first study of its kind, these findings present significant implications for nursing workforce policy and present an opportunity to not only improve safety, but potentially impact nurse retention.

## Introduction

Since the publication of To Err is Human (1), patient safety has received growing attention from researchers and policy makers worldwide (2). The NHS in England is no exception to this and remains the subject of numerous reports condemning the state of patient safety and highlighting insufficiencies that contribute to patient harm (3–5). Evidence suggests that one in ten patients come to harm as a result of healthcare, and almost half are considered avoidable (6). However, understanding of staff safety is far more limited (7). Staff safety is pertinent in the current climate within the NHS, which is facing a growing shortage of overall staff (8,9), and registered nurses in particular (10–13). Safety measures within healthcare have historically related to the patient. The collection of data on ‘violence and aggression’ in the Mental Health Safety Thermometer is recognition of this being a common harm experienced in this care setting (14). However, the metric does not distinguish between harm to other patients and harm to staff. There has been some qualitative work that highlights the concept of the absence of threat of harm to staff as being essential for nurses to be able to work effectively in the inpatient psychiatric setting (15), but there is little published research quantifying staff harm and exploring its relationship with staffing.

Of the limited research that is available regarding staff harm and safety measures, it has been found that feelings of safety correlate with experiences of stress and burnout. Evidence suggests that concern for personal safety is a high stressor and can contribute to burnout (16). Given that higher burnout has been linked with staff being more likely to view their work environment as ‘unsafe’ but reduces the likelihood of ‘near misses’ being reported (17), staff wellbeing is important in ensuring both an adequate workforce and being able to accurately capture the state of safety in healthcare.

The NHS currently utilises a number of operational databases for adverse events reporting and management of staffing levels, which collate routine patient and operational data that could be used for such purposes. The inherent availability of large, routinely collected datasets (18,19) alongside the emerging power of machine learning and knowledge-driven decision making tools (20), has been championed as a promising area for improving healthcare services (21,22). Indeed, the overlap between patient safety and staffing has been well established utilising such datasets (23–27).

This study utilises existing data from a large English mental healthcare provider of both acute and community services, focusing on the analysis of reported incidents to examine how such data can be used to monitor and predict where staff safety may be at risk. We examine the extent to which the perceived demand for registered and unregistered nurses, as well as the variations from the ‘clinically required’ levels, are correlated with increases in staff adverse event reporting. By utilising the clinically-perceived requirements in place of the pre-planned levels the analysis better reflects the shift by shift requirements.

This study utilises anonymised, routinely collected administrative data and therefore was not subject to NHS ethical approval HRA algorithm. Institutional ethical approval was sought and granted by the University and access granted by the participating trust’s research and development department following protocol review.

## Data selection criteria

Two databases were extracted from a large English mental healthcare provider in England providing both acute and community services – one containing safety incidents and one containing nurse staffing data. The nurse staffing dataset comprised a single data table detailing the ward-level nursing complement on each shift (early, late, and night) by nurse type (registered or unregistered) specifically for inpatient areas. Additionally, the dataset comprised the ‘planned level’ and ‘clinically required level’ of nurse staffing. The planned level was determined annually in line with available budgets, whereas the clinically required level allowed clinical staff to report additional staff requirements in accordance with demand and clinical judgements for safe staffing. The incident reporting system is a commonly used system for reporting adverse events comprising a web-form for data capture and SQL engine for data storage. The data storage element comprised two key data tables – one detailing the incident (including location, severity, likelihood to repeat, and date) and the other detailing those involved (patient vs staff, victim vs perpetrator vs witness, age). The database was initially aggregated to day and ward level, pooling adverse events to find the total daily rate. This selection was then aligned with the trust’s staffing database by date and ward ID.

The co-variates of interest selected from the staffing database were the clinically-required staffing level, and the variation of actual staffing from the required levels for each of the three shifts (referred to as the $Delta\_{Shift}^{Staff type}$ terms). This clinically required level showed a change (either more or less) from the planned level on 37% of the ward-shift pairings (and 72% of ward-days having at least one shift recommending a change). These clinically required levels have the advantage of reflecting both the perceived demand, combining case complexity and ward activity, and the confidence of the staff in delivering safe care.

The variation from the clinically-required staffing levels were used instead of the absolute staffing level to reduce the collinearity of the covariates. Where the level of staffing is less than the clinical demand, we can infer a situation in which staff members must optimise their time, with a greater risk of tasks going undone or work being rushed. Hence, where staffing levels are less than the demand requires we could expect to see low priority tasks (such as the reporting of near misses) going undone.

The other confounding variable included for analysis was the geographical location of the ward. The data from this trust was tagged within the incident reporting database for one of 10 geographical locations, which have been pseudonymised for this analysis. A brief description of the specialties and characteristics of each location are supplied in Table 1.

**Table 1: Location descriptions.**

|  |  |
| --- | --- |
|  | Code |
| Loc1 | Low security unit (Men) |
| Loc2 | Hospital with high security wards |
| Loc3 | Inpatient Adult ward |
| Loc4 | Adult working age  |
| Loc5 | Older people |
| Loc6 | Older people  |
| Loc7 | General in-patient including PICU |
| Loc8 | Medium security unit |
| Loc9 | Community Inpatient(autism and substance abuse assisted living) |
| Loc10 | Long Stay rehab |

Each feature of the constructed dataset was inspected as histograms in order to remove extreme outliers, which were then classified as miss-entered data, constituting 1.2% of the extracted data. The analysis subsequently addressed the relationship between the variation of nurse staffing (by shift) and perceived clinical demand on the rate at which members of staff were reported as victims of adverse events. The definition of ‘victim’ arises from the event reporting system where each individual is flagged for their involvement e.g. ‘victim’, ‘perpetrator’ or ’witness’. To allow for variation in ward sizes, the rate of reports was corrected for the total number of nurses on shift for a given ward and day. The analysis hence models the number of staff reported as victims of an adverse event per nurse on shift as a function of perceived clinical demand, and variation in staffing from the perceived clinical demand with location ID serving as a proxy for safety and reporting culture.

## Method

The data were analysed via a selection of count-based models including Poisson, zero-inflated Poisson, negative binomial and hurdle regression (28). The Poisson regression is the typical form of general linear model to apply to count data, with the others dealing with variations from the basic Poisson distribution. The negative binomial regression is designed to model count data with over-dispersion (i.e. larger error variance than allowed for by Poisson regression), while the hurdle and zero-inflated regression models count data that contains more zeros than is typical of the Poisson process. Each analysis routine used the well-documented implementations in the R language in the PSCL (29) and MASS (30) packages. Model comparisons were made using the Bayesian Inference Criteria (BIC) in which a smaller BIC score is indicative of a superior model (31).

Subset selection was carried out via LASSO-regularised regression routines (32), in particular the Poisson and Binomial-family glmnet algorithms implemented in R by the glmnet package (33). The LASSO-regularisation approach aims to extract the smallest subset of features that explain the model without the reduced error bars inherent in stepwise selection (34). The regularisation parameter λ was tuned via 10-fold cross validation (CV), and model coefficients are reported for the “one standard error” λ term, $λ\_{1SE}$. Using $λ\_{1SE}$ has been suggested to result in a simpler model, minimizing over-fitting error, and hence give the more parsimonious explanation of the system (33,35).

The Expectation-Maximization (EM) algorithm (36–38) used in the study makes use of the ‘emax.glm’ R implementation. The data analysis was limited to two competing Poisson regression models for simplicity, using 20 randomly selected starting conditions to perform early ending fits in order to explore the parameter space and check for an optimal starting point.

## Results

The aggregated dataset analysed comprised 40,123 observed days (total) divided between 51 wards and 10 locations (see Table 1 for descriptions). The data were recorded over three years as part of the routine function of the trust with the earliest observations taken on the 1st September 2014 until the 31st March 2017. Within the study there were 10,119 events reported, accounting for 19,693 members of staff being the victim of an adverse event, inclusive of near misses and non-harm incidents. The events reported have a raw prevalence of 0.252(± 0.004) events reported per ward per day and 0.491(± 0.009) members of staff a victim of an event per day per ward.

Prior to analysis, the rate of adverse events where staff were victims was first characterised before removing outliers. The majority of day-ward combinations resulted in no adverse events being reported, accounting for 84% of the dataset. Where events were reported, each was tagged for an incident type. Aggregating by incident type the most prevalent events were “Aggression by Patient
on Staff or Other” (6,520 events reported with 12,138 members of staff as victims), “Inappropriate Behaviour” (1,762 events reported with 4,058 members of staff as victims), “Self Harm” (429 events reported with 929 members of staff as victims) and “Sexual Incidents” (298 events reported with 493 members of staff as victims).

Initial trials of the Poisson model (BIC = 74,295) demonstrated evidence of over-dispersion, with the deviance residuals showing non-normal behaviour on a Q-Q plot (see Figure 1 A). The model dispersion was estimated at 3.07 via auxiliary OLS regression (with an associated p-value below numeric precision) implying over-dispersed errors. Trials of the dispersion-corrected count models showed an improvement in the BIC scores, with the negative binomial giving the optimal model (BIC = 52,726) compared to the zero-inflated (BIC = 56,928), and hurdle (BIC = 56,916) models. The negative binomial model resulted in a sizeable reduction in the BIC score in comparison to the Poisson model, indicating the reporting of adverse events is better explained by the negative binomial model.

Despite the improvement in BIC score, inspection of the negative binomial deviance residuals reveals a distinct structure. Notably, the model shows a bias to under-estimating larger values (Figure 1 B) and a strong discontinuity in the Q-Q plot (see Figure 1 A). This behaviour may well be explained by having two competing processes – one driving the inflated zero counts, and the other the expected Poisson behaviour.

**Figure 1. Comparison of residuals for Poisson and Negative Binomial regression.**

Two competing behaviours is relatively common, and should in theory be well modelled via the hurdle or zero-inflated models. However, if the variables being modelled do not account for which of the two behaviours dominates, these models will have only limited performance. As an alternative, we turn now to consider how well a latent variable model, such as the EM algorithm, explains the data.

Having fit the EM algorithm, the data were found to be well divided between two models. Dividing the data at a model-probability of 0.5, the two models show remarkably different average rates of adverse events per nurse on the ward of 0.175 (± 0.002) and 0.00087 (± 0.00004) for 5,908 and 34,079 observations respectively. With these values in mind, it appears the system is divided between adverse events being reported at a moderate rate, and at a low rate. We hence adopt the term “Moderate Reporting Model” (MRM) and “Low Reporting Model” (LRM) to describe the two competing scenarios.

To explore how well we can predict which model applies to a new observation, the LRM probabilities arising from the EM algorithm were fit using LASSO-logistic regression. By aggregating the data into two groups, LRM vs MRM, the regression becomes a binomial regression problem and is hence referred to as the Binomial Model (BM). The initial BM, performed using the same variables as the EM algorithm, showed some predictive power (‘Receiver Operator curve – area under curve’ score (*roc-auc*) = 0.7474). The *roc-auc* score takes values between 0.5 (unable to predict the correct model) and 1 (perfect prediction of the correct model), with a value of 0.75 suggesting only moderate predictive power.

To improve the predictive power of the BM, the covariates were expanded to include the last reported rate of adverse event on each ward. The addition of the extra term dramatically improved the *roc-auc* score (0.9983, near perfect prediction). This suggests that the model division learnt through the EM algorithm may be well predicted for new observations, not merely learnt retrospectively, and is highly dependent on the previous reporting rate. The parameter values for the optimal BM are given in Table 2 and the odds-ratio confidence-intervals for the staffing parameters are included in Figure 2 where a positive parameter value indicates a preference for the LRM.

**Table 2: Location parameters for the three models (with 95% confidence intervals and significance level)**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Binomial Model | Moderate Reporting Model | Low Reporting Model |
| Loc1 | 0.341 [0.0924, 1.17] (-) | -1.75 [-1.95, -1.47] (\*\*\*) | 0 [0, 0] (-) |
| Loc2 | 0 [-0.0808, 0.0808] (-) | -1.03 [-1.23, -0.749] (\*\*\*) | 0 [0, 0] (-) |
| Loc3 | 0.769 [0.352, 2.32] (-) | -1.08 [-1.62, -0.471] (\*\*\*) | 0 [0, 0] (-) |
| Loc4 | 1.6 [1.33, 2.63] (\*\*\*) | -1.46 [-1.66, -1.18] (\*\*\*) | 0 [0, 0] (-) |
| Loc5 | 1.97 [1.74, 2.9] (\*\*\*) | -1.39 [-1.6, -1.1] (\*\*\*) | 0 [0, 0] (-) |
| Loc6 | 1.26 [0.947, 2.38] (\*\*) | -1.48 [-1.69, -1.2] (\*\*\*) | 0 [0, 0] (-) |
| Loc7 | 0.931 [0.643, 1.89] (\*\*) | -1.47 [-1.69, -1.16] (\*\*\*) | 0 [0, 0] (-) |
| Loc8 | 0.299 [0.0703, 1.05] (-) | -1.58 [-1.8, -1.27] (\*\*\*) | 0 [0, 0] (-) |
| Loc9 | 2.39 [2.17, 3.25] (\*\*\*) | -1.42 [-1.64, -1.14] (\*\*\*) | -0.87 [-1, -0.745] (\*\*\*) |
| Loc10 | 2.96 [2.77, 3.85] (\*\*\*) | -1.37 [-1.6, -1.09] (\*\*\*) | 0 [0, 0] (-) |
| prev\_rate | -169 [-185, -161] (\*\*\*) |  |  |

N.B.:

|  |  |
| --- | --- |
| \* | significant at the 0.05 level |
| \*\* | significant at the 0.005 level |
| \*\*\* | significant at the 0.0005 level |

**Figure 2: Bootstrapped 95% Confidence intervals for the staffing parameters across the BM, MRM, and LRM.**

****

**Binomial Model Interpretation**

The Binomial Model (BM), which describes the likelihood of an observation belonging to the “Low Reporting Model” (LRM), shows a strong dependency on the location. Four locations (1, 2, 3 and 8) have no significant parameter, implying a baseline 50% probability of being in either model, if demand is not considered. Of the other locations there is a steadily growing likelihood of the ward following the LRM with Locations 6 and 7 having the lowest chance (75%), and Locations 9 and 10 the highest (93%).

Across the trust, the best predictor for which model will dominate (LRM or MRM) is the rate at which adverse events were previously reported. In general, the more events are reported by a ward on the previous day, the greater the probability the ward will remain in the MRM. The level of reporting needed varies, with Location 10 requiring 0.017 events reported per nurse on shift to give even odds of either model. To put this into context, across the dataset, the average rate of reporting is 0.0265(± 0.0005) events reported per nurse, so each location can attain the MRM scenario.

The clinically-required staffing levels show a marked effect on the BM as well. As the clinically-required levels of registered nurses (RNs) increases (i.e. as the demand for more clinical nursing skills increases) there is a growing preference for the LRM. Hence, a greater registered nurse demand appears indicative of a decrease in adverse events being reported. Considering this dataset includes near misses, this is not overly surprising. As workload increases there is a distinct possibility that staff will prioritise delivering care over reporting near misses or adverse events. The exception to this is the night shift for unregistered nurses, where a greater perceived demand leads to a preference for the MRM.

The delta staffing variables (variation from ‘clinically-required levels’) suggest a shift to the MRM when the number of RNs exceeds the clinically-required levels on the late and night shifts, and a shift to the LRM when the number of unregistered nurses exceeds the clinically-required levels on the early shift. The increase in reporting with more RNs is unsurprising, considering that the more staff members that are present, the easier it is for events to be observed and reported. The lower reporting with more unregistered nurses than required may be due to the division in responsibilities between registered and unregistered nursing staff – it is not merely a question of higher headcount but the nurse’s professional practice and education as well.

The strong dependency of the BM on location and previous rate of reporting suggests there is a cultural aspect to the divide between the MRM and LRM scenarios. Interpreting what a low reporting of adverse events and near misses means is complex, and we consider two competing hypotheses - a reduction either occurs because of less adverse events occurring, or because fewer adverse events are reported. If the later of these is true, and the variation in reporting is due to changes in reporting culture, we would expect these locations to be biased towards reporting higher severity incidents. The result of this would be a decrease in the relative proportion of “near miss” and “no harm” events compared to higher severity incidents. Table 3 summarises the relative event severity across the 10 locations, and we observe that three of the four locations with the highest rate of ‘no harm’ events reported (locations 1, 2 and 8) are amongst the three locations most favouring the MRM. It appears that a higher rate of reporting, as implied by favouring the MRM, is linked to a better reporting culture, as implied by the increase in “no harm” events, and hence the MRM should be considered the preferred behaviour we wish to instil on a ward.

**Table 3: Event severity by location.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | No harm (%) | Low (%) | Moderate (%) | Sever (%) |
| Loc1 | 97.8 | 2.0 | 0.2 | 0.0 |
| Loc2 | 94.8 | 4.2 | 0.8 | 0.2 |
| Loc3 | 83.3 | 16.7 | 0.0 | 0.0 |
| Loc4 | 89.7 | 10.0 | 0.3 | 0.0 |
| Loc5 | 83.8 | 16.0 | 0.1 | 0.0 |
| Loc6 | 87.7 | 11.7 | 0.5 | 0.1 |
| Loc7 | 89.8 | 9.6 | 0.7 | 0.0 |
| Loc8 | 90.5 | 9.3 | 0.2 | 0.0 |
| Loc9 | 92.5 | 6.7 | 0.9 | 0.0 |
| Loc10 | 85.7 | 14.3 | 0.0 | 0.0 |

**Moderate Reporting Model Interpretation**

Assuming a ward’s event reporting follows the MRM, the rate at which events are reported is reasonably consistent across locations. The only exceptions are: location 1, which shows the lowest baseline rate; and locations 2 and 3, which show the highest. Considering this model is consistent with a good reporting culture, we can assume the shift in rate is due to shifts in prevalence of adverse events. Location 1 being a low security location while location 2 being a high security location could be the reason for this shift in inherent safety of the location.

In the MRM scenario, the rate of reporting appears to be weakly linked to the clinically-required staffing levels. The only significant parameter is the required level of unregistered nurses on the night shift, where each extra unregistered nurse required reduces the rate of incidents reported by 15%. Fewer events being reported as the perceived demand grows implies more events reported at the lowest perceived demand. Considering that the night shifts generally run with the fewest staff members (between 1 to 2 RNs and 0 to 6 unregistered nurses) where perceived demand is lowest, the risk of very low numbers of staff on shift would inherently increase. It appears that even a small increase in unregistered nurses on the late shift above what clinical staff perceive as being required can result in a strong increase in staff safety.

The reporting of adverse events in the MRM scenario also appears to be only weakly linked to variation in staffing. The only two significant terms show a reduction in adverse events on staff being reported as the levels of RNs on the early and late shifts exceed the clinically-required levels. Given the assumption these observations arise from wards with a good reporting culture, the remaining explanation is that more RNs result in a safer environment for staff.

**Low Reporting Model**

Assuming a ward’s event reporting follows the LRM, the rate at which events are reported is relatively consistent across locations, with only location 9 showing a strong decrease in rate (58% decrease) compared to the others.

The LRM has the majority of its behaviour arising from the staffing parameters, with a strong link between increasing demand and decreased reporting on the night shift for RN staffing, and all shifts for unregistered nurses, though the signal arising from the night shift is by far the strongest. This supports the idea that as wards increase in demand, event reporting becomes increasingly rare, with staff members likely more focussed on the delivery of care than filling out near misses. However, the marked increase in signal for the night shift suggests that this may be in addition to the signal suggested for the MRM. A decrease in events being reported with more perceived demand means far more events reported when demand is low. Where the level of unregistered nurses on the night shift is low, the risk of low staffing levels would be enhanced, and more obvious accidents occur (e.g. of a higher severity than near miss).

The reporting of adverse events in the LRM scenario shows a similar link to variation in staffing as the MRM. The only two significant terms show a reduction in staff adverse events being reported as levels of RNs on the early and late shifts exceed the clinically-required levels. Whether this is a result of improved safety culture or indicative of something else is hard to say.

## Discussion

The marked improvement in BIC as the model moves from the standard count behaviours to the mixed Poisson behaviour suggests that the reporting of staff adverse events is relatively complex. There is strong evidence that the trust’s rate of staff being victims of adverse events is not continuous across the locations operated, with some locations reporting far fewer incidents than others. This variation in reporting appears to go hand in hand with an improved culture of reporting “no harm” events where a greater number of incidents are reported, and suggests that of the two models extracted, the “Moderate Reporting Model” represents a desirable culture.

Across all locations, there is strong evidence that the clinically-required level of unregistered nurses on the night shift is inversely proportional to the rate at which staff members are reported as the victims of adverse events. Both LRM and MRM scenarios suggest that as the clinically-required staffing levels of unregistered nurses decreases, the rate at which staff members are recipients of adverse events increases. The extent to which ward staffing varies from these required levels is relatively small, in general falling between 1 over or under the required level, and hence a low required level is indicative of a greater risk of very low staffing levels, which could explain why the number of adverse events on staff increases where the locations have a lower clinically-required level of staffing.

Considering the majority of adverse events suffered by staff members arise from aggressive/ inappropriate behaviour, and sexual incidents, the combination of low clinical demand and a greater risk of very low staffing levels establishes a clear image of a high risk environment for staff members. Such a risk is becoming increasingly well-understood as the growing demands on community nursing leads to higher lone working (39), however, the safeguarding needs may be unlikely to transport easily across disciplines.

The greater risk of harm that may be associated with very low staffing levels may also be countermanded by an increase in registered nurse staffing. The analysis shows that RN levels exceeding those recommended by clinical judgement on the early and late shift were linked to a decrease in adverse events on staff. Interestingly, equivalent behaviour on the night shift was not observed, although this may be due to the relatively low variation in RN staffing on the night shift in the dataset (83% of shifts running at the correct level with 12% running one under and 4% running one over). Alternatively, the increase in RN levels on the previous shifts may result in fewer tasks going undone and reducing the pressure on the night shift, resulting in a safer environment (40).

An increase in RN numbers has already been linked to improved patient outcomes (26,41), however, the effect of increased nursing levels may also lead to greater nurse retention. Considering the reported links between increased nurse stress with poor retention (42), and a poor perceived safety culture with burnout (17), it is foreseeable that staffing above the level set from clinical demand may have compounded benefits. If the wards considered to have the lowest clinical-requirements received just a small increase in RN levels, it may lead to not only fewer staff members experiencing harm, but potentially reduce staff turnover and improve the safety culture. Clinical judgement is clearly of use in determining staffing levels, however it needs to be complimented with operational considerations. Increasing the use of RNs is a growing ‘silver bullet’ by which we can fix the pressures faced by the NHS, but increasing demand does not fix the issue of supply (43,44).

Although one aspect of this study reveals that to increase staffing could improve staff safety, in reality, the system is far more complex than this and will unlikely be resolved by a single change. It is intriguing that shifts in unregistered nurse levels showed a far weaker link to the rate of events reported, suggesting that this isn’t a ‘bodies on the ground’ problem but rather something to do with the attributes of the registered nurse. The dynamics involved in the skill mix of a workforce are not linear, and safety cannot be improved by simply increasing the proportion of the unregistered workforce. Furthermore, increasing staffing indefinitely does not yield respective increases in safety indicating other factors are at play. Griffiths et al. (44) highlighted that wider leadership and environment factors played a significant role, alongside having the right numbers of staff. The absence of real-terms increases in healthcare budgets in the context of growing demand, during a time of austerity and government instability compounds the issue of growing demand, both in volume and complexity, and a solution to this isn’t imminent.

The exploration of this incident reporting system for staff harms is novel; traditionally, research in this area focuses on harm to patients. However, to not consider the implication of more than 10,000 reported incidents, resulting in nearly 20,000 staff victims over 32 months in the context of nurse retention would be a missed opportunity. Data warehoused in incident reporting systems are rarely considered in aggregate form, where the scale of incidents can be truly understood, and therefore actioned. Instead, organisations tend to deploy root cause analysis techniques to individual events that meet a particular threshold (typically severity), and this threshold may or may not include harm to staff. Given that it is known that poor staff safety is linked to burnout(16) and that there is a cyclical effect of burnout on perceived poor safety (17), it is important for organisations to reflect on the available intelligence housed in their data warehouses to ensure safe working environments for their staff.

Finally, when considering the usefulness of this analysis and the underlying role of routinely collected data within healthcare as a whole, we should reflect on the human factor. It has been shown that transparency within an organisation, when it comes to decision making, is known to have a positive impact on reducing staff absenteeism, and increasing staff retention rates. The databases that underpin this analysis have analogues in all NHS trusts, of varying quantity and quality, which have often remained as isolated silos of knowledge and yet it is feasible to mine them for insight. The data may contain inherent biases and validity issues, which will only become clear as the data is explored (45) but it is only as the value of the data to patient care is demonstrated to frontline staff that we can expect compliance to improve.

There are opportunities in other methodologies to explore the reasons behind the behaviours exhibited in these datasets, such as qualitative research methods. This sort of research could help generate hypotheses for areas of improvement, which could then be subjected to experimental design interventions to improve staff harm.

## Conclusion

This study demonstrates a method for modelling routinely collected data from a single mental health care provider in the NHS to explore the extent to which clinical judgement of demand and more notably variations from these clinically-required levels, align with adverse event reporting. There appears to be a consistent theme of reporting adverse events across the trust’s separate geographic locations, with different lessons to be learnt depending on location, ward acuity and shift.

The models suggest that the greatest risk to staff is present when night shift staffing levels are at their lowest, possibly as a result of very low staffing levels. This risk might be balanced by the addition of an RN, but does point to a limitation of clinical judgement. The required level to deliver care may be less than that required to protect staff. Establishing a good safety culture is inherently difficult, however, the reporting of no harm events does appear to be a key marker of culture. Once staff members have the resources to deal with the recognised clinical demand they appear to have better confidence in taking the time to report events, allowing for a trust to learn from near misses before they grow to actual harm.

The variation in unregistered nurse levels from the clinically-required levels showed the least effect on the adverse event rate, yet the inherent rate at which these staff members are required is linked. It appears that clinical staff recognise the task required of unregistered nurses but the wards show high resilience to under-staffing, while over-staffing gives little benefit. It is not only a question of increased headcount, but of the skill-set and expertise they bring to the environment.

Finally, the way in which we value healthcare staff has been brought to light here. Traditionally, the concept of safety is prepositioned by ‘patient’ and so safety in healthcare is conceptualised as patient safety; little consideration is given to the safety of the staff delivering care. In a time where workforce retention is proving difficult and the quality and safety of care to patients is compromised, greater efforts should be made to improve staff safety, which might improve retention and simultaneously patient safety.

The methods presented represent a flexible approach to mobilise knowledge from the growing silos of routinely collected data in order to shape local safe staffing practices. They demonstrate that the analysis of such data via the application of data mining techniques represent a currently untapped opportunity to improve staff safety in healthcare.

Data is not currently available in a suitable format for open access. If you would like to request access to the data, please contact the named authors directly.

## References

1. Institute of Medicine (US) Committee on Quality of Health Care in America. To Err is Human: Building a Safer Health System. Kohn LT, Corrigan JM, Donaldson MS, editors. Washington (DC): National Academies Press (US); 2000.

2. WHO. World Alliance for Patient Safety : forward programme 2005. Geneva : World Health Organization; 2004.

3. NHS England. Better Births: Improving outcomes of maternity services in England – A Five Year Forward View for maternity care. 2016.

4. Francis R. Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry - Executive Summary. London: H.M.S.O.;

5. National Advisory Group on the Safety of Patients in England. A promise to learn – a commitment to act – Improving the Safety of Patients in England. 2013 p. 46.

6. Vries EN de, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. BMJ Qual Saf. 2008 Jun 1;17(3):216–23.

7. Wagner A, Rieger MA, Manser T, Sturm H, Hardt J, Martus P, et al. Healthcare professionals’ perspectives on working conditions, leadership, and safety climate: a cross-sectional study. BMC Health Serv Res. 2019 Jan 21;19(1):53.

8. Oliver D. David Oliver: Falling immigration could destroy the NHS. BMJ. 2019 Jan 2;364:k5308.

9. Wray J, Aspland J, Gibson H, Stimpson A, Watson R. ‘A wealth of knowledge’: A survey of the employment experiences of older nurses and midwives in the NHS. Int J Nurs Stud. 2009 Jul;46(7):977–85.

10. Finlayson B, Dixon J, Meadows S, Blair G. Mind the gap: the extent of the NHS nursing shortage. BMJ. 2002 Sep 7;325(7363):538–41.

11. Nardi DA, Gyurko CC. The global nursing faculty shortage: status and solutions for change. J Nurs Scholarsh Off Publ Sigma Theta Tau Int Honor Soc Nurs. 2013 Sep;45(3):317–26.

12. Janiszewski Goodin H. The nursing shortage in the United States of America: an integrative review of the literature. J Adv Nurs. 2003 Aug;43(4):335–43.

13. Toh SG, Ang E, Devi MK. Systematic review on the relationship between the nursing shortage and job satisfaction, stress and burnout levels among nurses in oncology/haematology settings. Int J Evid Based Healthc. 2012 Jun;10(2):126–41.

14. Power M, Stewart K, Brotherton A. What is the NHS safety thermometer? Clin Risk. 2012;18(5):163–169.

15. Kanerva A, Lammintakanen J, Kivinen T. Nursing Staff’s Perceptions of Patient Safety in Psychiatric Inpatient Care. Perspect Psychiatr Care. 2016 Jan;52(1):25–31.

16. Burnard P, Edwards D, Fothergill A, Hannigan B, Coyle D. Community mental health nurses in Wales: self-reported stressors and coping strategies. J Psychiatr Ment Health Nurs. 2000 Dec;7(6):523–8.

17. Halbesleben JRB, Wakefield BJ, Wakefield DS, Cooper LB. Nurse burnout and patient safety outcomes: nurse safety perception versus reporting behavior. West J Nurs Res. 2008 Aug;30(5):560–77.

18. Lugg-Widger FV, Angel L, Cannings-John R, Hood K, Hughes K, Moody G, et al. Challenges in accessing routinely collected data from multiple providers in the UK for primary studies: Managing the morass. Int J Popul Data Sci. 2018 Sep 12;3(3).

19. Harris RJ, Harris H, Mandal S, Ramsay M, Vickerman P, Hickman M, et al. Monitoring the hepatitis C epidemic in England and evaluating intervention scale-up using routinely collected data. J Viral Hepat. 2019 Jan 21;

20. Hatton CM, Paton LW, McMillan D, Cussens J, Gilbody S, Tiffin PA. Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare. J Affect Disord. 2019 Mar 1;246:857–60.

21. Krumholz HM. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. Health Aff Proj Hope. 2014 Jul;33(7):1163–70.

22. Bates DW, Heitmueller A, Kakad M, Saria S. Why policymakers should care about “big data” in healthcare. Health Policy Technol. 2018 Jun 1;7(2):211–6.

23. Needleman J, Buerhaus PI, Stewart M, Zelevinsky K, Mattke S. Nurse Staffing In Hospitals: Is There A Business Case For Quality? Health Aff (Millwood). 2006 Jan 1;25(1):204–11.

24. Clarke SP, Donaldson NE. Nurse Staffing and Patient Care Quality and Safety. Agency for Healthcare Research and Quality (US); 2008.

25. Leary A, Cook R, Jones S, Smith J, Gough M, Maxwell E, et al. Mining routinely collected acute data to reveal non-linear relationships between nurse staffing levels and outcomes. BMJ Open. 2016 Dec;6(12):e011177.

26. Griffiths P, Ball J, Murrells T, Jones S, Rafferty AM. Registered nurse, healthcare support worker, medical staffing levels and mortality in English hospital trusts: a cross-sectional study. BMJ Open. 2016 Jan 1;6(2):e008751.

27. Leary A, Tomai B, Swift A, Woodward A, Hurst K. Nurse staffing levels and outcomes – mining the UK national data sets for insight. Int J Health Care Qual Assur. 2017 Apr 18;30(3):235–47.

28. Mullahy J. Specification and testing of some modified count data models. J Econom. 1986 Dec 1;33(3):341–65.

29. Zeileis A, Kleiber C, Jackman S. Regression Models for Count Data in R. J Stat Softw. 2008 Jul 29;27(1):1–25.

30. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth. New York: Springer; 2002.

31. Kuha J. AIC and BIC: Comparisons of Assumptions and Performance. Sociol Methods Res. 2004 Nov 1;33(2):188–229.

32. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Ser B Methodol. 1996;58(1):267–88.

33. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.

34. Thompson B. Stepwise Regression and Stepwise Discriminant Analysis Need Not Apply here: A Guidelines Editorial. Educ Psychol Meas. 1995 Aug 1;55(4):525–34.

35. Krstajic D, Buturovic LJ, Leahy DE, Thomas S. Cross-validation pitfalls when selecting and assessing regression and classification models. J Cheminformatics. 2014 Mar 29;6(1):10.

36. Marshall G, Henderson WG, Moritz TE, Shroyer AL, Grover FL, Hammermeister KE. Statistical methods and strategies for working with large data bases. Med Care. 1995 Oct;33(10 Suppl):OS35-42.

37. Stoica P, Selen Y. Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher. IEEE Signal Process Mag. 2004 Jan;21(1):112–4.

38. Graham JW, Hofer SM, Piccinin AM. Analysis with missing data in drug prevention research. NIDA Res Monogr. 1994;142:13–63.

39. Brennan W. Safer lone working: assessing the risk to health professionals. Br J Nurs. 2010 Dec 9;19(22):1428–30.

40. Ball JE, Griffiths P, Rafferty AM, Lindqvist R, Murrells T, Tishelman C. A cross‐sectional study of ‘care left undone’ on nursing shifts in hospitals. J Adv Nurs. 2016 Sep 1;72(9):2086–97.

41. Leary A, Punshon G. Determining acute nurse staffing, a hermeneutic review of an evolving science. BMJ Open [Internet]. 2019 Apr 1 [cited 2019 Mar 29]; Available from: http://researchopen.lsbu.ac.uk/3059/

42. Bartram T, Joiner TA, Stanton P. Factors affecting the job stress and job satisfaction of Australian nurses: Implications for recruitment and retention. Contemp Nurse. 2004 Oct 1;17(3):293–304.

43. Griffiths P, Maruotti A, Saucedo AR, Redfern OC, Ball JE, Briggs J, et al. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. BMJ Qual Saf. 2018 Nov 26;bmjqs-2018-008043.

44. Lawless J, Couch R, Griffiths P, Burton C, Ball J. Towards safe nurse staffing in England’s National Health Service: progress and pitfalls of policy evolution. Health Policy. 2019 Mar 19;

45. Verheij RA, Curcin V, Delaney BC, McGilchrist MM. Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. J Med Internet Res. 2018 May 29;20(5).