TITLE PAGE

Hospital Mortality and Trainee Experiences: How GMC Survey Findings Correlate with Summary Hospital-Level Mortality Indicator

Jack Barton (0000-0002-7376-4197), Daniel Richardson (0000-0001-8799-7509), Amber Corrigan (0000-0003-0636-0114), Ashvin Kuri (0000-0003-0125-4806), and Jonathan Round (0000-0002-7404-2006).

Authors

Dr Jack Charles Barton, MBBS. Academic Foundation Programme Year 2 Doctor. Critical Care and Perioperative Medicine Research Group, Royal London Hospital, Barts Health, London, United Kingdom, E1 1FR.

Dr Daniel Richardson, MBBS. Academic Foundation Programme Year 2 Doctor. King's College Hospital, Denmark Hill, Brixton, London, SE5 9RS.

Dr Amber Elyse Corrigan, MBBS. Academic Foundation Programme Year 1 Doctor. King's College Hospital, Denmark Hill, Brixton, London, SE5 9RS.

Mr Ashvin Kuri, MSc. Medical Student. Barts and The London School of Medicine and Dentistry, Garrod Building, Turner St, Whitechapel, London, E1 2AD.

Professor Jonathan Round, MBChB. Paediatric Intensive Care Consultant. Paediatric Intensive Care Department, St George's University Hospitals NHS Foundation Trust and St George's, University of London, London, United Kingdom, SW17 ORE.

Corresponding Author

Dr Jack Barton, MBBS. Critical Care and Perioperative Medicine Research Group, Royal London Hospital, Barts Health, London, United Kingdom, E1 1FR. jack.barton@nhs.net. 02035940353. No fax number available.

Funding Source

There was no funding source for this study.

Word Count of Full Text (exc. title page, abstract, references, figures, tables, and acknowledgements and placeholders) = 3672

ABSTRACT

Objectives

To investigate whether components of junior doctor satisfaction are associated with patient mortality within the United Kingdom.

Methods

We conducted a cross sectional study of publicly available data [the General Medical Council (GMC) National Survey and the Summary Hospital-level Mortality Indicator (SHMI)] pertaining to subjective physician trainee satisfaction and patient mortality within 80 United Kingdom based healthcare institutions. The direction and strength of correlation between components of the GMC National Survey and relative patient mortality as described by the SHMI were calculated. Additional outcomes included mean GMC survey scores for reported domains and mean SHMI by healthcare institution.

Results

SHMI for included healthcare institutions ranged from 0.69 to 1.21 (mean = 1.01; SD 0.1). Mean GMC domain scores ranged between 44.61 and 88.62 (mean = 71.16; SD = 10.84). Statistically significant correlations were observed for clinical supervision, clinical supervision out of hours, rota design, overall satisfaction, and teamwork. After application of Bonferroni correction statistically significant correlations remained for both clinical supervision and clinical supervision out of hours.

Conclusion

There is a significant association between components of subjective trainee satisfaction and patient mortality within the United Kingdom. Further investigation to examine these relationships, perhaps to target intervention, may prevent avoidable patient harm.

INTRODUCTION

Throughout this article we use the terms 'excess patient harm' and 'excess patient mortality' when referring to the analysed data and preceding literature. When doing so we are referring solely to harm above that which would be expected in an average National Health Service (NHS) hospital adjusted for the patient demographics of the hospital in question. We would like to emphasise that any harm caused by medical intervention is and should be considered excessive, however for the purposes of this analysis we are referring solely to relative incidence.

The 2018 'Caring for doctors Caring for patients' General Medical Council (GMC) review comprehensively outlined the implications of poor workplace satisfaction on patient safety, providing a stark warning for NHS providers.[1] The review not only suggested that workplace satisfaction was associated with patient outcomes, but that up to 50% of physicians within the United Kingdom (UK) were dissatisfied to a point that they were considering leaving their respective organisations.[1] Unfortunately, the COVID-19 pandemic has hindered an effective response to this review while also worsening physician dissatisfaction and mental health issues.[2]

Physician trainees provide both great insight into, and form a key component of, patient safety. However, although physician dissatisfaction is exacerbated by poor working conditions,[3] the data to support a direct association between this and worsened patient outcomes is limited in both quantity and quality. There are only a handful of high-quality analyses examining the influence of workplace satisfaction on patient outcomes such as harm and mortality .[4,5] Although the evidence base relating specifically to physician burnout and patient safety incidents is stronger. [6]

The GMC national training survey is an annual survey distributed to all physician trainees within the UK. Questions address 18 domains including supervision, education, training, and safety. It is used by arm's length bodies to highlight areas of concern for trainees,[7-8] as well as for quality monitoring and assessment. The uptake is high, most recently at 76% for UK trainees, so it is seen as a relatively accurate barometer of trainee satisfaction year on year.[9] The Summary Hospital-level Mortality Indicator (SHMI) reports the ratio of actual number of patient deaths during or within 30 days following hospitalisation, compared with expected deaths, calculated from mean figures for all NHS trusts and their patient population characteristics.[10] SHMI is calculated using hospital episode statistics and death registration data from the Office of National Statistics. In 2020 it was used to identify 12 NHS Foundation Trusts for potential intervention due to a higher-than-expected mortality rate. Both the GMC national training survey and SHMI provide departmental and hospital-level data respectively, allowing assessment of local, regional, and national relationships between the variables reported. Analysis of this data in combination is yet to be reported within the academic literature. One might logically conclude that such an endeavour would not be worthwhile, given the seemingly obvious connection between the variables reported within the GMC survey and patient mortality as reported by SHMI. However, this assumption is not borne out in evidence. Take physician experience for example, which outside of periods of trainee rotation [11] demonstrates no correlation with patient outcomes.[12] Measures of educational delivery such as curriculum development, regional and local teaching are associated with physician performance but not patient outcomes. [13] Procedural variables such as local induction and handover appear to demonstrate an association with patient mortality [14] but heterogeneity in design and delivery prevents widescale extrapolation of reported data. Commonly referenced overarching themes of physician satisfaction such as clinical supervision, teamwork, workplace culture, and physician workload are associated with patient outcomes, but to variable degrees depending on the healthcare institution and the population in question. [15-20] Thus, at a time of worsening workforce dissatisfaction within the context of increasing healthcare demand, it is of vital importance that we seek greater understanding of these relationships.

This review examines the relationship between SHMI and the 18 domains of physician trainee satisfaction reported within the GMC survey at a hospital-level. Given the available evidence we hypothesised that high levels of clinical supervision, clinical supervision out of hours, overall satisfaction, and teamwork would be negatively associated with excess patient mortality e.g. improved overall satisfaction within a hospital is likely to be associated with lower excess mortality. Whereas high reported workload would be positively associated with excess patient mortality. We hypothesised that the remaining 13 domains reported within the GMC survey would demonstrate no such association.

METHODS

Study design

A cross sectional study of two publicly available datasets.

Setting and Sample

A cross sectional study was conducted using two large publicly available datasets to retrospectively explore how the published SHMI for 80 healthcare institutions in the United Kingdom was correlated with variables reported in the GMC survey at the same institutions.

NHS digital defines the SHMI as a ratio of 'the actual number of patients who die following hospitalisation at the trust and the number that would be expected to die on the basis of average England figures, given the characteristics of the patients treated there'.

The 2018 GMC database contains 18 columns representing different 'domains' of experience, which the GMC and other related bodies identify as relevant to trainee experience, listed in Table 1. Data from the survey is generated through Likert responses to a series of questions such as "How would you rate the quality of experience in this post". Individual questions are collated to represent domains, and a scale (0-100) is generated for each with high scores indicating positive trainee responses, except for 'workload' where an ideal would score 50. For each year, variables are presented in columns whilst each row denotes a department in a healthcare institution. In 2018 the GMC dataset comprised 18 domains and 5718 rows (departments) in 490 different UK healthcare institutions: trusts, health boards, general practices or private businesses.

Data Collection

As part of a larger ongoing retrospective observational study, multiple years of GMC (2012-2019) and SHMI (2015-2019) data was downloaded from www.gmc-uk.org[9] and NHS Digital[10] respectively. SHMI data is collected and published on a monthly rolling basis. 2018 was the most recent year for which complete SHMI data was available at outset of the project. The 2018 GMC data was derived from trainee responses while the survey was open between 20th March 2018 and 2nd May 2018.

Data Analysis

Data presented here is an exploratory analysis or 'first look', however the entire data cleaning process of an ongoing project is described for completeness (Figure 1). Data was initially reorganised from individual year summaries to by-year-trends per domain. In the 2018 dataset, there were a possible 102924 datapoints, however 36478 (35.4%) of these were missing, giving a total of 66446 datapoints. In some cases, no data was available for entire departments, in other cases just a single datapoint was missing. Most commonly this is because no data is returned in the GMC survey if less than three respondents answer a question. All rows with missing data were then deleted so that each department retained had responses for all domains. A mean score for each GMC domain was then calculated for each healthcare institution.

Healthcare institutions with a score in every GMC domain were then matched to those with complete SHMI data from 2015-2019 producing 80 healthcare institutions with both complete GMC and SHMI data for analysis. The joint GMC/SHMI dataset was then further analysed.

INSERT FIGURE 1 HERE

Data processing was performed on Microsoft Excel (version 16.65), the final dataset was then transferred to Prism (version 9.1.2) for data analysis.

Descriptive statistics of the SHMI and each GMC domain were performed. Assumptions for normal distribution, linearity, homoscedascity and collinearity were met (see supplemental data sheet and supplementary figures). Pearson's correlation coefficients were calculated between matched healthcare institutions' SHMI and the corresponding GMC ratings for each domain across that healthcare institution. After an initial analysis, a post-hoc Bonferroni correction [21] was applied to achieve a more conservative alpha level (0.00278). In order to reduce the likelihood of a type I effort the decision was made to perform a Bonferroni correction given the multiple statistical tests performed during the exploratory analysis. However, given the evidence base guiding our statistical hypothesis testing, and the relative conservatism of the Bonferroni correction [22], significance levels both pre- and post-Bonferroni correction are presented to allow readers to interpret the data as they deem appropriate.

Given the data cleaning process resulted in the removal of SHMI data from 34 healthcare institutions, in an attempt to identify sources of bias, a student's two-tailed independent t-test was performed to ensure that the analysed group were not atypical. Q-Q plots and comparison of included vs. excluded GMC data and SHMI are also presented in supplementary figures A, B, and C.

Ethical considerations

Given that all data downloaded and analysed already exists within the public domain, ethical approval was not sought.

RESULTS

Healthcare Institution Characteristics

The final processed dataset represented trainee responses from 80 healthcare institutions. The data contained mean responses from 60 different specialties across primary and secondary care (Figure 2).

The mean (SD) SHMI for included healthcare institutions was 1.01 (0.1) with a range of 0.69 to 1.21 as demonstrated by Figure 3. Mean (SD) SHMI for excluded healthcare institutions was 0.98 (0.1) with a range of 0.73 to 1.19. There was no significant difference between those healthcare institutions included vs. those excluded due to incomplete GMC results (p=0.16).

GMC Domain Mean Scores

The mean (SD) GMC scores were 71.2 (10.84) with a range of 44.61 to 88.62. All mean scores for included GMC domains are displayed in Figure 4. Rota design had the lowest rating, with clinical supervision receiving the highest.

INSERT FIGURE 2 HERE

INSERT FIGURE 3 HERE

INSERT FIGURE 4 HERE

Correlation between 2018 GMC Domain and 2018 SHMI

Pearson correlation coefficients can be found in Table 1. Statistically significant correlations were observed for the GMC survey domains of clinical supervision, clinical supervision out of hours, rota design, overall satisfaction, and teamwork. After application of Bonferroni correction (alpha = 0.00278) statistically significant correlations remained for both clinical supervision and clinical supervision out of hours. Thus, lower reported scores for clinical supervision and clinical supervision out of hours was correlated with SHMI.

	r (95% CI)	R squared	p
Adequate Experience	-0·07 (-0·29 to 0·13)	0·005	0·516
Clinical Supervision	-0·34 (-0·51 to -0·13)	0·115	0·002**
Clinical Supervision Out of Hours	-0·47 (-0·63 to -0·28)	0·221	<0·001**
Curriculum Coverage	-0·1 (-0·31 to 0·12)	0·01	0·379
Educational Governance	-0·18 (-0·38 to 0·05)	0·031	0·118
Educational Supervision	-0·1 (-0·31 to 0·12)	0·01	0·386
Feedback	0·08 (-0·15 to 0·29)	0·006	0·506
Handover	-0·1 (-0·31 to 0·12)	0·01	0·377
Induction	-0·07 (-0·29 to 0·15)	0·005	0·517
Local Teaching	-0·1 (-0·31 to 0·13)	0·009	0·403
Regional Teaching	0·14 (-0·09 to 0·35)	0·019	0·226
Reporting Systems	-0·14 (-0·35 to 0·08)	0·02	0·208
Rota Design	-0·29 (-0·48 to -0·08)	0·086	0·008*
Satisfaction	-0·25 (-0·44 to -0·03)	0·061	0·027*
Study Leave	-0·18 (-0·39 to 0·04)	0·033	0·109
Supportive Environment	-0·17 (-0·38 to 0·05)	0·029	0·128
Teamwork	-0·32 (-0·51 to -0·12)	0·103	0·004*
Workload	0·02 (-0·2 to 0·24)	0·001	0·835

Table 1 – Pearson Correlation Coefficients of All 2018 GMC Domains with 2019 NHS Trust Summary Hospital-Level Mortality Indicator.

*Statistically significant without Bonferroni correction (p = 0·05)

**Statistically significant with Bonferroni correction (p = 0·0027778)

DISCUSSION

Physician trainees within the UK make up almost 50% of the hospital physician workforce, the majority of whom are also trainees on a recognised programme.[23] Their role in acute and out-of-hours settings renders them well placed to observe the workings of the NHS as it manages the most critically ill patients. The most widespread and consistent way that observations from trainees on recognised programmes are captured is the GMC survey. Through analysing the GMC survey alongside SHMI, a well-recognised and universal reporting tool for relative hospital mortality, we have shown that some aspects of the annual GMC survey correlate with mortality indices for healthcare institutions. Lower perceived overall clinical supervision and clinical supervision out of hours are particularly correlated with SHMI, with other areas showing weaker correlations. Importantly, we have also demonstrated that other, commonly referenced domains of physician satisfaction such as physician workload, have little such correlation. Of course, such observations imply no causality to the relationships reported, but given the existing evidence base one could make several suppositions with potentially important consequences on patient safety and healthcare delivery.

This is the first national, multi-centre study to demonstrate an association between clinical supervision and clinical supervision out of hours with mortality. The observation that clinical supervision is associated with patient outcomes is not new,[24-25]. Martin et al. [20] most recently summarised the international literature within the area, in doing so raising concerns regarding the external validity of available data given the heterogeneity in both definition and delivery of clinical supervision. Our data demonstrates a weak to moderate association. This supports the supposition that subjective perception of physician trainee supervision likely has implications on, or is influenced by, patient mortality. Several recent reviews have suggested that more direct or close supervision, particularly for surgical physician trainees, reduced the rate of complications and patient mortality.[19,24,25] However, extrapolating such data to suggest that greater intensity of clinical supervision is necessary for our more senior physician trainees is to neglect the potential training implications of doing so. It also fails to consider any potential reverse causality or confounding implicating the observed relationships. Tomlinson [26] insightfully discusses the need to separate educational and supportive supervision from more managerial and administrative components. Unfortunately, our data is not sufficiently granular to allow us to examine these relationships in such detail. Thus, we cannot at this stage comment on the direction of causality either through examination of our data in isolation, or when interrogating it alongside the academic literature.

We also observed a significant, weak association between rota design and SHMI, that was lost when applying Bonferroni correction. There is a well-established association between physician trainee working hours and medical error [27-28], yet the intricacies of rota design and implications on patient outcomes is poorly researched. A number of small-scale studies suggest that trainee perception of patient safety may improve with adaptations [29] but there is a paucity of supporting international literature. It is not unlikely that subjective trainee physician reports of rota design are confounded by other variables such as workload. It would be prudent when examining longitudinal data to adjust for said confounders.

Interestingly overall satisfaction was weakly associated with SHMI, with statistical significance lost after Bonferroni correction. It is challenging to contextualise these findings given that most published data focusses instead on components, or extremes, of overall satisfaction. For example, burnout is strongly associated with patient safety [6] suggesting that extreme lows of physician trainee satisfaction are detrimental to patient care. Yet we cannot say with confidence that there is a difference in patient outcomes when physician satisfaction is moderate instead of high. One may hypothesise that reverse causality is likely, with physician trainees exhibiting low satisfaction if their patient outcomes are worse. Our data does not support this hypothesis.

Evidence relating to the relationship between teamwork and patient mortality is relatively strong, particularly within surgical specialities.[16, 30, 31] Our data may provide further evidence as we demonstrate a weak association between perceived teamwork and SHMI, however the statistical significance of which was lost after Bonferroni correction. Examination of this relationship over multiple years data will allow us to confirm and/or contextualise this relationship further.

Our results do not show any relationship between several domains of the GMC survey and SHMI. While this might be unsurprising for subjective quality of teaching and education, the finding that perceived workload was unrelated to SHMI seems counterintuitive. This also contradicts multiple reports suggesting the opposite.[32] As workload tends to be the lowest rated component of the GMC survey, this literature has been used to guide workforce interventions in the NHS. Our findings suggest that perhaps these interventions are unlikely to have a widespread effect on mortality.

Notwithstanding the many limitations of this type of study (discussed below), the relationships highlighted provoke a response. As those completing this survey are those most closely working with the most unwell patients in the NHS at a time when the service has fewer professionals on duty, it is likely that aspects of their experience are related to the same factors that affect mortality. It would be tempting to suggest that lack of clinical supervision might reduce the quality of care delivered, hence increasing mortality. However if a patient were to die unavoidably, the attending team

may feel undersupported[33-34] and interpret this within the clinical supervision domain irrespective of objective supervision provided. Similar issues may link domains such as teamwork and rota design with SHMI.

One response would be to take these findings and use them to enforce changes in training conditions so as to improve patient safety, for instance ensuring 24 hour a day consultant presence. This would of course be inappropriate, as we are reporting an association. Also, few interventions within the workplace have been shown to be effective beyond improving the perceived intended parameter as opposed to affecting 'real-world' patient-centred outcomes. Most projects examining medical error[35] and redesigning rotas[29] have only explored subjective outcomes. The exceptions, linking rota design with error in particular, are largely not relevant to current UK working practices.[27-28] One such study unsurprisingly found that reducing working hours below 80 per week and removing 24 hour continuous shifts improved sleep and reduced inattention, diagnostic and other medical errors. Those exploring teamwork interventions have mainly reported improvements in process measures such as adherence to a checklist or time to CPR. Whilst there is significant evidence within surgical literature that checklists (such as the WHO checklist) are effective at reducing error,[36] it is unclear how translatable they are to other areas of medicine.

Evidence on the effect of interventions attempting to improve clinical supervision in a range of settings has been mixed. A large systematic review from 2012,[37] itself responding to changes in the US residency programme, found that while supervisors might change diagnoses or have different approaches to patient care or surgery, there were few instances where this led to a measurable change in patient outcomes. A more recent intervention study exploring differential levels of supervision found no difference in the rate of medical errors, although interns spoke less and felt less autonomous when more supervised.[38]

Increasing the availability or presence of senior clinicians especially 'out-of-hours' may improve perceived clinical supervision but may also have obvious as well as unintended consequences. Cost, workforce planning, job satisfaction and retention would be likely to suffer.[39-40] Consultant presence may inhibit the development of independent decision making, paradoxically making patient safety worse and creating a skill gap for the future.

Further understanding of the association between perceived clinical supervision and SHMI is urgently needed. This GMC domain is likely to be either reflective of a healthcare delivery issue or a marker of something driving both sub-optimal care and the feeling of poor supervision. An exploration of the relationships between actual supervision and perceived supervision, and what trainees value in supervision, especially out of hours will be important. Moving forward, a set of controlled interventions would be able to explore which best improved the supervision domain and patient safety outcomes. We acknowledge that any study is unlikely to demonstrate a clear immediate mortality reduction. There is also a need to investigate the same parameters in other staff groups, examining the subjective experience of doctors not currently within a training programme, nursing staff, or other members of the multi-disciplinary team. If the same findings are observed amongst other members of the multi-disciplinary team in trusts with high SHMI then one may argue that targeted investigation and intervention in such trusts may be needed.

LIMITATIONS

Despite concerns surrounding the validity of the GMC survey and SHMI as independent standalone metrics,[41-42] through compiling, averaging and comparing relative as opposed change to absolute data across these two datasets, inaccuracies may be minimised. However, reported findings are not without caveats. The GMC survey remains heavily influenced by uptake amongst departments and proportion of completion, and therefore the number of physicians it represents varies year on year. It is at best a subjective snapshot of rotating trainee satisfaction. With many physicians moving between healthcare institution, there is inevitably some spill-over of experience in previous healthcare institutions or rotations that may be difficult to quantify. This is despite the timing of the survey being in the middle of a placement for most specialities. The effect of this would be to dilute differences between institutions. Furthermore, one must consider the unique nature of the training pathway provided for physician trainees within the UK and system within which it is delivered. The external validity of our findings is debatable, even though they are in keeping with much of the international literature.

The process of cleaning and amalgamating the dataset also led to some unavoidable shortcomings. Firstly, it reduced the total number of physician responses by around 85%. Further, during the process of cleaning and amalgamating the data we were unable to weight the responses of individual departments, as there is no accessible information on the size of each department or number of respondents. To be included in the survey, and ensure confidentiality, a department must supply more than three respondents. Small departments with poor response rates are at risk of not being included. However, by calculating a mean score for whole institutions we have decreased the influence of trainees in large departments to the overall metric. Weighting departments in this manner may be a useful future endeavour, however the practicality of achieving this on a national scale, and risk of potential confounding likely outweighs the benefit.

The Bonferroni correction applied may also be considered a limitation, given the not unlikely possibility that it masked true associations between GMC variables and SHMI. Hence our decision to present the data as such within the manuscript. Readers are advised to consider this fact when interpreting the presented data.

Finally, the findings presented represent analysis of a single year's worth of data. Exploration of longitudinal data across different years may provide a useful means by which to consider trends and associated outcomes. There is also a prevalent need to attempt to identify, and control for, confounders within the dataset as best as possible. The results observed may simply reflect Simpson's paradox due to unknown and unaccounted for confounding variables. We believe that this data at the very least provides an impetus for further research into these variables, in addition to further interrogation of any causal relationship between the associations reported.

CONCLUSIONS

There appears to be a significant association between domains of subjective trainee satisfaction and standardised patient mortality within the UK. Perceived experience in domains such as clinical supervision and clinical supervision out of hours are particularly associated with inpatient mortality. Conversely some areas of the GMC survey that one may suspect to closely relate to outcomes, such as perceived workload, demonstrate no association.

These findings must be seen within the context of the many cofounding factors and previous approaches to address preventable patient mortality within the NHS. They provide an urgent stimulus for further research into the mechanisms of these relationships to identify targets for interventions and to prevent avoidable patient mortality.

Acknowledgements

With thanks to Professor Jaun Carlos Kaski and Dr Jeremy Isaacs for their feedback and critique of this manuscript.

Contributor Statement

All authors contributed to project conceptualisation, methodology, and manuscript development. Ms Corrigan, Dr Barton and Dr Richardson contributed to data collection. Dr Barton and Dr Richardson contributed to data curation, statistical analysis, and data visualisation. Professor Round supervised all stages of the project.

Dr Barton acts as project guarantor and accepts full responsibility for the work and/or the conduct of the study, has access to the data, and controlled the decision to publish.

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Competing Interests

All authors declare no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

REFERENCES

1. General Medical Council. Caring for doctors caring for patients. [GMC web site]. November, 2019. Available at: https://www.gmc-uk.org/about/how-we-work/corporate-strategy-plans-and-impact/supporting-a-profession-under-pressure/uk-wide-review-of-doctors-and-medical-students-wellbeing. Accessed January 12, 2022.
2. Scheepers RA, Boerebach BC, Arah AO, Heineman MJ, Lombarts K. A systematic review of the impact of physicians' occupational well-being on the quality of patient care. *Interventional Journal of Behavioural Medicine* 2015;22(6):683-98.
3. Brown M, Tucker P, Rapport F, et al. The impact shift of patterns of junior doctors' perceptions of fatigue, training, work/life balance and the role of social support. *Quality and Safety in Healthcare* 2010;19(6):e36.
4. Teoh KR-H, Hassard J, Cox T. Doctors' working conditions, wellbeing and hospital quality of care: A multilevel analysis. *Safety Science* 2021; 135:105115
5. Welp A, Meier L, Manser T. Emotional exhaustion and workload predict clinician-rated and objective patient safety. *Frontiers in Psychology* 2015;5:1573.

6. Hodkinson A, Zhou A, Johnson J, et al. Associations of physician burnout with career engagement and quality of patient care: systematic review and meta-analysis. *British Medical Journal* 2022;378:e070442.
7. Rimmer A. Employers must tackle high level of burnout among trainees, says GMC. *British Medical Journal* 2018;362:k2018
8. Rimmer A. Covid-19: Most trainees have faced disruption to their training, GMC survey shows. *British Medical Journal* 2020;371:m4093
9. General Medical Council. National training survey. [GMC web site]. January, 2022. Available at: https://www.gmc-uk.org/education/how-we-quality-assure/national-training-surveys. Accessed January 12, 2022.
10. NHS Digital. About the summary hospital-level mortality indicator (SHMI). [NHS Digital web site]. January, 2022. Available at: https://digital.nhs.uk/data-and-information/publications/ci-hub/summary-hospital-level-mortality-indicator-shmi. Accessed January 12, 2022.
11. Young JQ, Ranji SR, Wachter RM, et al. "July Effect": impact of the academic year-end changeover on patient outcomes. A systematic review. *Annals of Internal Medicine* 2011;155(5):309-15.
12. McAlister FA, Youngson E, Bakal JA, et al. Physician experience and outcomes among patients admitted to general internal medicine teaching wards. *Canadian Medical Association Journal* 2015;187(14):1041-48.
13. Cervero RM, Gaines JK. *Journal of Continuing Education in the Health Professions* 2015;35(2):131-38.
14. Delardes B, McLeod L, Chakraborty S. What is the effect of electronic clinical handovers on patient outcomes? A systematic review. *Health Informatics Journal* 2020;26(4):2422-34.
15. Hichtalk HJ, Yeh H-S, Pronovost PJ, et al. Impact of attending physician workload on patient care: a survey of hospitalists. *JAMA Internal Medicine* 2013;173(5):375-77.
16. Berry JC, Davis JT, Bartman T, et al. Improved safety culture and teamwork climate are associated with decreases in patient harm and hospital mortality across a hospital system. *Journal of Patient Safety* 2020;16(2):130-36.
17. Dewa C, Loong D, Bonato S, et al. The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: a systematic review. *BMJ Open* 2017;7(6):e015141.
18. Sexton JB, Schwartz SP, Chadwick WA, et al. The associations between work-life balance behaviours, teamwork climate and safety climate: cross-sectional survey introducing the work-life climate scale, psychometric properties, benchmarking data and future directions. *BMJ Quality and Safety* 2017;26(8):632-40.
19. Snowdon DA, Leggat SG, Taylor NF. Does clinical supervision of healthcare professionals improve effectiveness of care and patient experience? A systematic review. *BMC Health Services Research* 2017;28(4):447-55.
20. Martin P, Lizarondo L, Kumar S, et al. Impact of clinical supervision on healthcare organisational outcomes: A mixed methods systematic review. *PLoS One* 2021;16(11):e0260156.
21. Dunn O. Multiple comparisons among means. *Journal of the American Statistical Association* 1961;56(293):52-64.
22. VanderWeel TJ, Mathur MB. Some desirable properties of the bonferroni correction: is the bonferroni correction really so bad? *American Journal of Epidemiology* 2019;188(3):617-18.
23. The Nuffield Trust. The NHS Workforce in numbers. [Nuffield Trust web site]. May 8, 2019. Available at: https://www.nuffieldtrust.org.uk/resource/the-nhs-workforce-in-numbers. Accessed January 12, 2022.
24. Snowdon DA, Hau R, Leggat SG, Taylor NF. Does clinical supervision of health professionals improve patient safety? A systematic review and meta-analysis. *International Journal for Quality in Health Care* 2016;28(4):447-55
25. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: The effect of clinical supervision on patient and residency education outcomes. *Academic Medicine* 2012;87(4):428-42
26. Tomlinson J. Using clinical supervision to improve the quality and safety of patient care: a response to Berwick and Francis. *BMC Medical Education* 2015;15:103.
27. Lockley SW, Cronin JW, Evans EE, et al. Effect of reducing interns' weekly work hours on sleep and attentional failures. *New England Journal of Medicine* 2004;351(18):1829-37.
28. Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. *New England Journal of Medicine* 2004;351(18):1838-48.
29. Walton H, Uheba M. Improving patient safety by implementing a new general surgical on-call rota for FY1 doctors. *BMJ Open Quality* 2015;4:u206069.
30. Weller J, Boyd M, Cumin D. Teams, tribes and patient safety: overcoming barriers to effective teamwork in healthcare. *Postgraduate medical journal* 2014;90(1061):149-54.
31. Alsabri M, Boudi Z, Lauque D, et al. Impact of teamwork and communication training interventions on safety culture and patient safety in emergency departments: a systematic review. *Journal of Patient Safety* 2022;18(1):e351-61.
32. Jarman B, Gault S, Alves B, et al. Explaining differences in english hospital death rates using routinely collected data. *British Medical Journal* 1999;318(7197): 1515.
33. Moores TS, Castle KL, Shaw KL, Stockton MR, Bennett MI. 'Memorable patient deaths': Reactions of hospital doctors and their need for support. *Medical Education* 2007;41(10):942-6.
34. Redinbaugh EM, Sullivan AM, Block SD, et al. Doctors' emotional reactions to recent death of a patient: Cross sectional study of hospital doctors. *British Medical Journal* 2003;327(7408):185-9.
35. Ahmed M, Arora S, Tiew S, et al. Building a safer foundation: The Lessons Learnt patient safety training programme. *British Medical Journal Quality and Safety* 2014;23(1)78-86.
36. Russ S, Rout S, Sevdalis N, Moorthy K, Darzi A, Vincent C. Do safety checklists improve teamwork and communication in the operating room? A systematic review. *Annals of Surgery* 2013;258(6):856-71.
37. Farnan JM, Petty LA, Georgitis E, et al. A systematic review: The effect of clinical supervision on patient and residency education outcomes. *Academic Medicine* 2012;87(4):428-42.

38. Finn KM, Metlay JP, Chang Y, et al. Effect of increased inpatient attending physician supervision on medical errors, patient safety, and resident education: A randomized clinical trial. *Journal of the American Medical Association Internal Medicine* 2018;178(7):952-59.
39. Reid H, Wittkowski A, Vause S, Hezell A. 'Just an extra pair of hands'? A qualitative study of obstetric service users' and professionals' views towards 24/7 consultant presence on a single UK tertiary maternity unit. *British Medical Journal Open* 2018;8(3):e019977.
40. Burbeck R, Coomber S, Robinson S, Todd C. Occupational stress in consultants in accident and emergency medicine: A national survey of levels of stress at work. *Emergency Medicine Journal* 2022;19(3):234-8.
41. Woods S, Gough M. Online training surveys: Not worth the paper they're written on. *Bulletin of the Royal College of Surgeons of England* 2018;100(1):32-7.
42. Hogan H, Zipfel R, Neuburger J, Hutchings A, Darzi A, Black N. Avoidability of hospital deaths and association with hospital-wide mortality ratios: Retrospective case record review and regression analysis. *British Medical Journal* 2015;351:h3239.

Figure Legends

Figure 1 - Diagram of data reorganisation and cleaning process for 2018 GMC Survey and SHMI data.

Figure 2 – Number of Departments per Specialty Submitting GMC Data for included Healthcare Institutions. Median = 25; IQR = 61.25; min = 1 (allergy, audiovestibular medicine, community child health, occupational medicine, paediatric emergency medicine, paediatric immunology, infectious diseases and allergy, paediatric neurology, paediatric cardiology, sport and exercise medicine); max = 159 (anaesthetics).

Figure 3 – Frequency Distribution of 2018 SHMI by NHS Trust. Median = 1.00; IQR = 0.05; Min = 0.70; Max = 1.21. The SHMI data demonstrated a marginally negative skew with 38 of the included Healthcare institutions experiencing a SHMI <1 i.e. experiencing lower than mean mortality relative to patient population characteristics, and 48 of the included Healthcare Institutions experiencing a SHMI >1 i.e. experiencing a greater than mean mortality relative to the patient population characteristics.

Figure 4 – Mean 2018 GMC Score for Included Healthcare Institution by GMC domain. As demonstrated a wide variance in GMC survey scores were reported. Notably clinical supervision (mean = 88.63; SD = 2.43) and clinical supervision out of hours (mean = 85.99; SD = 3.19) received the highest ratings, and were most strongly correlated with SHMI as reported in Table 1.

498 **Tables**

	r	p
Adequate Experience	-0.07	0.516
Clinical Supervision	-0.34	0.002**
Clinical Supervision Out of Hours	-0.47	<0.001**
Curriculum Coverage	-0.1	0.379
Educational Governance	-0.18	0.118
Educational Supervision	-0.1	0.386
Feedback	0.08	0.506
Handover	-0.1	0.377
Induction	-0.07	0.517
Local Teaching	-0.1	0.403
Regional Teaching	0.14	0.226
Reporting Systems	-0.14	0.208
Rota Design	-0.29	0.008*
Satisfaction	-0.25	0.027*
Study Leave	-0.18	0.109
Supportive Environment	-0.17	0.128
Teamwork	-0.32	0.004*
Workload	0.02	0.835

Table 1 – Pearson Correlation Coefficients of All 2018 GMC Domains with 2019 NHS Trust Summary Hospital-Level Mortality Indicator. Clinical supervision and clinical supervision out of hours were significantly correlated with SHMI after Bonferonni correction demonstrating a weak and moderate strength association respectively. Prior to Bonferroni correction rota design, overall satisfaction, and teamwork were also significantly correlated with SHMI demonstrating very weak, weak, and very weak associations respectively.

*Statistically significant without Bonferroni correction (p = 0.05)

499

500

501