BMJ Open

Effect of implementing the NEWS2 escalation protocol in a large acute NHS trust: a retrospective cohort analysis of mortality, workload and ability of early warning score to predict death within 24 hours

Sarah Forster ,[1,2] Tricia M McKeever,[3] Dominick Shaw[1,2]

[1]Respiratory Medicine, University of Nottingham School of Medicine, Nottingham, UK
[2]Respiratory Medicine, Nottingham University Hospitals NHS Trust, Nottingham, UK
[3]Division of Epidemiology and Public Health, University of Nottingham, Nottingham, UK

Correspondence to
Dr Sarah Forster;
sarahforster@nhs.net

ABSTRACT

Objectives To describe the inpatient population, establish patterns in admission and mortality over a 4-year period in different cohorts and assess the prognostic ability and workload implications of introducing the National Early Warning Score 2 (NEWS2) and associated escalation protocol.

Design Retrospective cohort analyses of medical and surgical inpatient admissions.

Setting Large teaching hospital with tertiary inpatient care and a major trauma centre employing an electronic observations platform, initially with a local early warning score, followed by NEWS2 introduction in June 2019.

Participants 332 682 adult patients were admitted between 1 January 2016 and 31 December 2019.

Outcome measures Mortality, workload and ability of early warning score to predict death within 24 hours.

Results Admissions rose by 19% from 76 055 in 2016 to 90 587 in 2019. Total bed days rose by 10% from 433 382 to 477 485. Mortality fell from 3.7% to 3.1% and was significantly lower in patients discharged from a surgical specialty, 1.0%–1.2% (p<0.001). Total observations recorded increased by 14% from 1 976 872 in 2016 to 2 249 118 in 2019. 65% of observations were attributable to patients under medical specialties, 34% to patients under surgical specialties. Recorded escalations to the registrar were stable from January 2016 to May 2019 but trebled following the introduction of NEWS2 in June 2019.

Conclusions There was an increase in hospital inpatient activity between 2016 and 2019, associated with a reduction in mortality and percentage of observations calculated as reaching threshold NEWS2 score of 7 for escalation to the registrar. The introduction of the NEWS2, with a higher sensitivity and lower specificity, when allied to its escalation protocol, was associated with a significant increase in actual recorded escalations to the registrar. This was more marked in the surgical population and would support refining threshold scores based on admission characteristics when developing the next iteration of NEWS.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ Large dataset allowed adequate power for subdivision into medical and surgical cohorts.
⇒ Granularity of data allowed different early warning scoring systems to be compared retrospectively.
⇒ Findings are applicable to all hospitals using NEWS2.
⇒ Changes in hospital staffing and clinical policies that may have impacted outcomes could not be fully accounted for in the analysis.

INTRODUCTION

The NHS is facing unprecedented challenges with rising admissions and increasing patient need on a background of finite resources and staffing challenges all heightened by the demands due to the COVID-19 pandemic. Understanding hospital workload is key to managing resource allocation to provide safe and efficient patient care. A major factor determining workload is predicting and responding to the clinical deterioration of patients in hospital. This response is mostly driven by early warning scoring systems. Currently, NEWS2 (National Early Warning Score version 2), and its associated escalation protocol, is mandated across the NHS in England and Wales. Introduced in 2017, it uses routinely collected vital sign data to produce a combined score. If a set threshold score is reached, a change in frequency of vital signs monitoring, medical review and intervention is recommended by the associated escalation protocol.[1]

Despite NEWS2 being the principal early warning score used in the NHS, there is little evidence that introducing it into a hospital improves mortality and morbidity when compared with other methods of identifying

unwell patients. The introduction of early warning scores, and actions dictated at set cut points, is not without consequence; the workload generated by the escalation protocol based on the thresholds advised by NEWS2 may lead to alert fatigue[2] and diversion of a clinician's time from patients in whom intervention may alter trajectory, particularly in systems where automated escalations occur.

Although there is a single universal early warning score deployed in adult inpatients, a hospital population is not homogeneous. In order to understand the ability of NEWS2 to predict clinical deterioration, and the consequent workload generated by escalation of threshold scores as dictated by the associated protocol, the background mortality rate and variation within different populations of inpatients need to be explored. This is vital as both the positive predictive and the negative predictive values, and therefore potential workload implications of early warning scores, depend on the mortality rate (ie, prevalence) in the population in which they are being used.

Any changes to NEWS2 on its revision in 2023 are likely to have a significant impact on future hospital workload and staffing needs in the NHS. We set out to describe the inpatient population and different characteristics of the two cohorts described; to establish whether there were any changes in admission numbers and mortality in a large teaching hospital over a 4-year period; and to assess the impact of applying two different early warning scores, including analysis of the real-time introduction of NEWS2 and its associated protocol, on workload.

METHODS
Design and setting
This study consisted of retrospective cohort analyses of medical and surgical inpatient admissions. The study setting, Nottingham University Hospitals Trust (NUHT), consists of two hospitals with 1500–1700 overnight beds, depending on demand. NUHT is a regional referral centre for Neurosurgery, Cardiac and Thoracic Surgery as well as several medical specialties. In addition, it has a level 1 trauma centre ensuring a consistent flow of admissions, requiring flexible capacity, and placing significant demands on staff.

Data source
In the summer of 2014, an electronic observations software system (Nervecentre) enabling vital signs observations to be inputted at the bedside using handheld devices was deployed across the hospital. The same software collated the vital sign data into an early warning score. This local early warning score (LEWS) was similar to the later NEWS2 score (see online supplemental table 1 for comparison) but had scoring cut points that allowed slightly more deranged vital signs before scoring thresholds were reached, a graded score for both inspired oxygen and AVPU (A graded ssessment of conscious level using the categories of Alert, responds to Voice, responds

to Pain and Unresponsive), the inclusion of urine output and the exclusion of oxygen saturations. Another deviation from NEWS2 was the presence of different profiles to adjust for changes to baseline physiology in the setting of chronic disease. For example, being anuric for a period of more than 18 hours would score 3 on LEWS, but a chronic anuria profile was available for patients on renal replacement therapy to avoid unnecessary escalation. If the LEWS was elevated beyond set threshold scores, the software system automatically prompted clinical intervention/escalation through requests for medical review, transmitted to the clinical staff via a mobile phone. In June 2019, the software was amended to employ the NEWS2 algorithm; automatic escalation at set thresholds continued as dictated by the NEWS2 protocol published by the Royal College of Physicians.[1] The combined system captures all vital sign data, early warning scores and automated requests for clinical intervention/escalation.

This software system was used to create a database of vital signs observations, including heart rate, blood pressure, respiratory rate, oxygen saturation, fraction of inspired oxygen, temperature, conscious level and urine output, linked to outcome and demographics, admitted between 1 January 2016 and 31 December 2019, following Health Research Authority and Information Governance approvals.

Statistical analysis
Trends in admission rates, mortality, length of stay and early warning scores were analysed, both within the whole population and subpopulations of surgical and medical specialties defined by admission specialty. Patients admitted to specialised day-case areas were removed from the analysis as these represent a very different population subset and are managed in a different manner.

Defining risk factors for mortality and length of stay
Univariate logistic regression analysis was used to identify which variables (from those available at admission) had a significant association with in-hospital mortality or longer length of stay, beyond the population median. Variables assessed included age, gender, discharge from hospital in the preceding 30 days, NEWS2 score at admission and time, day and month of admission. These were built into multivariable models. Variables that remained significant for inclusion were inputted into the final models in order to illustrate differences in behaviour between the two cohorts.

Evaluation of workload
Workload (defined as request for medical review, or escalation) as protocoled by the early warning scoring system was assessed in several ways.

Recorded escalations
Data on the number of actual requests for review/escalations were measured both pre (using LEWS) and post the change to NEWS2 in June 2019.

Observations reaching threshold score for escalation

In addition to recorded activity, predicted escalations based on the scoring thresholds for both early warning scores used during the study period were calculated. An assumption that all scores reaching the protocoled threshold for intervention/escalation would have led to an intervention was made for the purposes of this study and any observation with a score at or above that thresholds was considered a predicted positive. That is, for LEWS all scores of ≥4 were counted as flagged to a junior doctor and scores of ≥6 counted as flagged to registrars; and for NEWS2 all scores of ≥5 were counted as being flagged to a junior doctor and scores of ≥7 as being flagged to registrars.

Where the specified score was not in use at the time of the observation set being recorded, as was the case with NEWS2 before June 2019 and LEWS score after June 2019, the score was calculated from the component vital signs. NEWS1 was also calculated to determine the level of difference when compared with NEWS2. The individual components to calculate NEWS2, including the new confusion component of ACVPU (AVPU with the additional category of new confusion), were all recorded as part of the system prior to its introduction in June 2019 and are therefore present consistently in the dataset. When applying oxygen saturation targets as part of NEWS2, scale 2 was employed in patients with a diagnosis of COPD (Chronic Obstructive Pulmonary Disease) and scale 1 to those without. Although this is not in line with Royal College of Physicians guidance for the use of scale 2 in NEWS2, there is precedent in the literature that applying target saturations of 88%–92% in all patients with a diagnosis of COPD requiring oxygen improves outcome.[3] In addition, this allowed for a consistent approach both before and after the introduction of NEWS2.

Analysis of score performance

These data were used to assess the performance of the score across the different cohorts. This included calculation of sensitivity and specificity at the protocol cut points applicable to each score, that is, 4 and 6 for LEWS and 5 and 7 for NEWS2, as well as area under the receiver operating characteristic (ROC) curve for the score as a whole for identifying the outcome of death within 24 hours of an observation.

The final metric calculated was the number needed to evaluate (NNE or workup detection ratio). NNE has been proposed as a method for comparing the ability of early warning scores to accurately predict clinical deterioration in the context of the workload they generate and provide an indirect measure of the cost-efficiency of each alert and of the early warning score employed. In this paper, the method of Kipnis et al[4] was used and is defined as the number of observations that is necessary to respond to in order to pick up one outcome of death within 24 hours.[5 6]

$$NNE = \frac{FP + TP}{TP} = \frac{1}{PPV}$$

Here, false positives refer to observations reaching threshold not followed by outcome of death within 24 hours and true positives refer to observations reaching threshold that were followed by death within 24 hours.

It should be noted that in both LEWS and NEWS2 the nursing and clinical staff looking after a patient have the ability to pause escalation or adjust parameters based on the clinical situation. Therefore, the number of recorded escalations is expected to be lower than the number of observations which met the threshold for escalation.

All analyses were carried out in STATA V.17. Approval was given by the UK Health Research Authority (IRAS ID 270837) and Nottingham University Hospitals Trust's Caldicott guardian, Research and Innovation team and Information Governance department (Ref: DG20-000049-D and IG0025). As the study did not involve human participants and was limited to routinely collected data anonymised prior to extraction, the HRA did not require research ethics committee review.

In addition to these analyses, a freedom of information request was sent to all acute trusts in the NHS in England to further determine the applicability of these findings. The information requested included whether the trust employed electronic observations systems, what platforms were used, whether automatic escalations were triggered at threshold scores and what those thresholds were.

Patient and public involvement

None.

RESULTS

Admissions and length of stay

332 692 adult patients were admitted between 1 January 2016 and 31 December 2019 (table 1). This excluded 23 156 patients admitted under obstetrics who were managed using a different scoring system and 22 138 patients admitted to day-case units. Median age at admission was static at 61 years throughout the 4-year study period. 8 788 (2%) were discharged from the emergency department without specialty referral and were included in total numbers but not analysed separately; 198 300 (60%) were admitted under medical specialties and 125 604 (38%) under surgical specialties (flow diagram included as online supplemental figure 1).

Admissions rose by 19% from 76 055 in 2016 to 90 587 in 2019. Total bed days rose 10% from 433 382 to 477 485. Readmissions also rose—4.6% of admissions in 2019 had been discharged in the preceding 30 days compared with 3.5% in 2016, accounting for 5.3% of surgical admissions and 3.8% of medical admissions. There was a small decrease in median length of stay in patients under medical specialties from 2.1 days in 2016 to 1.9 days in 2019. Length of stay was static among patients admitted under surgical specialties as shown in table 1.

The reduction in length of stay seen in patients under a medical specialty was partly attributable to a reduction in bed days by patients who had been declared medically

Table 1 Hospital admissions, medicine and surgery between 2016 and 2019

	2016	2017	2018	2019
Admissions				
Total	76 055	81 379	84 671	90 587
Surgery	29 120	31 006	31 890	33 588
Medicine	46 182	49 034	51 483	51 601
Admissions within preceding 30 days (% of total admissions)				
Total	2646 (3.5)	3274 (4.0)	3811 (4.5)	4196 (4.6)
Surgery	1259 (4.3)	1537 (5.0)	1795 (5.6)	1780 (5.3)
Medicine	1353 (2.9)	1643 (3.4)	1951 (3.8)	1967 (3.8)
Bed days	453 382	451 482	456 417	477 485
Median age (IQR)	61 (46–76)	61 (46–76)	61 (41–76)	61 (41–76)
Median LOS (IQR)				
Total	2.1 (0.8–6.5)	2.0 (0.7–5.9)	2.0 (0.7–6.0)	1.9 (0.5–5.7)
Surgery	1.3 (0.4–4.1)	1.3 (0.4–4.2)	1.4 (0.4–4.3)	1.4 (0.4–4.2)
Medicine	3.1 (1.0–8.2)	2.7 (0.9–7.5)	2.7 (0.9–7.4)	2.9 (0.9–7.5)
Mortality (% of admissions)				
Total	2821 (3.7)	2853 (3.5)	2772 (3.3)	2818 (3.1)
Surgery	335 (1.2)	352 (1.1)	319 (1.0)	347 (1.0)
Medicine	2481 (5.3)	2493 (5.1)	2449 (4.8)	2368 (4.8)

Bed days, total beds occupied for 24 hours a day; LOS, length of stay; Medical, all admissions under medical specialties; Surgery, all admissions under surgical specialties; Total, Surgical + Medical + Emergency department.

fit for discharge and were waiting placement (a bed in a care home or social service input). The equivalent of 174 beds were occupied for an entire year in 2016 and 156 in 2019 by patients who had been declared medically fit for discharge; even at the lower level seen in 2019 this equates to almost six 28 bedded wards being occupied for a whole year by patients ready for discharge. Sixty-three per cent (97) of these bed years were accounted for by patients aged over 75 years.

The multivariable logistic regression analysis demonstrated that across all admissions, patients were more likely to have a longer length of stay if older, female, had a NEWS2 score of 5 or more at admission, presented overnight or were admitted in winter (see online supplemental table 2). In patients under a surgical specialty, having been discharged in the preceding 30 days was associated with lower likelihood of length of stay greater than 2 days, whereas in medicine a previous discharge within 30 days was associated with greater risk of length of stay longer than 2 days.

Mortality
Mortality rates varied with season when examined monthly but remained relatively constant over the study period. However, because of the increase in overall admissions the percentage mortality decreased year on year. Mortality was significantly lower in patients discharged from a surgical specialty at 1.0%–1.2% compared with 4.8%–5.3% in patients discharged from a medical specialty (p<0.001).

Several variables were associated with risk of mortality and were common to both medical and surgical patients (table 2). Patients were more likely to die in hospital if they were older, had a NEWS2 score of 5 or more at admission, had been discharged in the preceding 30 days or presented in the winter. Surgical patients were more likely to die in hospital if presenting overnight or were female. Medical patients were more likely to die if male and risk of mortality in medical patients overall decreased with admission year. The degree to which each of these variables was associated with risk of mortality was different when comparing medical and surgical patients, with NEWS2 of 5 or more and older age associated with a higher risk of mortality in a surgical than medical population.

Observations and early warning scores
Over the 4 years of the study, total observations recorded increased by 14% from 1 976 872 to 2 249 118 as shown in table 3 below, with median observations per patient per day rising from 3 to 4 (online supplemental figure 2). If time taken to record observations is assumed to be 3 min 45 s,[7] this equates to an increase of 85 000 minutes a month. Sixty-five per cent of observations were attributable to patients under medical specialties, 34% to patients under surgical specialties and 1% to patients discharged by the emergency department. The median admission NEWS2 remained stable at a score of 1.

Workload
Recorded escalations to the medical registrar were relatively stable between January 2016 and June 2019.

Table 2 Multivariate analysis for factors associated with a significantly higher or lower risk of mortality

Baseline variables	OR	Adj OR	95% CI	P value
Total population				
Age (years)				
18–40	0.07	0.09	0.08 to 0.11	<0.001
41–60	0.37	0.42	0.39 to 0.46	
61–75	1.00			
76–85	1.76	1.70	1.62 to 1.80	
86+	2.97	2.87	2.72 to 3.03	
NEWS2>5	8.57	6.41	6.14 to 6.69	<0.001
Admissions preceding 30 days	1.77	1.62	1.55 to 1.70	<0.001
Admission quarter				
Dec–Feb	1.00			<0.001
Mar–May	0.80	0.86	0.81 to 0.91	
Jun–Aug	0.77	0.87	0.83 to 0.93	
Sep–Nov	0.80	0.92	0.87 to 0.97	
Sex (females vs male)	0.92	0.91	0.87 to 0.95	<0.001
Presenting overnight (1700–0800)	1.28	1.14	1.10 to 1.19	<0.001
Year of admission				
2016	1.00	1.00		0.003
2017	0.94	0.97	0.92 to 1.03	
2018	0.88	0.92	0.87 to 0.98	
2019	0.83	0.91	0.86 to 0.96	
Surgery				
Age (years)				
18–40	0.07	0.07	0.05 to 0.12	<0.001
41–60	0.30	0.32	0.26 to 0.39	
61–75	1.00			
76–85	2.34	2.29	1.99 to 2.63	
86+	5.48	5.19	4.75 to 6.40	
NEWS2≥5	11.63	9.49	8.24 to 10.92	<0.001
Sex (females vs male)	1.13	1.14	1.02 to 1.28	0.018
Admissions preceding 30 days	1.42	1.43	1.25 to 1.64	<0.001
Admission quarter				
Dec–Feb	1.00			<0.001
Mar–May	0.68	0.69	0.59 to 0.81	
Jun–Aug	0.76	0.79	0.68 to 0.92	
Sep–Nov	0.74	0.80	0.69 to 0.94	
Presenting overnight (1700–0800)	1.81	1.75	1.56 to 1.96	<0.001
Medicine				
Age (years)				
18–40	0.10	0.10	0.08 to 0.12	<0.001
41–60	0.41	0.46	0.43 to 0.50	
61–75	1.00			
76–85	1.51	1.49	1.41 to 1.58	
86+	2.28	2.25	2.12 to 2.38	
NEWS2>5	6.18	5.01	4.78 to 5.23	<0.001
Sex (females vs male)	0.78	0.82	0.79 to 0.86	<0.001
Admissions preceding 30 days	1.62	1.56	1.49 to 1.64	<0.001

Table 2	Continued				
Baseline variables	OR	Adj OR	95% CI	P value
Admission quarter				
Dec–Feb	1.00			<0.001
Mar–May	0.83	0.89	0.84 to 0.95	
Jun–Aug	0.79	0.89	0.84 to 0.95	
Sep–Nov	0.83	0.94	0.89 to 1.00	
Admission year				
2016	1.00			0.046
2017	1.04	0.97	0.91 to 1.03	
2018	0.93	0.93	0.87 to 0.98	
2019	0.82	0.93	0.88 to 0.99	

However, there was an increase of approximately 300% following the change from LEWS to NEWS2 in June 2019 (see figure 1 and table 4), when registrar escalations rose approximately 932 a month to over 3000 a month. This could mean an estimated additional 172–2068 hours a month depending on whether a 5 min review of the observations and notes or a full assessment and management plan, taking an estimated 60 min, is required. On reviewing the number of times that a patient was escalated to the registrar within a 24-hour period, an increase was seen across the spectrum. Patients escalated once in a 24-hour period rose from 2500 a month before the introduction of NEWS2 to 5000 a month after the introduction of NEWS2 and its associated escalation protocol. At the other end of the range, patients escalated more than 10 times in a 24-hour period rose from an average of 110 a month to an average of 486 a month following the introduction of NEWS2 and its escalated protocol (see online supplemental figure 3).

This rise in recorded escalations is not reflected in patterns of predicted escalations calculated from retrospective analysis of LEWS and NEWS2 when each is looked at independently over the 4 years of the study. Both scores show a trend towards decrease in vital signs observation scores reaching the respective threshold scores for junior doctor and registrar escalation (table 4). When examining percentage of scores above cut point that were escalated, using the score in use at the time the observation set was recorded just over 40% of observations with LEWS of 6 or more were escalated to the registrar while more than 60% of observations with a NEWS2 of 7 or more were escalated to the registrar.

When using NEWS2 retrospectively to calculate NNE, for every outcome of death within 24 hours detected at a threshold cut point of 7, 14.2–15.4 observations sets met the threshold for escalation (compared with 14.7–15.6 when applying NEWS1). The NNE for surgical patients at a threshold score of 7 was 16.8–21.8 observation sets for every death detected within 24 hours. In medical patients, 13.8–14.7 observation sets met the threshold for escalation for every death within 24 hours. LEWS had a lower NNE at the threshold for registrar escalation. However, actual escalations did not match the predicted number of observations reaching threshold for escalation using either the LEWS prior to June 2019 or NEWS2 after June 2019 (table 4) as a proportion of escalations was stopped by the clinical team if felt to be unnecessary or inappropriate.

The area under ROC curve for NEWS2 in predicting outcome of death within 24 hours was similar between the two patient populations (0.910, 95% CI 0.908 to 0.911 in medicine and 0.912, 95% CI 0.907 to 0.917 in surgery). These values are comparable with similar study populations[8] and with the original NEWS1 protocol (0.911, 95% CI 0.090 to 0.913 in medicine and 0.919, 95% CI 0.914 to 0.924 in surgery). There was a significant reduction in area under ROC curve over time suggesting that by this measure NEWS2 was less able to predict which observations would be followed by death in 24 hours in 2019 than in 2016.

National use of NEWS2 and application of escalation protocols
Seventy-four trusts covering over 100 hospitals across England responded to the freedom of information request sent out in March 2022. Sixty-five out of the 74 trusts employed electronic observation platforms (see online supplemental figure 4 for distribution of trusts responding) to deploy NEWS2 and its associated escalation protocol, with 24 different platforms reported as being in use. Two further trusts indicated that they were looking to deploy electronic observations in the future. Twelve of these trusts reported employing automatic escalation of observations to the registrar. The cut points reported for escalation varied, with reported thresholds scores of 4–5 to more junior doctors, 5 or 7 to the registrar and 5 or 7 to critical care outreach teams (CCOTs). Trusts not deploying automated escalation reported relying on a combination of nursing staff escalation based on advised actions at set scores, and dashboards displaying threshold scores across the hospital to highlight high scores. One trust reported using an additional risk assessment based on highest score in the preceding 12 hours alongside current NEWS score to assist in clinical judgement.

Table 3 Patterns of early warning score by specialty group and year

	2016	2017	2018	2019
Number of observations				
Total	1 976 872	1 995 823	2 067 015	2 249 118
Surgery	627 359	651 865	672 519	720 919
Medicine	1 345 812	1 337 457	1 388 273	1 515 547
Median observations per patient per day (IQR)				
Total	3 (2–5)	3 (2–5)	4 (2–5)	4 (2–6)
Surgery	4 (2–5)	4 (2–5)	4 (2–6)	4 (3–6)
Medicine	3 (2–5)	3 (2–5)	3 (2–5)	4 (3–6)
Median admission NEWS2 (IQR)				
Total	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–2)
Surgery	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–1)
Medicine	1 (0–3)	1 (0–3)	1 (0–3)	1 (0–3)
Median NEWS2 (IQR)				
Total	1 (0–2)	2 (0–2)	1 (0–2)	1 (0–2)
Surgery	1 (0–2)	1 (0–2)	1 (0–2)	1 (0–1)
Medicine	1 (0–3)	1 (0–3)	1 (0–3)	1 (0–3)
% Observations followed by death in 24 hours				
Total	0.48	0.50	0.46	0.40
Surgery	0.20	0.22	0.17	0.15
Medicine	0.59	0.63	0.60	0.51
Sensitivity for death in 24 hours NEWS2 5				
Total	0.81	0.81	0.79	0.74
Surgery	0.79	0.78	0.74	0.64
Medicine	0.82	0.81	0.80	0.76
Specificity* for death in 24 hours NEWS2 5				
Total	0.89	0.89	0.89	0.92
Surgery	0.93	0.93	0.93	0.95
Medicine	0.87	0.87	0.87	0.90
Sensitivity for death in 24 hours NEWS2 7				
Total	0.62	0.62	0.60	0.57
Surgery	0.58	0.61	0.53	0.47
Medicine	0.63	0.63	0.61	0.58
Specificity for death in 24 hours NEWS2 7				
Total	0.96	0.96	0.96	0.97
Surgery	0.98	0.98	0.98	0.99
Medicine	0.95	0.95	0.95	0.96
Area under ROC curve for death in 24 hours of NEWS2 (95% CI)				
Total	0.921 (0.918 to 0.924)	0.918 (0.915 to 0.921)	0.909 (0.906 to 0.913)	0.910 (0.907 to 0.914)
Surgery	0.928 (0.920 to 0.936)	0.920 (0.912 to 0.929)	0.901 (0.890 to 0.913)	0.891 (0.878 to 0.903)
Medicine	0.915 (0.912 to 0.919)	0.913 (0.909 to 0.916)	0.904 (0.900 to 0.908)	0.908 (0.904 to 0.912)

*Specificity here refers to the percentage of observations not followed by death in 24 hours which fell below the threshold for escalation.

DISCUSSION

Our data reveal an increase in the number of admissions year on year, with a smaller increase in bed days, associated with a trend towards a decrease in length of stay. This is consistent with figures reported by the King's Fund into NHS activity which described a reduction of 3000 beds across the NHS from 2016 to 2019.[9] Over the same period, inpatient elective and emergency attendances rose by 9%
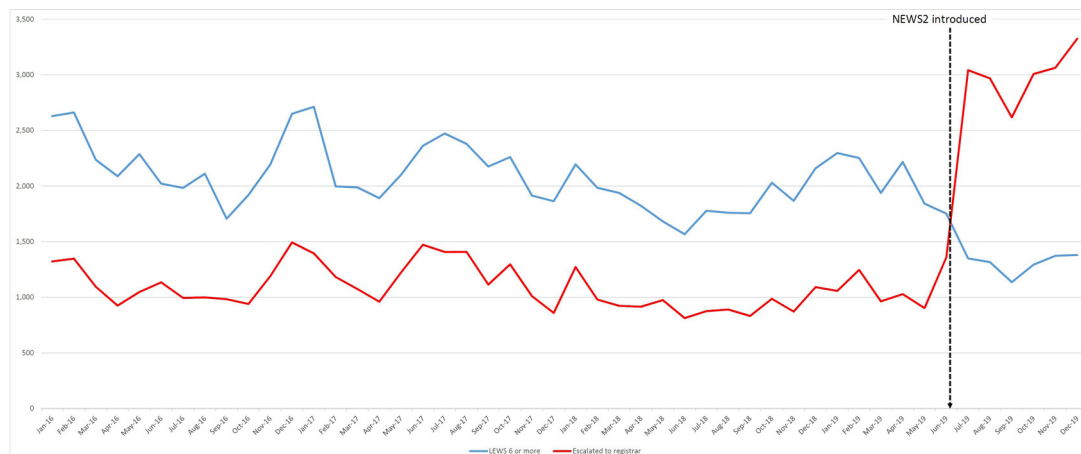
Figure 1 Pattern of actual number of observations escalated to the registrar versus e number of observations reaching LEWS escalation threshold by month and year. LEWS-is the local early warning score in use prior to the introduction of NEWS2 in June 2019.

nationally compared with 19% in this dataset.[10] Mortality reduced from 3.7% to 3.1% in the overall hospital inpatient population between 2016 and 2019. This fall in mortality is consistent with overall patterns of mortality in Nottinghamshire between 2016 and 2019 and is not offset by a higher proportion of deaths in the community[11]

The two different early warning scores, LEWS and NEWS2, had varying effects on workload as defined by recorded escalations to the on-call team or the registrar. Despite both scores showing downwards trends in observations reaching the threshold for escalation over the course of the study, the recorded escalations to the registrar more than trebled partway through 2019 from an average of 932 a month in the 6 months July to December 2018 to an average of 3062 a month in the 6 months July to December 2019, when NEWS2 was introduced. It is not possible to match the cut points of the two scores as the shapes of the receiver operating curves means they do not overlap except at extremes of sensitivity and specificity (online supplemental figure 5), where any cut point would be meaningless. However, NEWS2 has a higher sensitivity and lower specificity than LEWS at the escalation thresholds with equivalent actions, resulting in a higher number of escalations to the registrars, including a rise in patients being escalated multiple times in a 24-hour period (online supplemental figure 3). In addition to the statistical performance, the human factors element of introducing a new early warning score should also be considered. A higher proportion of observations reaching threshold score was escalated to the registrar following the introduction of NEWS2 than had been the case with the previous local score. One explanation for this is familiarity with the score. It is possible that given lack of experience with a new score staff felt less able to use their own judgement where it contradicted the protocol. It is also possible that the Hawthorne effect played a role due to the increased monitoring that was carried out in the months after NEWS2 and its associated protocol was introduced.

It could be argued that an increase in NNE reflected a reduction in adverse outcomes as more patients were reviewed by senior staff. However, the corollary is that reviewing all patients all the time will reduce adverse events, but in a resource-limited system this is not possible and a doctor reviewing a patient with a high score cannot perform a task elsewhere.

The increase in escalations is problematic from a workforce perspective as there has been no associated increase in registrar numbers in training. Without any changes to either workforce numbers or NEWS2 escalation thresholds, either there will be a delay in the clinical review for some patients or an impact on the ability of registrars to complete other aspects of their job important to training and development. One way of responding to this challenge is through the use of specialist nurses acting as a first point of call for deteriorating patients and liaising with CCOT.

Studies of CCOT introduction found improvement in staff in decision making and early access to ICU.[12] However, robust evidence remains lacking[13] and CCOT comes at an increased cost. Another response to rising workload is through greater empowerment of nursing staff in terms of their assessment of patient condition.[12–14]

The prognostic ability of NEWS2, as measured by area under ROC curve for outcome of death within 24 hours, reduced between 2016 and 2019. One contributory factor could be a change in the way conscious level was recorded on the Nervecentre platform. Analysis of vital signs patterns before and after the change in early warning score to NEWS2 demonstrated a drop in observations coded as having reduced conscious level according to ACVPU (a tool that rates conscious level based on whether someone is Alert, has new Confusion, is responsive to voice or pain or is unresponsive), following the introduction of NEWS2. The reduction in mortality over time may also have influenced the decline in the performance of NEWS2.

Table 4 Predicted escalations by scores reaching threshold and actual recorded escalations by year

	2016	2017	2018	2019*
Actual recorded registrar escalations				
Total	13 468 (0.7)	14 399 (0.7)	11 420 (0.6)	24 577 (1.1)
Surgery	3161 (0.5)	3410 (0.5)	2658 (0.4)	5173 (0.7)
Medicine	10 304 (0.8)	10 988 (0.8)	8754 (0.6)	19 376 (1.3)
Median recorded registrar escalations per day				
Total	14 (6–23)	13 (5–22)	11 (5–18)	16 (7–33)
Surgery	5 (3–9)	5 (3–10)	4 (2–8)	6 (3–11)
Medicine	13 (5–23)	12 (5–22)	10 (4–17)	15 (7–30)
NEWS2 scores reaching threshold of 7 (% total)†				
Total	80 505 (4.3)	83 732 (4.3)	80 438 (3.9)	63 085 (3.0)
Surgery	13 757 (2.3)	14 188 (2.2)	11 917 (1.8)	9680 (1.4)
Medicine	66 731 (5.2)	69 515 (5.3)	68 504 (5.0)	53 335 (3.8)
LEWS scores reaching threshold of 6 (% total)				
Total	26 484 (1.3)	26 116 (1.3)	22 537 (1.1)	20 144 (0.9)
Surgery	4732 (0.8)	4545 (0.7)	3436 (0.5)	3192 (0.4)
Medicine	21 734 (1.6)	21 566 (1.6)	19 091 (1.4)	16 937 (1.1)
Number needed to evaluate for outcome of death in 24 hours NEWS2 score of 7†				
Total	15.4	14.2	15.1	14.5
Surgery	20.2	16.8	21.8	20.8
Medicine	14.7	13.8	14.3	13.8
Number needed to evaluate for outcome of death in 24 hours LEWS 6				
Total	7.9	7.0	7.1	8.2
Surgery	9.6	8.0	9.4	11.6
Medicine	7.6	6.8	6.8	7.8
Actual number needed to evaluate from recorded escalations to registrar				
Total	10.3	9.4	10.9	12.4
Surgery	9.3	8.9	9.8	16.3
Medicine	10.7	9.5	11.2	11.7

*Following introduction of NEWS2, LEWS high scores may be overestimated as AVPU recording changed.
†NEWS2 introduced to NUHT in June 2019—before this point calculated retrospectively.
AVPU, A graded ssessment of conscious level using the categories of Alert, responds to Voice, responds to Pain and Unresponsive; LEWS, local early warning score; NUHT, Nottingham University Hospitals Trust.

To fully understand the opportunity cost of NEWS2, it is helpful to examine the differences between patients admitted under surgical specialties in comparison to those admitted under medical specialties, given differences in underlying pathology, treatment strategies and trajectory in these patient groups. This comparison reveals distinct differences. Patients admitted under surgical specialties accounted for 37% of admissions, 32% of observations, but only 16% of scores over 7 and 12% of deaths. Medical inpatients had a longer length of stay, higher mortality and higher number of admissions.

In analysing the factors associated with length of stay and mortality, several were common to both medical and surgical inpatients. Advancing age, high NEWS score at admission and time of year were risk factors for increased mortality and longer length of stay in all populations.

However, there were differences in terms of the impact of gender, readmission and time of day admitted. For example, being female was associated with a higher mortality risk in surgery,[15] potentially contributed to by the inclusion of cardiac surgery patients,[16] but a lower risk in medicine. Admission overnight, between 5 pm and 8 am, was associated with an increased mortality risk in surgery but not medicine when other factors were adjusted for in the multivariate model. A discharge within the 30 days preceding the current admission was associated with a shorter than median length of stay in surgery and a longer than median length of stay in medicine.

These differences in outcome could be used to influence the composition and application of early warning scoring systems in terms of thresholds for escalation; however, any difference in the applicability of NEWS2

has to be balanced against the benefits of having a single standardised score for any deteriorating adult patient in terms of familiarity with score and benchmarking of care. Moreover, monitoring of patients is reliant not just on the score used, but how the absolute score, and any need for clinical review is communicated to medical staff and the clinical response to it. The Freedom of Information response shows significant variation in how NEWS2 has been adopted across the NHS. This includes the use of different threshold cut points in different trusts, a differential response to the threshold scores and varying staffing responses in term of seniority/experience. It is apparent that despite a single mandated national scoring system the response to a deteriorating patient is still varied.

This could suggest that, as with the chronic respiratory scale, many hospitals developed alongside the first NEWS. NHS hospitals are finding ways to use NEWS2 that are compatible with their system and staffing resources. This highlights the fact that although a single system has benefits, there may need to be refinement regarding the cut points applied in the real world and, ideally, prospective studies to refine implementation. This would confirm whether specific cut points for patients in different specialty areas, for example medical versus surgical specialties, would be beneficial as our data and others[6] suggests.

There is also disparity in deployment and make up of rapid response teams (acute response teams—RRT/medical emergency teams (MET)). These generally consist of a number of on-call doctors including those with critical care or airway skills. The MERIT study was a cluster randomised control trial of the introduction of MET to 23 hospitals. It reported that despite a higher number of emergency referrals, the introduction of an MET did not lead to a reduction in mortality,[17] although this may reflect length of follow-up, as a further study reported lower mortality after a longer period and a change in team composition.[12 13]

This study has several strengths. The size and completeness of this dataset make it possible to include multiple variables in analyses while maintaining statistical power. The inclusion of four full years mitigates the impact of a single unusual year of either higher or lower mortality. The inclusion of the period in which NEWS2 was introduced allows actual rather than predicted impact of the score on escalations to be analysed. The use of multiple measures of efficiency not only allows comparison with previous studies but also gives a more complete picture of the situation.

Only one previous study has examined the differing performance of NEWS in medical versus surgical inpatients.[8] The proportion of observations followed by death within 24 hours was 0.21% in surgery and 0.69% in medicine comparable to our study. The primary method for judging score performance in these populations was area under ROC curve for outcome of death, ICU admission, cardiac arrest and combined within 24 hours of an observation. By this measure, the performance of NEWS2 was not significantly different between the two groups and performed at least as well in surgical patients as medical patients, a result replicated in this study. The authors also used a measure of workload and detection (sensitivity), both of which clearly showed a difference between the populations. Again, results were comparable to this study, despite their use of combined outcomes to report this metric in comparison to our use of death within 24 hours as an outcome. This supports the view that these two cohorts represent distinct populations with different characteristics requiring different management.

The data on the mortality trend in the second half of 2019 was in keeping with the downward trend seen in previous years. However, multiple new ways of working have been introduced into hospitals over the years, including focusing on falls, pressure care, sepsis 6, early consultant review, etc; consequently, it is very difficult to establish which of these may or may not have affected mortality. In addition, having only a few months of trend after the introduction of NEWS2, due to the emergence of COVID-19 in 2020, means it is not possible to distinguish any effect of the new score from seasonal fluctuations in disease. As with all studies analysing vital signs linked to outcome, it is only possible to see the association with scores and that outcome. It is not possible to determine where a high score has triggered an intervention that averts an outcome as intended or where factors known to impact outcome such as staffing[18] are not available. In order to establish a causative link between use of NEWS2 and mortality, a randomised control trial would be needed. The use of death within 24 hours as an outcome, rather than cardiac arrest, ICU admission or combined, means this study cannot be compared directly with the only previous study examining NEWS in surgical and medical populations.[8]

In conclusion, our study illustrates clear differences in population characteristics and mortality between patients admitted under medical and surgical specialties and an associated difference in ability of NEWS2 to predict outcome at the current protocol thresholds. The increase in escalations following switch to NEWS2 also highlights the potential workload impact of changes to scores and associated escalation protocols. These factors should be taken into account when developing the next iteration of NEWS2 and while a single scoring system has many benefits, there is need for ongoing refinement following real-world evaluation, ideally using prospective studies that can accurately observe and evaluate the response to triggers.

in final approval of submission.TMMcK contributed to design of study, statistical support and interpretation of data, editing of article and was involved in final approval of submission. DS conceived the idea for the work, contributed to design of study, interpretation of data, editing of article and was guarantor author.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Due to the nature of the data sharing agreements and DPIA agreed with the Nottingham University Hospitals Trust Information Governance Department, the University and the HRA the study data cannot be shared outside of the study group.

ORCID iD
Sarah Forster http://orcid.org/0000-0001-9424-6543

REFERENCES

1 Royal College of Physicians of, L. National early warning score (news) 2: standardising the assessment of acute-illness severity in the NHS; 2017.
2 Yiu CJ, Khan SU, Subbe CP, *et al*. Into the night: factors affecting response to abnormal early warning scores out-of-hours and implications for service improvement. *Acute Med* 2014;13:56–60.
3 Echevarria C, Steer J, Wason J, *et al*. Oxygen therapy and inpatient mortality in COPD exacerbation. *Emerg Med J* 2021;38:170–7.
4 Kipnis P, Turk BJ, Wulf DA, *et al*. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform* 2016;64:10–19.
5 Romero-Brufau S, Huddleston JM, Escobar GJ, *et al*. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care* 2015;19:285.
6 Pankhurst T, Sapey E, Gyves H, *et al*. Evaluation of NEWS2 response thresholds in a retrospective observational study from a UK acute Hospital. *BMJ Open* 2022;12:e054027.
7 Dall'Ora C, Griffiths P, Hope J, *et al*. How long do nursing staff take to measure and record patients' vital signs observations in hospital? A time-and-motion study. *Int J Nurs Stud* 2021;118:103921.
8 Kovacs C, Jarvis SW, Prytherch DR, *et al*. Comparison of the National early warning score in non-elective medical and surgical patients. *Br J Surg* 2016;103:1385–93.
9 Ewbank L TJ, Mckenna H, Anandaciva S, Ward D. NHS Hospital bed numbers: past; present, future The Kings Fund; 2021.
10 Fund TKs. The NHS in a nutshell, 2020. Available: https://www. kingsfund.org.uk/projects/nhs-in-a-nutshell
11 Statistics, O.o.N, 2022
12 Valentine J, Skirton H. Critical care outreach--a meaningful evaluation. *Nurs Crit Care* 2006;11:288–96.
13 Tillmann BW, Klingel ML, McLeod SL, *et al*. The impact of delayed critical care outreach team activation on in-hospital mortality and other patient outcomes: a historical cohort study. *Can J Anaesth* 2018;65:1210–7.
14 Kenward G, Hodgetts T. Nurse concern: a predictor of patient deterioration. *Nurs Times* 2002;98:38–9.
15 Chana P, Joy M, Casey N, *et al*. Cohort analysis of outcomes in 69 490 emergency general surgical admissions across an international benchmarking collaborative. *BMJ Open* 2017;7:e014484.
16 Eifert S, Guethoff S, Kaczmarek I, *et al*. Applying the gender lens to risk factors and outcome after adult cardiac surgery. *Viszeralmedizin* 2014;30:99–106.
17 Hillman K, Chen J, Cretikos M, *et al*. Introduction of the medical emergency team (Met) system: a cluster-randomised controlled trial. *Lancet* 2005;365:2091–7.
18 Needleman J, Buerhaus P, Pankratz VS, *et al*. Nurse staffing and inpatient hospital mortality. *N Engl J Med* 2011;364:1037–45.